

Theoria cum praxi

Reinhold Scholl

Das Motto des Universalgelehrten Gottfried Wilhelm Leibniz, der als einer der Väter der Digital Humanities gelten kann, weil er die Voraussetzungen für die binäre Welt der Einser und Nullen mitgeschaffen hat, gilt auch für diese Zeitschrift.

Die in diesem Heft versammelten Beiträge eint, daß sie es mit der Anwendung der Arbeitsweisen der Digital Humanities auf bestimmte Fragestellungen zu tun haben. Das geht von der Forderung nach einer sinnvollen und nachvollziehbaren Erstellung von Stoppwörtern, über die händische Codierung von Texten, um sie über den Umweg der Maschinenlesbarkeit wieder den Menschen verständlich zu machen, der Auswertung einer personenbezogenen Datenbank im unteren Donauraum in griechisch-römischer Zeit auf der Basis von Inschriften bis hin zu einer textminigbasierten Wortanalyse zur Klärung einer Ernährungsfrage in Athen in klassischer Zeit.

Der Beitrag von Patrick J. Burns beschäftigt sich mit dem Erstellen von sogenannten Stoppwörtern. Alle Tools, die die automatische Bearbeitung von größeren Textcorpora zum Ziel bzw. als Hilfsmittel haben, waren von Anbeginn an daran interessiert, um Rechen- und Rechnerkapazität zu sparen sowie um aussagefähige Suchergebnisse zu erhalten, Stoppwörter aus ihren Analysen heraus zu nehmen. Burns bemängelt in seinem Artikel u.a., daß die Kriterien für diese Listen oft nicht transparent gemacht, nicht immer eingehalten und auch nicht dokumentiert wurden. Oft kommt zu den Algorithmen noch eine subjektive Bearbeitung hinzu. Zu beachten ist ebenfalls, daß die Stoppwörterlisten, auch wenn Sie nach klaren Algorithmen gebildet werden, immer abhängig sind von dem jeweils zugrundeliegenden Textcorpus, aus dem sie gewonnen werden, nämlich von dessen zeitlicher Erstreckung und von ihrem Inhalt. Diesen undurchschaubaren Stoppwörterlisten für einzelne Sprachen und besonders für Latein stellt er seinen Beitrag gegenüber. Sein "paper describes the development of a generalizable method for building stoplists in the Classical Language Toolkit (CLTK), an open-source Python platform for natural language processing research on historical languages." Das Plädoyer für mehr Transparenz und auch Dokumentation kann man nur unterstützen. Andererseits stellt sich einem mehr geisteswissenschaftlich vorgebildeten interessierten IT-Laien die Frage, ob bei der heutigen Kapazität der Rechner Beschränkungen wie die Herausnahme von Stoppwörtern in negativen oder positiven Stoppwörterlisten noch notwendig sind oder besser, ob man es nicht lieber den Anwenderinnen und Anwendern überlassen sollte, ob sie dieses Tool der Stoppwörterlisten nutzen wollen, zumal es von der Fragestellung abhängig ist, welche Wörter zur Bedeutung und Deutung beitragen. Stoppwörter erfüllen zwar per definitionem in erster Linie grammatikalische und syntaktische Funktionen und tragen weniger zum Inhalt, wohl aber zum Sinn und zur Bedeutung bei. Das versteht sich z.B. bei Verneinungen von selbst und Konjunktionen können komplexe Zusammenhänge verdeutlichen. Bei einer Untersuchung zum Sprachstil (z.B. Parataxe-Hypotaxe, Litotes, usw.) sind diese Stoppwörter zumindest für die Statistik nicht irrelevant. Strenggenommen müßte für jede Untersuchung eine eigene Stoppwörterliste erstellt werden.

Bei dem Beitrag „Digital Fabius Pictor“ von Kevin Strassburger könnte man ketzerisch die Frage stellen, ob man wirklich einen Mehrwert für die Leserinnen und Leser generiert, wenn man den Text einer analogen Ausgabe, in der sich die Schriftsetzer alle Mühe gegeben haben, die unterschiedlichen Informationen auch drucktechnisch anschaulich zur Verfügung zu stellen, wie z.B. die verschiedenen Schriften (griechische und lateinische Buchstaben), Markierungen mit Linien, um den Originaltext von der Übersetzung zu trennen, wie in der Ausgabe von Müller geschehen, in eine digitale Form bringt oder Peters Ausgabe, der statt der Übersetzung einen zweigeteilten Apparat bietet, der unter dem griechischen Originaltext durch Absatzstriche voneinander getrennt einen lateinischsprachigen Kommentar und davon wiederum getrennt einen textkritischen Apparat mit all den für Nichtphilologen geheimnisvollen bzw. kryptisch anmutenden Abkürzungen, die bisweilen Anfänger zum Verzweifeln bringen können, bietet, digitalisiert. Die Abkürzungen in diesen Ausgaben, die in gewisser Weise Herrschaftswissen symbolisieren, sind u.a. verwendet worden, um Platz- und Druckkosten zu sparen. Diese hätte man im Zuge der Digitalisierung auflösen und mit einem gesonderten `Tag` versehen können, um das Original von der Bearbeitung zu unterscheiden. Ein weiterer Schritt wäre dann die Umwandlung des maschinenlesbaren Textes wieder in einen für Menschen lesefreundlichen Text mit all den zusätzlichen Annotationen, die, wie der Verfasser auch schreibt, durchaus in unterschiedlicher Form – nach Möglichkeit sogar wahlweise – angezeigt werden könnten. Wenn dann noch alle Wörter eine morphologische Analyse erhalten und eventuell ihre syntaktische Funktion erklärt wird und nicht nur wie hier die Personennamen (antike und moderne, wobei aber kein Unterschied zwischen ihnen gemacht wird), Ortsnamen und Literatur und die mit dem Layout verbundenen Dinge annotiert sind, ist der Mehrwert einer digitalen Ausgabe sicherlich größer und es werden zukünftig bessere Suchen und auch Auswertungen ermöglicht. Dazu war diese praktische Handarbeit der Digitalisierung und deren Beschreibung ein erster Schritt.

Die Auswertung der Personendatenbank (<http://romans1by1.com>) durch Rada Varga und Imola Boda ist die Fortsetzung eines früheren Beitrages in dieser Zeitschrift (<https://journals.ub.uni-heidelberg.de/index.php/dco/article/view/35822>). Auf der Homepage der Datenbank (zuletzt eingesehen am 5.11.2018) steht: “Romans 1by1 is a population database recording people identified in Greek and Roman epigraphy. The areas integrally covered at this point are the provinces of Moesia Inferior, Moesia Superior and Dacia. From Pannonia Superior, we have integrated so far the urban centers of Poetovio, Neviodunum, Siscia, Savaria, Brigetio and Scarbantia. Romans 1by1 is a dynamic tool and can be considered a constant work in progress.” Diese sehr gut strukturierte und dokumentierte Datenbank liefert nun erste Auswertungsergebnisse hinsichtlich der Art der Inschriften, ihrer Sprache, der Zusammensetzung der Bevölkerung in Bezug auf Alter, Geschlecht, Rechtsstatus, Tätigkeiten in Verwaltung, Militär, Kult und Vereinswesen. Im Zuge der Personendatenbank wurde sicherlich auch mehr oder weniger ein Corpus der untersuchten Inschriften erstellt, die – sofern bereits digitalisiert – über Verlinkungen für die Nutzerinnen und Nutzer als Volltext zur Kontrolle einzusehen sind. Was spricht eigentlich dagegen, diese Inschriften zukünftig auch als Gesamtkorpus zur Verfügung zu stellen?

Der Titel des Beitrages von Sven Brandt „κρίθη oder πύρος? - Eine systematische Annäherung im klassischen Athen mit Hilfe der Kookkurrenzanalyse von eAQUA und den Word Frequency Statistics der Perseus Digital Library“ deutet bereits an, daß die beiden im Titel genannten Verfahren der Digital Humanities nicht ohne weitere „menschliche“ Zuarbeit zu hilfreichen Argumenten oder Ergebnissen bei der Beantwortung der Frage führen. Es wird deutlich gemacht, daß diese Verfahren noch nicht ganz ausgereift sind, sondern der Überprüfung und Kontrolle durch die Anwenderinnen und Anwender bedürfen. „Blended Readings“ heißt das neue Schlagwort. *Distant reading* und *close reading* (vgl. [Schubert DCO 1,1 2015](#)) werden dabei

miteinander durch klassische Diskursanalyse und Hermeneutik ergänzt. Provozierend gesagt, was man in den Altertumswissenschaften methodisch, aber noch nicht computerunterstützt, bisher fast immer getan hat: Zu einer Fragestellung, einem Problem, das man lösen wollte, hat man die entsprechenden Quellen gesammelt, gesichtet, geordnet, analysiert und ausgewertet und schließlich in adäquater Form und nachvollziehbar präsentiert. Genau das geschieht auch im Beitrag von Sven Brandt. Die Quellensammlung erfolgt über die entsprechenden Termini in den zur Verfügung stehenden Datenbanken. Die Auswahl der Autoren beschränkt sich auf ihre Wirkungszeit im 5. und 4. Jh. v. Chr. in Athen. Dazu ist das Frequenzanalysetool von Perseus ein gutes Hilfsmittel. Das Problem ist allerdings, daß dieses Tool bei Homonymen (Homographen) nicht unterscheiden kann, zu welchem Wort jeweils die gleichaussehenden Grapheme gehören. Hier muß dann der menschliche Verstand die Zuordnung vornehmen. Das gleiche Problem der Homographie/Homonymie taucht auch in der Kookkurenzanalyse von eAQUA auf. Allein schon aus diesem Grund können diese beiden Tools nicht allein ohne Überprüfung und Kontrolle angewandt werden, ohne zu verfälschenden Ergebnissen zu gelangen. Textminingtools ersetzen nicht Sprachkompetenz, im Gegenteil, sie setzen diese in einem hohen Maße voraus. Das Ergebnis des Beitrages zeigt, daß in den literarischen Quellen nicht der Gegensatz Gerste oder Weizen, sondern das Gemeinsame, nämlich Weizen und Gerste betont wird. In der Tat bringt die textmininggestützte Analyse der literarischen Quellen der klassischen Zeit in Athen nur eine Annäherung – aber immerhin die – an den Komplex zur Getreidefrage in Athen (Bedarf, Anbau, Import, Konsum), aber natürlich keine Lösung. Eine weitere Annäherung könnte darin bestehen, dieses Verfahren auch auf die Inschriften anzuwenden und auch auf Autoren, die zwar später lebten, aber die klassische Zeit behandeln. Und schließlich könnte man als weitere Ergänzung die entsprechenden „Halb-“ und „Fertigprodukte“ wie ἄλφιτον (Gerstenmehl) und μᾶζα (Gerstenkuchen) auf der einen und ἄλευρον (Weizenmehl) und ἄρτος (das Brot (aus Weizen)) auf der anderen Seite mit in die Untersuchung aufnehmen. Denn die meisten Stadtbewohner Athens kamen lediglich mit diesen Halb- und Fertigprodukten in Kontakt und für deren Herstellung gab es die στροποιοί (Müller/Bäcker). Auch diese Vorgehensweise ist ein Beispiel für die Sinnhaftigkeit einer ganzheitlichen Betrachtungsweise.

Alle Beiträge dieses Heftes zeigen anschaulich und auf unterschiedliche Art und Weise, daß die Zuhilfenahme von Verfahren und Arbeitsweisen der Digital Humanities den Classics helfen können.

Autorenkontakt¹

Prof. Dr. Reinhold Scholl

Universität Leipzig
Historisches Seminar, Alte Geschichte
Beethovenstr. 15
04107 Leipzig

Email: scholl@uni-leipzig.de

¹ Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren. Alle Inhalte dieses Beitrages unterstehen, soweit nicht anders gekennzeichnet, der Lizenz CC BY 4.0.

Constructing Stoplists for Historical Languages¹

Patrick J. Burns

Abstract: Stoplists are lists of words that have been filtered from documents prior to text analysis tasks, usually words that are either high frequency or that have low semantic value. This paper describes the development of a generalizable method for building stoplists in the Classical Language Toolkit (CLTK), an open-source Python platform for natural language processing research on historical languages. Stoplists are not readily available for many historical languages, and those that are available often offer little documentation about their sources or method of construction. The development of a generalizable method for building historical-language stoplists offers the following benefits: 1. better support for well-documented, data-driven, and replicable results in the use of CLTK resources; 2. reduction of arbitrary decision-making in building stoplists; 3. increased consistency in how stopwords are extracted from documents across multiple languages; and 4. clearer guidelines and standards for CLTK developers and contributors, a helpful step forward in managing the complexity of a multi-language open-source project.

1. Introduction

Stoplists are lists of words that have been filtered from documents prior to text analysis tasks, usually words that are either high frequency or that have low semantic value.² Such words have been described as lexical “noise” which prevents “signal,” that is semantically or thematically significant content, from being accurately discriminated.³ For this reason, stoplists have been called “negative dictionaries,” that is lexica of unwanted terms.⁴ Stopwords tend to be articles, particles, prepositions, conjunctions, pronouns, and other, often indeclinable, function words, but depending on the disposition of the corpus used as the basis for the stoplist, they can include other parts of speech. For example, forms of the verb “to be” and similar high frequency verbs are not uncommon. Through the removal of such words, text analysis tasks, such as text classification, text summarization, and information retrieval, benefit in areas like noise reduction, feature reduction, or speed optimization.

1 An early version of this paper was presented at the Global Philology: Big Textual Data workshop at Universität Leipzig in the summer of 2017. The author would like to thank the workshop’s organizer, Thomas Koentges, and the workshop’s participants for feedback and suggestions. The paper has also benefited from the feedback of the Classical Language Toolkit’s community of open-source contributors.

2 Manning et al. (2008) 26, Rasmussen (2009). On the origins of the terms “stop word/stopword” and “stop list/stoplist,” see Flood (1999).

3 Luhn (1957), Rijsbergen (1975) 14–17, Salton/McGill (1983) 71–72.

4 Salton/McGill (1983) 71.

Stopwords are extracted from a document or document collection on the basis of some formalization of the relative importance of individual words. Christopher Manning, for example, offers the following strategy for building stoplists:

Sort the terms by collection frequency (the total number of times each term appears in the document collection), and then...take the most frequent terms, often hand-filtered for their semantic content relative to the domain of the documents being indexed, as a stop list, the members of which are then discarded during indexing.⁵

Yet “collection frequency” is not the only basis used for constructing stoplists, and research has shown that it is also not necessarily the best basis. Accordingly, several other methodologies have been proposed. Stoplists have been developed, for example, based on inverse document frequency and entropy measures in addition to or in combination with collection frequency. Furthermore, research in this area has moved decidedly in the direction of limiting, or in some way formalizing, the arbitrary decision-making—that is, the “hand-filtered” stage of Manning’s method—that often went into earlier stoplist construction.⁶

In addition to the basis by which stopwords are extracted from a document collection, there are additional considerations that go into constructing these lists. Stoplists can be generic or domain-specific, that is based on a broad cross-section of the language or based on a restricted, well-defined collection of texts, respectively.⁷ Different domains have different definitions of signal and noise, and building domain-specific lists decreases the likelihood that important terms will be lost in stopword removal. Considerations of preprocessing are also involved in stoplist construction. Accordingly, decisions such as whether to remove numbers from documents or retain case need to be made when constructing stoplists. Lastly, a decision needs to be made about the size of the list, a consideration for which there is no consensus and which varies widely by application.⁸

The Classical Language Toolkit (CLTK) is an open-source Python platform for natural language processing (NLP) research on historical languages.⁹ The project offers NLP support for the languages of Ancient, Classical, and Medieval Eurasia with the goals of compiling analysis-friendly corpora, collecting and generating linguistic data, and acting as a free and open platform for generating scientific research for fields such as historical linguistics and comparative philology among others. The CLTK offers a panoply of NLP resources and tools in service of these goals. Stoplists are among these offerings and the following languages are represented in some capacity at the time of writing: Arabic, Classical Greek, Classical Hindi, Biblical Hebrew, Latin, Marathi, Middle English, Old English, Old French, Old Norse, Punjabi, and Sanskrit. That said, the aim of the CLTK is to offer as complete a set of tools as possible for each language in its scope, as well as a set of tools and resources that is, within reason, consistent across languages. In this respect, there is room for both increased coverage and improved quality.

5 Manning et al. (2008) 27.

6 Fox (1989) 19, concerned about “arbitrary decisions,” was already noting the importance of consulting “empirical studies of word frequencies.”

7 Zaman et al. (2011) 133 uses the terms “arbitrary” and “tailored.”

8 Fox (1989) considers 421 words to be “maximally efficient” for general English application, but the reasons for this are not entirely clear. Saini/Rakholia (2016) report an average length of 200 words for stoplists across 42 languages in both Latin and non-Latin scripts.

9 Johnson et al. (2018).

At present, and excluding CLTK offerings, NLP resources are not always readily available for historical languages, stoplists included.¹⁰ Generic stoplists have long been available online for Classical Greek and Latin, for example via the Perseus Project and stopwords-json.¹¹ Recently, Aurélien Berra has published a repository of Greek and Latin stoplists for use with Voyant Tools, which represents a big step forward in both the presentation and documentation of historical-language stoplists.¹² There has also been recent interest in building stoplists for Sanskrit.¹³ But for the most part, historical languages lack this kind of support completely and, when it is available, the absence of documentation, provenance, and related issues of reproducibility and citation hamper their use.

The Perseus Project stoplists can serve as an example of the difficulties present when working with existing lists. There is no documentation to explain how these lists were constructed; that is, there is no information about the corpus used, no information about the basis by which it was constructed, and no explanation for the length of the list, among other considerations. To this can be added that, by the standards of current code-driven research, the code used to construct the lists is not available.¹⁴ This is compounded by the fact that the Perseus lists are not published as self-contained datasets: they do not provide version control or persistent identifiers for proper citation. All of these factors should concern a historical-language researcher faced with the need to use a stoplist for a text analysis task.

The goal of this paper is to present a generalizable method in the CLTK for the construction of stoplists and to describe the implementation of this method in the CLTK Stop module. By abstracting stoplist construction and adding language-agnostic stopword extraction tools to the project, the CLTK is able to offer consistency and standardization in historical-language tool development. These tools can then be customized to support language-specific requirements. In addition, by providing a generalizable stoplist construction module to CLTK contributors, the project is better able to manage the complexity of a multi-language open-source project.

The remainder of this paper is organized as follows: Section 2 offers a literature review of computational work on building stoplists, with special attention to languages other than English. Section 3 describes the architecture of the CLTK and where the Stop module fits, as well as the logic used in writing classes and methods for stoplist construction. Section 4 presents the results of this method for a sample language, namely Latin. By way of conclusion, Section 5 presents the benefits of the Stop module and proposes future directions for development.

10 Interestingly, Weinberg (2004) 129 traces the origin of the stoplist to a 16th-century biblical Hebrew concordance that included a one-page list of “function” words to be excluded. On the limited availability of resources, this could in fact be broadened to include lower-resourced modern languages as well; for example, see Makrehchi/Kamel (2008) 224, who identify the lack of non-English resources for text-retrieval tasks as a primary motivation for developing stopword extraction algorithms.

11 These lists can be found at <https://www.perseus.tufts.edu/hopper/stopwords> and <https://github.com/6/stopwords-json/blob/master/dist/la.json>, respectively. The Tesseract project (<https://tesseract.caset.buffalo.edu/>) uses a default stoplist based on corpus frequency. Note too that there is a desideratum for Greek and Latin stoplists, to say nothing of other historical languages, in other prominent distributions, for example in the Python package stop-words, available at <http://pypi.python.org/pypi/stopwords>.

12 Berra (2018). For Voyant Tools, see Rockwell et al. (2012).

13 Raulji/Saini (2016), Raulji/Saini (2017).

14 See, for example, Peng (2011).

2. Related Research

Applications for stopword removal include improving information retrieval, document classification, sentiment analysis, and readability assessment, among others.¹⁵ Much of the early research in this area was done on the English language, beginning in particular with H. P. Luhn's theorization of information retrieval.¹⁶ But more recently, stopword studies have been published for a variety of languages, including Ahmaric, Arabic, Chinese, Greek, Gujarati, Hebrew, Malay, Mongolian, Punjabi, Sanskrit, Thai and Yoruba, among others.¹⁷ Some studies have produced generic stoplists for the languages in question. In English, for example, the Brown stoplist and the Van stoplist have been used widely and have been referred to as "classical" stoplists.¹⁸ In the language-specific studies listed above, stoplists have been derived both from simple word frequency as well as from other weighted document measures (for example, inverse document frequency, mean probability, variance probability, entropy measures, and combinations thereof). Other methods have also been proposed, including "term-based random sampling," the use of "combinatorial values," and the use of "term adjacency."¹⁹ Another study employed a rules-based approach so as to avoid entirely a statistical approach to stopword extraction for the Gujarati language.²⁰

Some issues that appear often in stoplist research include the issue of generic versus domain-specific list construction, preprocessing, and stoplist size. Early stoplist work focused on the construction of generic reference lists.²¹ But recent trends in stoplist research favor the development of generalizable methods that can construct lists on-the-fly for any given domain.²² There is no consensus as to preprocessing decisions for stoplist construction, and, at any rate, these decisions are prone to be highly language-specific. That said, papers such as Savoy (1999) and Zaman et al. (2011) offer examples of how such early decisions can be formalized before further processing. Lastly, there is the issue of stoplist size. This appears to be art as often as science. The DIALOG search interface, for example, took a minimalist

15 See, among many possible examples, Fox (1989), Silva/Ribeiro (2003), Lo et al. (2005), Daowadung/Chen (2012), Saif et al. (2014).

16 Luhn (1957), Luhn (1958), Luhn (1960).

17 Ahmaric: Miretie/Khedkar (2018); Arabic: El-Khair (2006); Chinese: Zou et al. (2006); Yao/Ze-wen (2011); French: Savoy (1999); Greek: Lazarinis (2007); Gujarati: Rakholia/Saini (2017); Hebrew: HaCohen-Kerner/Shmuel (2010); Malay: Chekima/Alfred (2016); Mongolian: Zheng/Gaowa (2010); Persian: Sadeghi/Vegas (2014); Punjabi: Puri et al. (2013), Kaur/Saini (2016); Sanskrit: Raulji/Saini (2016), Raulji/Saini (2017); Thai: Daowadung/Chen (2012); Yoruba: Tijani et al. (2017).

18 See, for example, Lo et al. (2005) 17. The Perseus Project stoplists have attained a kind of "classical" status within the Greek and Latin research community, though their limitations are noted above.

19 Term-based random sampling: Lo et al. (2005); combinatorial values: Choy (2012); term adjacency: Rose et al. (2010).

20 Rakholia/Saini (2017).

21 See, for example, Fox (1989).

22 For example, Lo et al. (2005) 17 note in their stoplist study: "Each collection of documents is unique. It is therefore sensible to automatically fashion a different stopword list for different collections in order to maximise the performance of an IR system."

approach, capping their stoplist at just nine words: “an and by for from of the to with.”²³ Berra pushes in the other direction, including 4,001 stopwords for Latin and 6,696 for Greek in the most recent versions of the Voyant Tools stoplists.²⁴

3. Description

This study describes the development and implementation of a generalizable method for building stoplists in the CLTK, namely the Stop module. The code for the Stop module, like all code in the CLTK, is written in Python (version 3.7 at the time of writing).²⁵ The Stop module is available beginning with version 0.1.84 of the CLTK and is available as part of the free, open-source platform under an MIT license.

The CLTK increasingly aims at the following development strategy: build language-agnostic abstract classes for major NLP tasks, such as word tokenization or lemmatization, which can then be inherited as necessary by language-specific subclasses. The CLTK Stop module follows this “object-oriented philology” development pattern.

23 Harter (1986) 88. In digital classics, this can be compared to the minimalist approach taken by the Tesseract Project in their default search settings, where the stoplist, as given in exported files, is also composed of nine words: *qui, quis, sum, et, in, is, non, hic, ego, and ut*. With respect to minimalist stoplists, note also Schofield et al. (2017) on stopword effectiveness in topic modeling: “Except for the dozen or so most frequent words, removing stopwords has no substantial effect on model likelihood, topic coherence, or classification accuracy.”

24 Berra’s lists are greatly enlarged by the inclusion of inflectional variants; so, for example, all morphological possibilities of the verb *sum* (“to be”) are included in the Latin list without consideration of the relative frequency of individual forms. It is worth noting that Nothman et al. (2018) suggest that the inclusion of inflectional variants “generally seems an advisable path towards improved consistency.”

25 The complete code for the Stop module is available in the CLTK GitHub repository at <https://github.com/cltk/cltk/blob/master/cltk/stop/>.

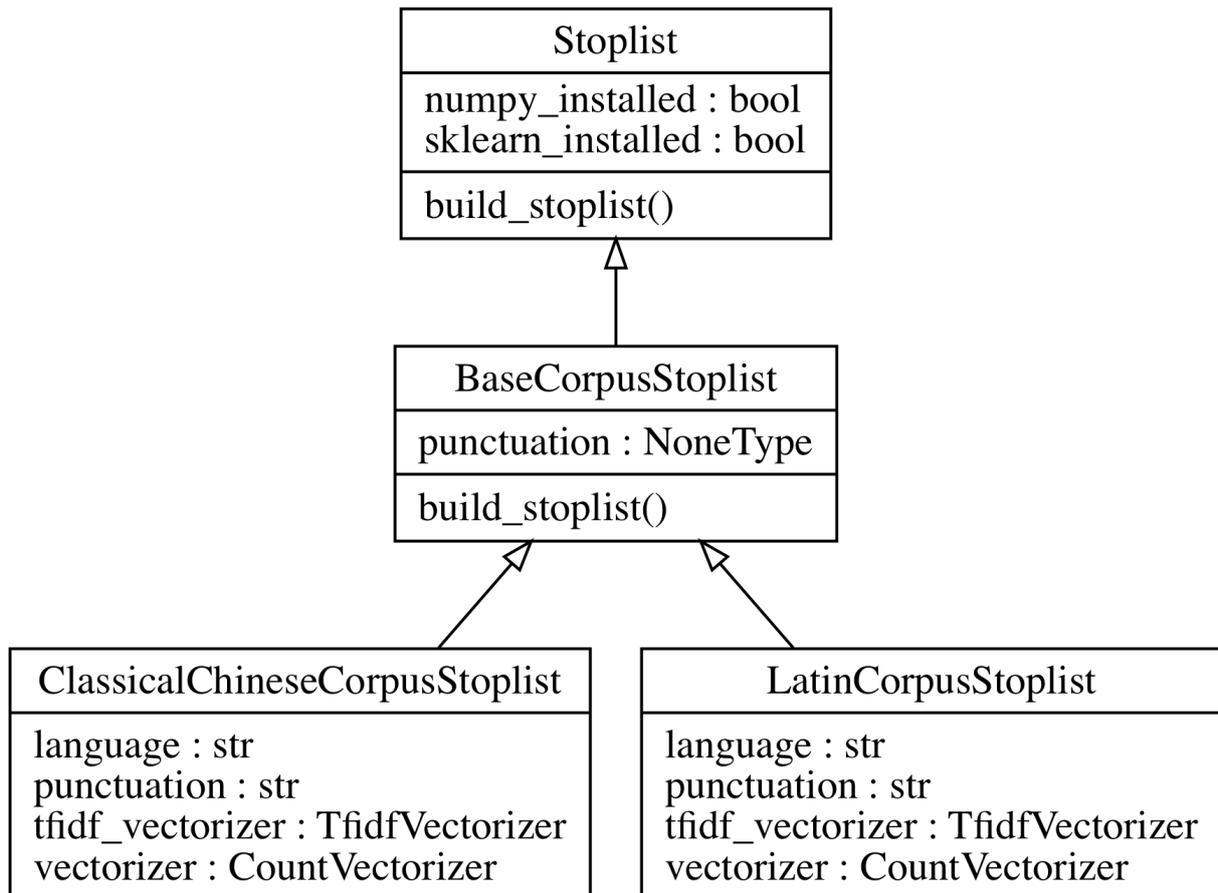


Fig 1: The inheritance tree for the CLTK Stop module. Note how language-specific classes, such as `ClassicalChineseCorpusStoplist`, inherit the `build_stoplist` method from `BaseCorpusStoplist` while overriding the language and punctuation attributes as necessary.

The module has the following structure, as shown in Figure 1:

1. There is a `Stoplist` abstract class that sits at the top of the inheritance hierarchy. Its main role is basic initialization: first, it allows a language to be set in case there are language-specific parameters that need to be set, and, secondly, it confirms that package dependencies for the class, namely Numpy and Scikit-learn, are installed.²⁶ `Stoplist` also contains an abstract method called “`build_stoplist`,” which serves as a placeholder method to be overridden by subclasses of `Stoplist`. At the module level, that is in the file `stop.py` in the Stop module, the `Stoplist` abstract class has an additional subclass: `BaseCorpusStoplist`, which will be the focus of the remaining discussion in this section.
2. `BaseCorpusStoplist` inherits and overrides the “`build_stoplist`” method from `Stoplist`. Furthermore, it is customized by way of an array of parameters that need to be set for the task of stoplist construction. As Fox mentioned in his landmark stoplist publication for English: “Selecting a stop word list is more difficult than it appears.”²⁷ Likewise, “selecting” the parameters that need to be considered when constructing a Python class and related functions for this task poses similar challenges. A primary goal in writing a generalizable method then is to eliminate, or at least formalize, the “various arbitrary

²⁶ Numpy: Oliphant (2006); Scikit-learn: Pedregosa et al. (2011).

²⁷ Fox (1989) 21.

decisions” that are often required.²⁸ Providing a class dedicated to producing domain-specific stoplists in a systematic manner reduces the number of arbitrary decisions that need to be made, and when decisions are made, it allows these decisions to be declared explicitly in parameter selection or otherwise documented directly in supporting code. Accordingly, stoplists built using the “build_stoplist” method meet two of the goals for stoplist construction stated in the introduction, that is they are well documented and replicable.

The “build_stoplist” method takes the following parameters: “texts”, “size”, “sort_words”, “inc_values”, “lower”, “remove_punctuation”, “remove_numbers”, “include”, “exclude”, and “basis.”

The parameter “texts” take a list of strings that serve as the source texts or document collection from which stopwords should be extracted. These texts are split into tokens and it is these tokens from which stopwords are drawn; no lemmatization is performed by the “build_stoplist” method.²⁹ The parameter “size” is the number of words included in the output list. The parameter “inc_values” is a Boolean variable that allows users to produce either a bare wordlist for output (inc_values = False) or include a value related to how words in the document collection are measured (the default setting, True). So, if words are measured by simple frequency, a word appearing 99 times in the “texts” would have a value of 99. The parameters “lower,” “remove_punctuation,” and “remove_numbers” are built-in preprocessing parameters for handling case (that is, whether to make all words in the collection lowercase), punctuation, and numbers; also Booleans, they are all set to True by default.

Since even in domain-specific applications, there may be an argument for including additional words or for “culling important terms,”³⁰ there are two parameters that allow users to formalize the “hand-filtered” aspect of stoplist construction, namely “include” and “exclude.” Both parameters take a list of strings. Any words included in the former are added to the final output regardless of their value; words in the latter are removed. Both parameters are set to empty lists by default.

Lastly, the “build_stoplist” method allows users to set the basis on which words are measured. At present, the bases that are supported include: frequency, mean probability, variance probability, entropy, and a composite basis drawn from a 2006 paper on constructing stoplists by Zou et al.³¹

28 Savoy (1999) 949. For example, Rijsbergen (1975) 14, on the boundary decisions for cutting off both insignificant very high frequency and very low frequency words, writes: “A certain arbitrariness is involved in determining the cut-offs. There is no oracle which gives their values.” That said, the goal here is to make explicit in code any decisions that might otherwise fall into the category of “certain arbitrariness.”

29 It should also be noted that “build_stoplist” uses the default tokenization method of Scikit-learn’s CountVectorizer class, namely whitespace tokenization. There is provisional support for character-based languages, such as Classical Chinese; in this case, a language-specific subclass of BaseCorpusStoplist is created with CountVectorizer’s “analyzer” parameter set to analyze features at the character level. Since both CountVectorizer’s “tokenizer” and “analyzer” parameters can take a callable, there is the potential for custom tokenizers and segmenters to be used in these language-specific subclasses. This is an area of current CLTK development.

30 Fox (1989) 19–20.

31 Zou et al. (2006).

3. BaseCorpusStoplist can be further subclassed by language, as for example with ClassicalChineseCorpusStoplist and LatinCorpusStoplist. These can be kept in language-specific files (for example, classical_chinese.py or latin.py), with certain parameters, such as the punctuation set for that language or whether to analyze texts by word or character, set in advance.

As noted above, the method that has been used as the default for classes in the CLTK Stop module is that of Zou et al. (2016). In their paper, Zou et al. present an “automatic aggregated methodology” which produces a stoplist through a combination of three measures, two with a document-statistical model (namely, mean probability and variance probability) and one with an informational model (namely, entropy). The document-statistical model measures mean probability to account for overall frequency in a document collection and variance probability to account for distribution across the documents. The information model uses entropy as a measure of how informative a word is within a given document. Lists of top terms are produced for each measure and the lists are aggregated using a Borda count to produce a composite list. As shown by Zou et al., we should expect stopwords to have a high mean probability and a low variance probability.³² That is, they should both occur with great frequency in individual documents in the collection and their distribution across the documents should be stable. The CLTK Stop module, by using Zou et al.’s method as the default basis, formalizes this model, allowing users to generate statistically based composite stoplists for any document collection.

Zou et al. developed their method for use in Chinese, but their method has been used for languages such as Arabic, Persian, Punjabi, Thai and Yoruba.³³ Since it has already been shown to be effective for stoplist construction in several languages and across multiple scripts, this method seemed particularly well suited for general application in the CLTK Stop module.

4. Example: Constructing Latin Stoplists³⁴

In this section, I provide an example of historical-language stoplists created using the CLTK Stop module for the Latin language using the default ‘Zou’ basis. This list is then compared to existing lists.

I have used the CLTK’s Latin Library corpus (LL) to construct the example stoplist.³⁵ This is a corpus of plaintext files, scraped from the website “The Latin Library.”³⁶ It consists of 2,152 files, representing Latin texts in a wide range of authors, works, and genres, and spanning chronologically from the earliest Latin authors to 20th-century Neo-Latin authors. For this

32 Zou et al. (2006) 1012: “Intuitively, the probability of a word to be a stop word is directly proportional to the mean of probability, but inversely to the variance of probability.”

33 Alajmi et al. (2012), Daowadung/Chen (2012), Puri et al. (2013), Sadeghi/Vegas (2014), Tijani et al. (2017).

34 All of the code used to generate text statistics or to build and analyze the stoplists presented in this section of the paper can be found in the Jupyter notebook in the GitHub repository associated with this paper, available at <https://github.com/diyclassics/stopwords-paper>. The version of this repository at the time of publication has been archived with Zenodo at doi:10.5281/zenodo.1477208.

35 Available at https://github.com/cltk/latin_text_latin_library.

36 Available at <http://www.thelatinlibrary.com/>.

test, the texts have been minimally preprocessed; specifically, 1. HTML entities have been converted to the corresponding unicode character (that is, entities such as “>” have been converted to the “>” character), and 2. the texts have been truncated to start at the first word appearing 1,000 characters from the beginning of the text and to end at the word appearing 1,000 characters from the end in order to remove paratextual material included in “The Latin Library” texts, such as title and author information and text related to web navigation.³⁷ Case has been retained, as have numbers and punctuation. There are approximately 16.2 million tokens in this corpus.³⁸

Using the “Zou” basis with the “size” parameter set to 100 in “build_stoplist,” the Latin-specific instance of BaseCorpusStoplist produces the following stoplist for LL:

*ab ac ad ante apud atque aut autem causa cui cuius cum de dei deus dum ea ego ei eius enim eo erat ergo esse esset est et etiam eum ex fuit haec hic his hoc iam id igitur illa ille in inter ipse ita me mihi modo nam ne nec neque nihil nisi nobis non nos nunc omnes omnia omnibus per post potest pro qua quae quam quem qui quia quibus quid quidem quis quo quod quoque res se secundum sed si sibi sic sicut sine sit sub sunt tamen te tibi tu tunc ubi uel uero uos ut.*³⁹

The LL stoplist, drawn as it is from the largest collection with widest variety of genres, can be taken as the closest thing to a generic stoplist for Latin and it compares favorably to existing lists.⁴⁰ It contains 55.4% of the words (41/92) from the Perseus stoplist and 77.5% of the words (31/40) from the stopwords-json stoplist. While on the surface, these numbers may appear low, ready answers appear upon review of the lists. The Perseus list, for example, consists primarily of lemmas (that is, dictionary headwords), while the LL list contains any token that meet the statistical criteria of the stopword extraction algorithm.⁴¹ So, for example, the nominative forms of several personal pronouns, such as *ego* (“I”), *tu* (“you”), and *nos* (“we”), appear on both lists, but oblique forms, such as *mihi* (“to me”), *me* (“me”), *tibi* (“to you”), *te* (“you”, accusative) *sibi* (“himself, herself, itself”), and *nobis* (“to us”), only appear on the LL list and so count against the 100-word limit set for this example. Another factor is that there are no nouns or adjectives in the Perseus list, while by strict statistical determination certain forms of the nouns *causa* (“cause, reason”), *deus* (“god”), and *res* (“thing”) and certain forms of the adjective *omnis* (“all”) appear on the LL list. Interestingly, the situation is reversed in the stopwords-json list where the largest point of difference when compared to the LL list is the

37 Because of this truncation, files consisting of less than 1000 characters have been removed from the fileset all together.

38 The WordPunctTokenizer from the Natural Language Toolkit has been used to tokenize the texts; cf. https://www.nltk.org/_modules/nltk/tokenize/regexp.html#WordPunctTokenizer.

39 Specific information about the measurements underlying the “Zou” basis, such as mean probability, variance probability, and entropy values, can be found in Appendix A of the Jupyter notebook at <https://github.com/diyclassics/stopwords-paper/blob/master/notebooks/stopwords-paper.ipynb>.

40 Complete results of comparison between the LL stoplist and the Perseus, stopwords-json, and Voyant Tools list can be found in Appendix B of the Jupyter notebook at <https://github.com/diyclassics/stopwords-paper/blob/master/notebooks/stopwords-paper.ipynb>.

41 There are some non-lemma forms such as *es* (“you are”) and *est* (“he, she, it is”) in the Perseus list, but these words are the exception.

inclusion of three oblique forms of *res* (*re*, *rem*, and *rebus*). It is unclear why these are included (and why other forms of *res* like *rei* and *rerum* are not).⁴² As noted in Section 1, while stoplists tend to include mostly function words, this is only a tendency and perhaps a subjective one prone to the caveats against arbitrary decision-making noted by Manning. The direction taken in this paper tends towards a statistical definition of the stopword extraction process and these nouns meet the statistical criteria of the algorithm. Arguments for excluding these words can (and should) be made by other stoplist builders, but the burden will be on them to argue against the statistically derived description of a corpus of Latin text.

Conversely, because of the large size of Berra's Voyant Tools Latin stoplist, it makes more sense to reverse the direction of the comparison with this list. The Voyant Tools list contains 98% of the words in the LL stoplist. As seen in the previous comparisons, noun forms account for the missing forms: *deus* ("god") and its genitive form *dei* ("of god"). (Interesting to note that forms of *causa* and *res* have made Berra's cut.) Again, there are good arguments for exclusion of these terms from stoplists. But the comparison of the two lists demonstrates the importance of domain in creating (and by extension applying) stoplists. The Latin Library contains a large amount of post-Classical, Christian writing, in which "god" is a very high frequency token. Berra uses the Packard Humanities Institute Latin Texts version 5.3, which for the most part only covers the Latin language up to 200 CE, explaining the difference.⁴³ If a researcher planned to use a Latin stoplist with a set of Classical texts, the recommendation would be to build a custom list on a representative sample of Classical texts. As noted in the Introduction, this flexibility in stoplist construction is one of the main benefits of the CLTK Stop module.

42 Words like this can be compared to what Nothman et al. (2018) 9 call "controversial" words, that is words that appear in a stoplist though without clear statistical support.

43 See <http://latin.packhum.org/about> for a description of the PHI Latin Texts project. In Appendix C of the Jupyter notebook (<https://github.com/diyclassics/stopwords-paper/blob/master/notebooks/stopwords-paper.ipynb>), I have provided examples of three domain-specific stoplists constructed from subsets of the Latin Library corpus: one for the works of Cicero, one for the Biblia Sacra, or Latin text of the bible, and one for collections of Roman legal texts. It should be noted that *deus* is extracted by the 'Zou' algorithm from the Biblical texts, though it is not when applied to the Cicero texts or the legal texts.

5. Conclusion

The use of stoplists is not without its detractors. The origin of their use stems from physical and computational limitations, such as the avoidance of paper waste on the one hand and low processing power and memory on the other, that are increasingly obsolescent. Furthermore, their use is subject to issues of linguistic ambiguity that can be difficult to control for, like homonymy.⁴⁴ It has also been suggested that constructing domain-specific stoplists is time-consuming and subject to arbitrary interventions by the creator of the list, while at the same time not being particularly effective.⁴⁵ Nevertheless, there are strong arguments in favor of developing the CLTK Stop module.⁴⁶

First, there are many existing NLP text processing tools, largely aimed at modern language applications, that make use of stoplists or include stoplist parameters. For example, the Natural Language Toolkit provides default stoplists for 21 languages.⁴⁷ Historical-language researchers learning NLP basics using the NLTK book find themselves at a loss when the example code asks them to import a list for an unsupported language.⁴⁸ Another example is the CountVectorizer class in Scikit-learn, which allows a stoplist to be designated in its parameter list.⁴⁹ This list is by default an English stoplist. Researchers wishing to apply machine-learning methods to historical-language text through Scikit-learn have the option of using a stoplist not appropriate for their research question or are otherwise obligated to design their own list or simply ignore the parameter. By using a list created with methods from the CLTK Stop module, users can be confident that this resource is appropriate for their language, not to mention the specific domain that they are working on. Moreover, by offering methods for reading document collections and systematically building stoplists, the module pushes back against criticisms about the “time-consuming” or “arbitrary” nature of the process.

44 Savoy (1999) 945.

45 Schofield et al. (2017).

46 In addition to the arguments presented below, there are other positive uses for statistically based word extraction methods similar to or identical to stopword extraction. These lists can be used, for example, as the basis for stylometric analysis and author-attribution studies; see Arun et al. (2009).

47 The NLTK Stopwords corpus includes Arabic, Azerbaijani, Danish, Dutch, English, Finnish, French, German, (modern) Greek, Hungarian, Indonesian, Italian, Kazakh, Nepali, Norwegian, Portuguese, Romanian, Russian, Spanish, Swedish, and Turkish; see https://www.nltk.org/nltk_data/.

48 Bird et al. (2015); see section 4.1 on “Wordlist Corpora.” The development of the CLTK Stop module described in this paper has its origins in using the instructions for adding a new language to the spaCy NLP platform, which asks contributors to define a language-specific stoplist as their first task; see <https://spacy.io/usage/adding-languages>.

49 Pedregosa et al. (2011). For parameter definitions, see http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html.

The CLTK Stop module similarly pushes back against the criticisms of available historical-language stoplists mentioned in the Introduction, most especially the lack of documentation, provenance, sources, and construction methodology behind many existing resources.⁵⁰ Stoplists created using the CLTK Stop module have a clear advantage. The software itself is open source, documented, version-controlled, and citable. Moreover, any code written by a user that uses the CLTK Stop module to build a stoplist can be shared and cited, and as such the document collection used as the source will be explicitly defined in this code as will the parameters and any pre- or postprocessing applied to the lists. Accordingly, the CLTK Stop module offers better support for well-documented, data-driven, and replicable results.⁵¹

There are two additional benefits to the CLTK community in the offering of a generalizable method for stoplist construction. The first pertains to CLTK users, namely that they can be assured of consistency in how stopwords are extracted from documents across multiple languages. As I mentioned in the Introduction, the resources currently offered in the CLTK, specifically generic reference lists, have been cobbled together from a variety of sources with no underlying and unifying methodology for how these resources were developed. By offering an abstract class with language-agnostic methods that can be subclassed and then overridden and reparameterized as necessary, CLTK users can be confident that the stoplists that they create are based on research that has been shown to be applicable across a wide variety of languages but also take into account language-specific demands. So, for example, the Latin and Classical Chinese submodules both inherit from `BaseCorpusStoplist` but the former is subclassed in such a way that it tokenizes documents based on words and the latter on characters.

The second pertains to CLTK contributors. Stoplists are one element of the project's basic language resource kit for each language.⁵² That said, they are currently available for only a limited number of languages in the project's scope. Before the development of the new Stops module, contributors who wished to add a stoplist for a given language did so with minimal direction. That is, there was no attempt at consistency in extraction methodology across languages, nor was there a "quality control" standard by which existing CLTK stoplists could be evaluated.⁵³ With the introduction of the Stop module, contributors now have a clear starting point for development, a helpful step forward in managing the complexity of a multi-language open-source project.

50 This problem is not limited to historical-language NLP stoplists as shown in Nothman et al (2018). For example, on the Scikit-learn default stoplists, Nothman et al. (2018) 7 write: "Despite its popular use, the current Scikit-learn maintainers cannot justify the use of this particular list, and are unaware of how it was constructed."

51 In this respect, it is compatible with the RAD (replicable, aggregable, and data-driven) paradigm described in Haswell (2005). Note also that Nothman et al. (2018) 11 propose the following strategies for stoplist provisioning in open-source software: better documentation, dynamic adaptation to specific NLP tasks, increased quality control in assessing the inclusion of controversial terms and evaluating completeness, and wider availability of tools for automatic list generation. CLTK Stop development aims at everything on this list, especially the last item.

52 Krauwer (2003).

53 See Nothman et al (2018) 11.

This paper describes the first version of a generalizable method for building stoplists for historical languages with the CLTK. That said, work on the module continues. Here are some concluding thoughts on future directions for the module.

Zou et al.'s method has been used in stoplist research for several languages and, for this reason, is an appropriate starting point for developing generalizable methods. Still, as noted above, it is not the only method for stoplist construction. Future work will investigate how other methods for systematically building both generic and domain-specific stoplists, including term-based random sampling and term adjacency among others, could be effectively implemented in the CLTK. Further discussion of parameters available in “build_stoplist” and their applicability within the language-specific subclasses is also necessary.⁵⁴ Another area that will require additional research is the evaluation of historical-language stoplists. Just as these low-resource languages often lack stoplists all together, they also largely lack the tagged document collections necessary to serve as benchmarks in evaluation. Only with the development of more benchmark collections can we truly begin to evaluate the effectiveness of historical-language stoplists.⁵⁵ In the meantime, the CLTK Stop module offers a significant step forward in formalizing stoplist construction methodology and, in doing so, promotes consistency and maintains standards across languages for both users and developers, which in turn facilitates documentation and reproducibility of an NLP resource for historical-language research.

54 Discussions about best practices for stoplist construction in the CLTK are ongoing in the project's GitHub ‘Issues’; see, for example, the following thread on “improvements to multilingual stop module,” through which character-based tokenization for Classical Chinese was introduced to the Stop module: <https://github.com/cltk/cltk/issues/743>.

55 So, for example, supervised classifier-based stopword extraction methods, such as those used in Makrehchi/Kamel 2008, are not yet practical for most languages represented in the CLTK.

6. References

- Alajmi et al. (2012): A. Alajmi, E. M. Saad and R. R. Darwish. “Toward an Arabic Stop-Words List Generation,” *International Journal of Computer Applications* 46 (8), 8–13.
- Arun et al. (2009): R. Arun, R. Saradha, R., V. Suresh, M. Narasimha Murty and C. E. Veni Madhavan, “Stopwords and Stylometry: A Latent Dirichlet Allocation Approach,” in: *NIPS Workshop on Applications for Topic Models: Text and Beyond*, 1–4.
- Berra (2018): Aurélien Berra, *Ancient Greek and Latin Stopwords for Textual Analysis*, version 2.1.0. <https://github.com/aurelberra/stopwords>. (accessed on 30 August 2018)
- Bird et al. (2015). Steven Bird, Ewan Klein and Edward Loper, “Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit,” <https://www.nltk.org/book/>. (accessed on 30 August 2018)
- Chekima/Alfred (2016): Khalifa Chekima and Rayner Alfred, “An Automatic Construction of Malay Stop Words Based on Aggregation Method,” in: *Proceedings of Soft Computing in Data Science: Second International Conference*, 180–189.
- Choy (2012): Murphy Choy, “Effective Listings of Function Stop Words for Twitter,” arXiv preprint, <http://arxiv.org/abs/1205.6396>.
- Daowadung/Chen (2012): Patcharanut Daowadung and Yaw-Huei Chen, “Stop Word in Readability Assessment of Thai Text,” in: *2012 IEEE 12th International Conference on Advanced Learning Technologies*, 497–99.
- El-Khair (2016): Ibrahim Abu El-Khair, “Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study,” *International Journal of Computing and Information Sciences* 4 (3), 119–133.
- Flood (1999): Flood, Barbara J., “Historical Note: The Start of a Stop List at Biological Abstracts,” *Journal of the American Society for Information Science* 50 (12), 1066.
- Fox (1989): Christopher Fox, “A Stop List for General Text.” *SIGIR Forum* 24 (1–2), 19–21.
- HaCohen-Kerner/Shmuel (2010): Yaakov HaCohen-Kerner and Yishai Blitz Shmuel, “Initial Experiments with Extraction of Stopwords in Hebrew,” in: *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, 449–453.
- Harter (1986): Stephen P. Harter, *Online Information Retrieval: Concepts, Principles, and Techniques*. San Diego.
- Haswell (2005): Richard H. Haswell, “NCTE/CCCC’s Recent War on Scholarship,” *Written Communication* 22 (2): 198–223.
- Johnson (2018): Kyle P. Johnson and the Classical Language Toolkit contributors, “The Classical Language Toolkit,” <http://cltk.org/>. (accessed on 30 August 2018)
- Kaur/Saini (2016): Jasleen Kaur and Jatinderkumar R. Saini, “Punjabi Stop Words: A Gurmukhi, Shahmukhi and Roman Scripted Chronicle,” in: *Proceedings of the ACM Symposium on Women in Research*, 32–37.
- Krauwer (2003): Steven Krauwer, “The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap.” *Proceedings of International Workshop Speech and Computer (SPECOM)*, 8–15.

Lazarinis (2007): Fotis Lazarinis, “Engineering and Utilizing a Stopword List in Greek Web Retrieval,” *Journal of the American Society for Information Science and Technology* 58 (11), 1645–1652.

Lo et al. (2005): Rachel Tsz-Wai Lo, Ben He and Iadh Ounis, “Automatically Building a Stopword List for an Information Retrieval System,” in: *5th Dutch-Belgium Information Retrieval Workshop*, 17–24.

Luhn (1957): Hans Peter Luhn, “A Statistical Approach to Mechanized Encoding and Searching of Literary Information,” *IBM Journal of Research and Development* 1 (4): 309–17.

Luhn (1958): Hans Peter Luhn, “The Automatic Creation of Literature Abstracts.” *IBM Journal of Research and Development* 2 (2): 159–65.

Luhn (1960): Hans Peter Luhn, “Key Word-in-Context Index for Technical Literature (KWIC Index).” *American Documentation* 11 (4): 288–95.

Makrehchi/Kamel (2008): Masoud Makrehchi and Mohamed S. Kamel, “Automatic Extraction of Domain-Specific Stopwords from Labeled Documents,” in: *Advances in Information Retrieval*, 222–33.

Manning et al. (2012): Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge.

Miretie/Khedkar (2018): Sileshi Girmaw Miretie and Vijayshri Khedkar, “Automatic Generation of Stopwords in the Amharic Text,” *International Journal of Computer Applications* 180 (10), 19–22.

Nothman et al. (2018): Joel Nothman, Hanmin Qin and Roman Yurchak, “Stop Word Lists in Free Open-Source Software Packages,” in: *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, 7–12.

Oliphant (2006): Travis E. Oliphant, *A Guide to NumPy*, vol. 1, Spanish Fork, UT.

Pedregosa et al. (2011): Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss and Vincent Dubourg, “Scikit-Learn: Machine Learning in Python,” *Journal of Machine Learning Research* 12: 2825–2830.

Peng (2011): Roger D. Peng, “Reproducible Research in Computational Science,” *Science* 334 (6060): 1226–27.

Puri et al. (2013): Rajeev Puri, R. P. S. Bedi and Vishal Goyal, “Automated Stopwords Identification in Punjabi Documents,” *Research Cell: An International Journal of Engineering Sciences* 8, 119–125.

Rakholia/Saini (2017): Rajnish M. Rakholia and Jatinderkumar R. Saini, “A Rule-Based Approach to Identify Stop Words for Gujarati Language,” in: *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, 797–806.

Rasmussen (2009): Edie Rasmussen, “Stoplists,” in: Ling Liu and M. Tamer Özsu (eds.), *Encyclopedia of Database Systems*, Boston, 2794–96.

Raulji/Saini (2016): Jaideepsinh K. Raulji and Jatinderkumar R. Saini, “Stop-Word Removal Algorithm and Its Implementation for Sanskrit Language,” *International Journal of Computer Applications* 150 (2), 15–17.

Raulji/Saini (2017): Jaideepsinh K. Raulji and Jatinderkumar R. Saini, “Generating Stopword List for Sanskrit Language,” in: 2017 IEEE 7th International Advance Computing Conference (IACC), 799–802.

Rijsbergen (1975): C.J. van Rijsbergen, Information Retrieval, Newton, MA.

Rockwell et al. (2012): Geoffrey Rockwell, Stéfan Sinclair and the Voyant Tools team, “Voyant Tools,” <https://voyant-tools.org/>. (accessed on 30 August 2018)

Rose et al. (2010): Stuart Rose, Dave Engel, Nick Cramer, Wendy Cowley, “Automatic Keyword Extraction from Individual Documents,” in: Michael W. Berry and Jacob Kogan (eds.), Text Mining, Chichester, UK, 1–20.

Sadeghi/Vegas (2014): Mohammad Sadeghi and Jésus Vegas, “Automatic Identification of Light Stop Words for Persian Information Retrieval Systems,” Journal of Information Science 40 (4), 476–487.

Saif et al. (2014): Hassan Saif, Miriam Fernández, Yulan He and Harith Alani, “On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter,” in: Proceedings of the 9th Language Resources and Evaluation Conference (LREC), 810–817.

Saini/Rakholia (2016): Jatinderkumar R. Saini and Rajnish M. Rakholia, “On Continent and Script-Wise Divisions-Based Statistical Measures for Stop-Words Lists of International Languages,” in: Procedia Computer Science 89, 313–19.

Salton/McGill (1983): Gerard Salton and Michael J. McGill, Introduction to Modern Information Retrieval, New York City.

Savoy (1999): Jacques Savoy, “A Stemming Procedure and Stopword List for General French Corpora,” Journal of the American Society for Information Science 50 (10), 944–52.

Schofield et al. (2017): Alexandra Schofield, Måns Magnusson and David Mimno, “Pulling Out the Stops: Rethinking Stopword Removal for Topic Models,” in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 432–36.

Silva/Ribiero (2003): Catarina Silva and Bernardete Ribeiro, “The Importance of Stop Word Removal on Recall Values in Text Categorization,” in: Proceedings of the International Joint Conference on Neural Networks, 1661–1666.

Tijani et al. (2017): Olatunde D. Tijani, A. T. Akinwale, Saidat A. Onashoga and E. O. Adeleke, “An Auto-Generated Approach Of Stop Words Using Aggregated Analysis,” in: Proceedings of the 13th International Conference of the Nigeria Computer Society, 99–115.

Weinberg (2004): Bella Hass Weinberg, “Predecessors of Scientific Indexing Structures in the Domain of Religion,” in: Second Conference on the History and Heritage of Scientific and Technical Information Systems, 126–134.

Yao/Ze-wen (2011): Zhou Yao and Cao Ze-wen, “Research on the Construction and Filter Method of Stop-Word List in Text Preprocessing,” in: Fourth International Conference on Intelligent Computation Technology and Automation, 217–21.

Zaman et al. (2011): A. N. K. Zaman, Pascal Matsakis and Charles Brown, “Evaluation of Stop Word Lists in Text Retrieval Using Latent Semantic Indexing,” in: 2011 Sixth International Conference on Digital Information Management, 133–36.

Zheng/Gaowa (2010): Gong Zheng and Guan Gaowa, “The Selection of Mongolian Stop Words,” in: IEEE International Conference on Intelligent Computing and Intelligent Systems, vol. 2, 71–74.

Zou et al. (2006): Feng Zou, Fu Lee Wang, Xiaotie Deng, Song Han and Lu Sheng Wang, “Automatic Construction of Chinese Stop Word List,” in: Proceedings of the 5th WSEAS International Conference on Applied Computer Science, 1010–1015.

7. Author⁵⁶

Patrick J. Burns

New York University
Institute for the Study of the Ancient World

Email: patrick.j.burns@nyu.edu

⁵⁶ The rights pertaining to content, text, graphics, and images, unless otherwise noted, are reserved by the author. This contribution is licensed under CC BY-SA 4.0 International.

Digital Fabius Pictor

Kevin Straßburger

Abstract: The following essay describes and explains digital working at the text fragments of Fabius Pictor published by Karl Müller in 1849. For comparison reasons and as a control mechanism we also encoded the 2nd edition in 1914 of Hermann Peter. The texts were processed with XML Oxygen Editor in order to create digital editions of all of his known fragments. This approach allows for filtering the semantical information about persons, places, editorial techniques and characteristics. Also problems, questions and ideas of solutions were enumerated, everything in view of the coding of the text.

Der folgende Aufsatz ist eine Zusammenstellung von Problemen, Lösungswegen sowie Vor- und Nachteilen bei der Erstellung digitaler Editionen¹ – in diesem Fall einer Edition der Fragmente des römischen Historikers Fabius Pictor – und richtet sich in seiner Ausführung an Historikerinnen und Historiker mit Ambitionen, diese mit Extensible Markup Language (XML) unter Einbindung des kritischen Apparates in Angriff zu nehmen. Hier fehlt es nach wie vor an standardisierten Vorgehensweisen und vor allem an am konkreten Text orientierten Einführungen oder Beispielen für Altertumswissenschaftlerinnen und Altertumswissenschaftler. Ihnen sollen mit den nachfolgenden Beispielen praktische Hinweise für die eigene Arbeit gegeben werden.

Die hier beschriebene Digitaledition ist nicht als eigenständiger Typus der Edition antiker Texte zu betrachten, sondern vielmehr eine Entwicklung aus den vorangegangenen Printeditionen mit eigener Charakteristik.² Im vorliegenden Fall sind die Teilüberlieferungen des Fabius Pictor – ein Autor, dessen Wirken sich auf den Beginn des 2. Punischen Krieges beschränkt – als Basis benutzt worden. Sein Werk ist nur in Fragmenten (Zitaten, Erwähnungen, Bezügen) aus späteren Autoren erhalten. Bekannt ist, dass er bevorzugt auf Griechisch schrieb, um der kulturell einflussreichen Avantgarde Griechenlands die römische Sichtweise näher zu bringen.³

1 Dass die Auseinandersetzung mit der digitalen Edition an Aktualität nichts verloren hat, zeigen auch Beiträge in diesem Journal wie Meins (2016): S. 52–57 (<http://journals.ub.uni-heidelberg.de/index.php/dco/article/viewFile/24491/21869>), der sich mit der Dissertation von Sahle (2013): *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Norderstedt 2013 auseinandersetzt. Beide diskutieren in ihren Texten den Umgang mit der textkritischen Edition in der Altertumswissenschaft. Dieser Beitrag dagegen soll lediglich darauf aufmerksam machen, welche konkreten Schwierigkeiten auftreten (können), wenn man sich dazu entschließt, eine Edition mittels XML in die Digitalität zu übertragen.

2 Dazu Pierazza (2015), S. 74.

3 Timpe (1972), S. 928ff.

Die Idee „Digital Fabius Pictor“ wurde an der Universität Leipzig am Alexander von Humboldt Chair of Digital Humanities unter dem Dach des Projekts Digital Fragmenta Historicorum Graecorum (DFHG) erarbeitet und durchgeführt.⁴

Als Textgrundlage der digitalen Edition sind die entsprechenden Fragmente von Karl Müller in seiner Printedition *Fragmenta Historicorum Graecorum*, Bd. 3, verwendet worden, welche er 1849 veröffentlichte.⁵ Neuere Editionen⁶ sind aufgrund der Beschränkungen des Copyrights nicht verwendbar.⁷ Nachdem mit Hilfe der Auszeichnungssprache XML alle Informationen aus den Fragmenten hierarchisch strukturiert wurden,⁸ ist eine weitere Edition zum Vergleich erstellt worden, um unterschiedliche Herangehensweisen der Editoren stärker herauszuarbeiten, Differenzierungen vorzunehmen und so auch anderen Editoren Anregungen zu geben. Hierfür wurden die Fabius-Fragmente der Edition von Hermann Peter aus dessen Werk *Historicorum Romanorum reliquiae* in der 2. Auflage von 1914, das ebenfalls keinem Copyright mehr unterliegt, verwendet.⁹ Sein Werk ist gleichermaßen online einsehbar,¹⁰ so dass man die originale Printedition mit der von mir erstellten digitalen Edition vergleichen kann. Den Kern dieser Arbeit bilden die Fragmente des Fabius Pictor. Eine ernsthafte Auseinandersetzung mit den dabei entstandenen Schwierigkeiten kann nur durch einen Blick in die Originaldateien¹¹ erfolgen. An dieser Stelle ist auf ein Problem hinzuweisen, das sich aus der Anlage von Müllers Edition ergibt. So sind Textpassagen, die er als Fragmente definiert, in seinen Zusammenstellungen aus dem ursprünglichen Kontext gelöst und oft in einen neuen Kontext eingebettet. Insofern ist der Blick in diejenigen Editionen, aus denen er die Textpassagen entnommen hat, unerlässlich. Denn gerade die systemische Zuordnung der Fragmente (zu Autoren, Gruppen, Kontexten) wird heute z. T. anders gesehen. Für das hier beschriebene Vorgehen und die Methodik ist dies weniger relevant, muss jedoch für weitere Arbeiten im Blick behalten werden.

XML - Erstellen, Probleme, Lösungen

Bevor man sich der digitalen Edition widmen kann, sind nicht selten einige weitere Arbeitsschritte einzuplanen, die größtenteils als selbstverständlich vorausgesetzt werden, hier der Vollständigkeit halber jedoch aufgeführt werden. Die vorliegenden Werke in Printform müssen zunächst eingescannt werden, damit sie digital (PDF) vorliegen. Das weitere Vorgehen löst die Texte aus dem PDF-Dokument mit Hilfe von OCR (Optical Character Recognition), womit man in der Lage ist, diese digital zu bearbeiten und – wie in diesem Fall – im Oxygen

4 <http://www.dfhg-project.org/>. Diejenigen XML-Dateien, die im Folgenden behandelt werden, unter: https://github.com/kevinstrassburger/dfhg-dev/tree/master/Volume_3. Dort die Dateien (dfhg3FabPic_grc.xml, dfhg3FabPic_la.xml, Peter1FabPic.xml, Referenzliste.xml). An dieser Stelle sei Fr. Dr. Monica Berti und Fr. Simona Stoyanova ganz herzlich für ihre überaus wertvolle Unterstützung gedankt.

5 <https://archive.org/stream/fragmentahistori03mueluoft#page/80/mode/2up>

6 Beispielhaft: Die frühen römischen Historiker, hrsg., übers. und komm. von Hans Beck und Uwe Walther, Darmstadt 2. vollständig überarb. Aufl. 2005 oder *The Fragments of the Roman Historians*, ed. Tim J. Cornell, Oxford 2013.

7 Dies ist für Altertumswissenschaftlerinnen und Altertumswissenschaftler noch immer ein wunder Punkt. Dass das Arbeiten mit modernsten Editionen für die Digitalisierung nicht möglich ist, ist nahezu schändlich.

8 Das ist der Grundgedanke des DFHG-Projekts.

9 Besondere Aufmerksamkeit verlangt der Aspekt, dass Müller mit seinem Werk griechischsprachige Autoren sammeln wollte, während Peter römische Autoren zusammenzustellen plante. Fabius ist in beiden Sammlungen berücksichtigt. Die unterschiedliche Zielsetzung erklärt vermutlich auch die unterschiedliche Herangehensweise der beiden Editoren.

10 <https://archive.org/stream/historicorumroma01peteuoft#page/n395/mode/2up>

11 Vgl. Fußnote 4.

XML Editor zu nutzen, der als Werkzeugkasten bei der Bearbeitung des Textes zur Verfügung steht. Die Standards von EpiDoc TEI¹² und deren Guidelines, welche sich für antike Dokumente bereits etabliert haben, bilden die Grundlage. Zusätzlich mussten aber eigene Richtlinien erstellt werden,¹³ da nicht alle aufgetretenen Schwierigkeiten von EpiDoc abgedeckt wurden. Gerade neuere Projekte haben neue Fragen und Probleme aufgeworfen.¹⁴ Einige Unterschiede sind sehr einfach und leicht zu erkennen, wie die unterschiedlichen Vorgehensweisen von Müller und Peter in ihren jeweiligen Editionen belegen. Während Müller alle griechischen Fragmente in das Lateinische übersetzt, fehlt diese Art der Textbearbeitung bei Peter, während jener mit einem kritischen Apparat arbeitet, auf den wiederum Müller verzichtet.¹⁵ Hinzu kommt, dass Müller sich selbst unsicher in seiner Einschätzung ist, ob alle Fragmente zu Fabius Pictor gehören und so ordnet er zwei¹⁶ der 29 Fragmente dem Numerius Pictor zu. Eine Sichtweise, die schon bei Peter nicht mehr zur Anwendung kommt und der sich auch kein moderner Editor mehr anschließt. Auch die Länge der genutzten Fragmente unterscheidet sich teils erheblich. Dies ist – wie bereits festgestellt – durch die Entscheidung des Editors geprägt. Heute tendiert man zu längeren Fragmenten, um den Kontext zu wahren.¹⁷ Durch den kritischen Apparat von Peter ist es ihm ferner möglich, bestimmten Referenzstellen, die sich nicht klar Fabius zuordnen lassen, keinen eigenen Raum geben zu müssen. Beispielsweise hat Eusebius gerade nicht bei Pictor direkt kopiert, sondern seine Textpassage bei Dion. Hal. 1, 79 entnommen, der sie wiederum Pictor entlehnt hat. Peter kann dieses Vorgehen im kritischen Apparat dokumentieren. Auffällig ist auch, dass Peter in Summe dem Fabius mehr Fragmente zugeordnet hat.¹⁸ Die oben beschriebene Arbeitsweise erklärt, weshalb Müller auf 29 Fragmente kommt, während Peter 28 angibt und andere Passagen wiedergibt. Besonders häufig scheint Dionysios von Halikarnassos Pictor als Quelle benutzt zu haben. Dies rührt daher, dass er sich bemühte, Quellen zum römischen Gründungsmythos zu finden; und da Fabius Pictor als ältester römischer Annalist gilt, kann der Vertreter des Attizismus in seinem Werk *Ρωμαϊκή Αρχαιολογία* nicht auf ihn verzichten. Ebenfalls findet sich Livius auffällig häufig in der Fragmentsammlung. Sein Werk reicht ebenso von der Gründung Roms bis in seine Gegenwart, so dass sich für die Verwendung des Fabius Pictor gleich zwei Zeiträume eignen: die erwähnten Anfangsjahre der Stadt durch Romulus und Remus und der 2. Punische Krieg, in dem Pictor als Zeitgenosse wirkte. Die häufigste Verwendung des römischen Schriftstellers erfolgte aber tatsächlich immer dann, wenn ein anderer Autor versuchte, die Idee einer Darstellung von Roms Gründung/der Königszeit umzusetzen. 17 der von Müller genannten 29 Stellen beschäftigen sich mit der Gründung, den ersten Königen und dem Sabinerfeldzug. Die Bedeutung des Pictorischen Werkes wird unter den antiken Autoren nicht

12 Text Encoding Initiative; <http://www.tei-c.org/guidelines/>.

13 https://docs.google.com/document/d/1T6_2QXx1NVRE5iriQtYKtgkZhJTCq07_wjrWhoRpO1k/edit?pli=1#018.

14 Die Defizite und ihre Folgen benennt Damon (2016) in ihrem Fazit und kommt zu dem Schluss, dass gerade im kritischen Apparat Nachholbedarf im Hinblick auf die entsprechende Codierung besteht bzw. es überhaupt an einer einheitlichen Idee in der Vorgehensweise mangelt.

15 Das zeigt die Individualität, die jede Edition mit sich bringt. Probleme entstanden bei der Bearbeitung dadurch gerade nicht, da sowohl für die lateinische Übersetzung als auch für den kritischen Apparat Möglichkeiten der Codierung existieren. Bei Peter besteht der Apparat sogar aus zwei Teilen, einem lateinischen Kommentar und einem textkritischen Apparat.

16 Frg. 20, Cic. de div. 1, 21 und Frg. 21, Diod. 7, 3, 3.

17 Beispielsweise die Fragmente aus Polyb. 3, 8, Plut. Rom. 14, Dion Hal. 1, 79 bei Beck und Walther.

18 Peter hat Serv. auct. ad. Verg. Aen. 5, 73; Plut. Rom. 3 und Isid. Etym. 4, 7, 34 gefunden, De origin. gent. Rom. 20 weggelassen und Eus., Eutrop. 3, 8, Diod. 7, 5, 4 in den app. crit. verschoben. Beck und Walther verzichten auf die Erwähnung des Eusebius, ebenso auf Syncellus, der sich ebenfalls bei Diodor bediente. Hinzu kommt ein Fragment, Fab. Pic. F1; Quint. inst. 1, 6, 12; Non. p. 834L; Gell. 5, 4, 3 und Serv. auct. ad. Verg. Aen. 8, 630. Das bedeutet, dass es dringend geboten ist, auch die neueren Editionen der digitalen Arbeitsweise zugänglich zu machen, damit nicht nur die Methode auf dem neuesten Stand ist, sondern auch der Inhalt. Sonst ist vieles unter Umständen vergebens. Eine Übersicht findet sich in der Datei Referenzliste.xml.

angezweifelt, doch keineswegs immer alles widerspruchslös übernommen. Dionysios kritisiert wiederholt chronologische Ungenauigkeiten. Dies liegt nicht zuletzt daran, dass auch Pictor schon nicht mehr auf zuverlässige Quellen zurückgreifen konnte, während er, zumindest was den Informationsgehalt in den Passagen über Hannibals Feldzug angeht, als Zeitgenosse wohl als glaubwürdig kopiert wurde. Gerade Livius im 22. Buch übernimmt auch die Deutungen Pictors recht sorglos. Pictor wurde sowohl hinsichtlich der Frühzeit als auch für den Zeitraum der punischen Kriege vermutlich mangels Alternativen oft ausgeschrieben. Die Kritik jedoch wird nur an seiner Arbeit während der Königszeit geübt. Als Zeitzeuge der Kriege gegen die Karthager schien er dagegen große Autorität genossen zu haben.

Neben diesen inhaltlich-strukturellen Problemen können für den Ersteller antiker Editionen aus analogen Vorlagen auch Graphievarianzen, wie die unterschiedliche Verwendung der Buchstaben u und v Konsequenzen beispielsweise in der Wortsuche haben, die für jedes digitale Dokument zum Pflichtrepertoire gehört.¹⁹ Ebenfalls von Bedeutung sind Unicode Characters²⁰ und Worttrennungen am Zeilenende, da der Rechner – sofern man ihm dies nicht mittels Befehl mitteilt – diese Wörter nicht finden würde – ein Problem auf das man auch heute noch immer stößt. Das Wort *Ἀσδρούβου* besitzt z. B. für den Rechner nicht die gleiche Struktur wie das Wort *Ἀσδρού-βου*. Im vorliegenden Editionsprojekt ist der Einfachheit halber die Worttrennung schon innerhalb des XML-Dokuments beseitigt worden, so dass diese Frage im weiteren Verlauf unbeachtet bleiben kann, auch wenn damit ein Eingriff in das Dokument erfolgt. Problematisch an dieser Stelle – wir haben dafür noch keinen überzeugenden Lösungsansatz gefunden – ist der textkritische Apparat bei Peter. Ein direkter Eingriff in die Wortstruktur hätte einen Eingriff in die Satzstruktur und damit Zeilenlänge zur Folge, womit aber die Zugehörigkeit der Angaben des textkritischen Apparates unter Umständen verfälscht würden.²¹ Bevor es möglich ist, die annotierten Daten aus dem Text herauszuarbeiten, ist eine weitere Fehlerquelle auszuschalten. Der mittels OCR bearbeitete Text weist einige Fehler auf, die von Hand korrigiert werden müssen. Ähnlich aussehende Buchstaben werden oft verwechselt, Beispiele bilden die folgenden Paare: y/v, α/κ, u/v, b/h, nr/m, l/l, lc/k, g/y. Auffällig ist, dass besonders lateinische, weniger griechische Buchstaben von solchen Fehlern betroffen sind.

Nachdem die genannten Korrekturen im Text durchgeführt wurden, ist die eigentliche (historisch-) editorische Arbeit möglich. Dazu gehört, die Metadaten des Werkes in den **teiHeader** einzufügen, d.h. im Dokumentkopf die Daten über den Editor, den Übersetzer, den Herausgeber, das Datum und den Ort der Veröffentlichung sowie eventuell Lizenzen anzuzeigen.²² Der Titel des Textes – meist der Autorenname – ist von uns mit einem Verweis zum Perseus-Katalog²³ versehen worden, in dem Informationen über Werk und Autor zu finden sind.

Der Text beginnt dann innerhalb der Tags²⁴ **text** und **body**; sowohl bei Müller als auch bei Peter folgt in deren Edition zunächst eine Einleitung, die mit dem Tag **p** gekennzeichnet wird, welcher in diesem Projekt einen Text markiert, der von dem Editor stammt.²⁵ Dies dient zur

19 Dieses spielt zwar in der Bearbeitung des Textes mittels XML keine Rolle, darf aber am Ende bei der Programmierung einer Suchmaske nicht vergessen werden, da mittels einfacher Zeichensuche sonst keine Ergebnisse geliefert werden.

20 Zum Beispiel kann die Ligatur æ nicht im Dokument wiedergegeben werden, sondern wird vom Editor mit einem Zeichencode ersetzt, in diesem Fall “æ“ so dass aus quæ im Dokument quæ wird.

21 Wörter verschieben sich so beispielsweise von Zeile x in Zeile x+1, wohingegen im kritischen Apparat auf Zeile x verwiesen wird.

22 Veranschaulicht in dhfg3FabPic_grc.xml und Peter1FabPic.xml (vgl. Anm. 5), jeweils in den Zeilen 7–124.

23 <http://catalog.perseus.org/>.

24 Ein Tag ist eine Auszeichnung eines Datenbestandes mit weiteren Informationen. Es beginnt mit einem Anfangstag <tag> und endet mit einem Endtag </tag>.

25 Die Guidelines von TEI geben für p an: „p (paragraph) marks paragraphes in prose“.

Unterscheidung der Fragmente, die mit dem Tag **cit** gekennzeichnet wurden, also als Zitat antiker Autoren wie Plutarch u.a., die sich bei Fabius Pictor bedient haben. Der einleitende Abschnitt bei Müller enthält bereits zahlreiche verschiedene Informationen, die einen historisch-semanticen Sinn aufweisen. Dazu gehören Referenzstellen, Autoren sowohl antiker als auch moderner Natur, Namen, Datumsangaben oder Werktitel. Referenzstellen sowohl antiker als auch moderner Autoren werden mit dem Tag **bibl** ausgestattet. Um besser zu unterscheiden bzw. überhaupt filtern zu können, benötigt jede Referenzstelle ein Attribut, das es erlaubt, weitere Informationen hinzuzufügen. Im Fall des Livius erscheint folgende Konstellation:²⁶ `<bibl type="primary"><author>Livius</author></bibl>`. Das Attribut **type** wird mit dem Wert **primary** gekennzeichnet. Es weist dem Wort Livius also die Eigenschaft eines antiken Autors zu, was dem Leser in den meisten Fällen zwar bewusst ist, jedoch hier auch maschinenlesbar dargestellt werden muss. Das Attribut erlaubt damit, eine eventuelle Suche zu verfeinern, da nun nicht mehr alle Autoren angezeigt werden, sondern, wenn dies gewünscht ist, in antike und sekundäre Literatur unterteilt werden kann. Dementsprechend würde die Angabe eines modernen Autors den Wert **secondary** bekommen: `<bibl type="secondary"><author>A. Knause</author></bibl>`, wobei der Wert einfach bestätigt, dass Knause nicht in die Gruppe der antiken Autoren gehört, sondern eben zur modernen Forschungsliteratur. Natürlich benötigt man für Referenzstellen weitere Angaben. Eine vollständige Stelle des Livius wird – wir befinden uns noch immer in der Einleitung – wie folgt dargestellt:

```
<bibl type="primary"><author>Livius</author> <biblScope unit="volume">I</biblScope>,
<biblScope unit="chapter">44</biblScope></bibl>
```

Detailliert bedeutet das: **bibl** signalisiert, dass es sich hier um eine Referenzstelle handelt, das Attribut **type** mit dem Wert **primary** zeichnet die Stelle als eine aus einem antiken Werk entnommene aus, dann folgt Livius, der mithilfe des Tags **author** als Autor deutlich gemacht wird. Der Tag **biblScope** definiert den Umfang einer bibliographischen Referenz, ihm weist man unter Einbindung des Attributes **unit** unterschiedliche Werte zu. Für die erste angegebene Zahl „I“ wird der Wert **volume** ausgegeben, entsprechend der Nummer des Buches, die nächste Zahl, die „44“, erhält den Wert **chapter**, entspricht also dem Kapitel. In anderen Fällen können noch Seiten (Wert=**page**) oder Zeilenangaben (Wert=**line**) hinzugefügt werden. Moderne Autoren werden von Müller mit Werktiteln zitiert:

```
<bibl type="secondary"><author>A. Knause</author> <title>Vitae et Fragmenta vett. Historicorum Roman.</title> Berlin. 1833 <biblScope unit="page">38</biblScope></bibl>
```

Auch hier wird die Gesamtstelle mit **bibl** umfasst. Der Wert **secondary** weist sie als Stelle in einem sekundären Werk aus, es folgen der Name des Autors, der wiederum mit **author** gekennzeichnet wurde, und der Werktitel mit dem Tag **title**. Der Veröffentlichungsort und das Jahr erhielten noch keinen Tag, da zunächst anders priorisierte Informationen gekennzeichnet werden mussten.²⁷ Die Seitenangabe innerhalb des Werkes von Knause erhält wie schon bei Livius den Tag **biblScope** mit dem Attribut **unit** und dem Wert **page**, um die von Müller angegebene Stelle auch im Werk finden zu können.

Um auch Abkürzungen einbinden zu können, muss diese Abkürzung auch maschinenlesbar gemacht werden, ebenso wie die ausgeschriebene Form. Überhaupt ist der elementarste Ansatz, Aneinanderreihungen von Zeichen mit Hilfe von Tags einen Sinn zu verleihen. Im Falle der Abkürzungen bietet sich ein Blick auf die Situation der Namensschreibweise unter Müller an.

²⁶ Das Dokument von Müller gilt als Maßstab für die Analyse, Peter als Vergleichsdokument.

²⁷ Das ist unbedingt nachzuholen, können doch das Veröffentlichungsjahr bzw. die Auflage große Unterschiede bedeuten.

Zunächst ein Exkurs zur Codierung von Namen: Ein „gewöhnlicher“ römischer Name besteht zumeist aus drei Teilen – dem Praenomen, dem Gentilnomen und dem Cognomen. Quintus Fabius Maximus würde regulär wie folgt codiert:²⁸

```
<persName>
<name type="praenomen">Quintus</name>
<name type="gentilicium">Fabius</name>
<name type="cognomen">Maximus</name>
</persName>
```

Der gesamte Name wird von **persName** eingefasst, um zu verdeutlichen, dass es sich um eine Person handelt, dann wird jeder Name einzeln unter dem Tag **name** zusammengefasst, der wiederum mit dem Attribut **type** ausgestattet wird und dem ein Wert des jeweiligen Namens zugewiesen wird; das kann entweder **praenomen**, **gentilicium** oder **cognomen** sein.

Nun sind Namen bekanntlich nicht immer im Nominativ geschrieben; wie soll das Suchprogramm erkennen, dass Quinto Fabio Maximo dieselbe Person wie Quintus Fabius Maximus ist? Die Auszeichnung erfolgt deshalb nach diesem Schema:

```
<persName>
<name type="praenomen" nymRef="Quintus">Quinto</name>
<name type="gentilicium" nymRef="Fabius">Fabio</name>
<name type="cognomen" nymRef="Maximus">Maximo</name>
</persName>
```

Im Vergleich zum vorherigen Beispiel wurde jeder Name im Dativ um das Attribut **nymRef** erweitert und ihm als Wert der Nominativ des jeweiligen Namens gegeben. Das wurde für jeden Casus in den Fragmenten durchgeführt.

Für die Abkürzungspraxis ist folgende Auszeichnung gewählt worden:

```
<persName>
<name type="praenomen" nymRef="Quintus">Quinti</name>
<name type="gentilicium" nymRef="Fabius">Fabii</name>
<persName><expan><abbr>Q</abbr><ex>uinti</ex></expan>
<expan><abbr>F</abbr><ex>ilii</ex></expan></persName>
<persName><expan><abbr>Q</abbr><ex>uinti</ex></expan>
expan><abbr>N</abbr><ex>epotis</ex></expan></persName> <name>
type="cognomen">Pictoris</name> </persName>29
```

Dieses auf den ersten Blick gewiss verwirrende Schema folgt einem klaren Muster. Der ursprüngliche von Müller verfasste Name lautet: „Quinti Fabii Q. F. Q. N. Pictoris“, demnach Quinti Fabii Q(uinti) F(ilii) Q(uinti) N(epotis) Pictoris.³⁰ Wenn eine Abkürzung auftaucht, benötigt sie den Tag **expan** (Inhalt/Erweiterung einer Abkürzung) und den Tag **abbr** (Abbreviation=Abkürzung). Wenn man nun noch den fehlenden Teil ergänzen möchte, benötigt man weiterhin den Tag **ex**, um deutlich zu machen, dass dieser Teil im Original eben nicht vorkommt. **expan** schließt sowohl **abbr** als auch **ex** ein. Mit **abbr** selbst wird die Abkürzung als solche sichtbar gemacht, also jener Teil, der übrig geblieben ist.

²⁸ Die Codierung ist für den Fall, dass die römischen Namen in Griechisch niedergeschrieben wurden, die gleiche. Ursprünglich griechische Namen werden nur mit dem Tag **persName** ausgestattet, da sie keine gesonderte Einteilung benötigen (jedenfalls nicht in diesem Dokument).

²⁹ Vgl. `dfhgFabPic_grc.xml` Zeile 219.

³⁰ Theoretisch müsste auch hier der Nominativ addiert werden, was einen für dieses Beispiel zu großen Aufwand bedeutet hätte. Dies wird in nächster Zeit automatisch möglich sein.

Ortsangaben wie Stadtnamen o.ä. werden mit dem Tag **placeName** und dem Attribut **ref**³¹ (für Referenz) versehen. Der Wert des Attributes ist ein Link der Seite <http://pleiades.stoa.org/>, auf der sich antike Orte leicht sichtbar machen lassen. Beispielhaft dafür ist `<placeName ref="http://pleiades.stoa.org/places/423080">Τίβεριν</placeName>`: hier wird man per Link auf eine Karte gebracht, auf der die Lage des Flusses Tiber angegeben wird. Diese Verlinkung wurde mit allen Ortsangaben durchgeführt. Die Erläuterungen zu weiteren Tags sind den Guidelines von TEI³² zu entnehmen.

Nachdem in der Einleitung Müllers, die er zur näheren Betrachtung über Fabius Pictor schrieb, im Wesentlichen alle Informationen kenntlich gemacht wurden, soll nun ein Fragment exemplarisch erläutert werden, nämlich sein Fragment 11.³³

Das Originalfragment, sofern in Griechisch verfasst, erhielt von Müller eine lateinische Übersetzung, die in seiner Ausgabe unten auf der Seite angebracht ist, welche aber im Rahmen des Projektes in einem separaten Dokument untergebracht ist.³⁴

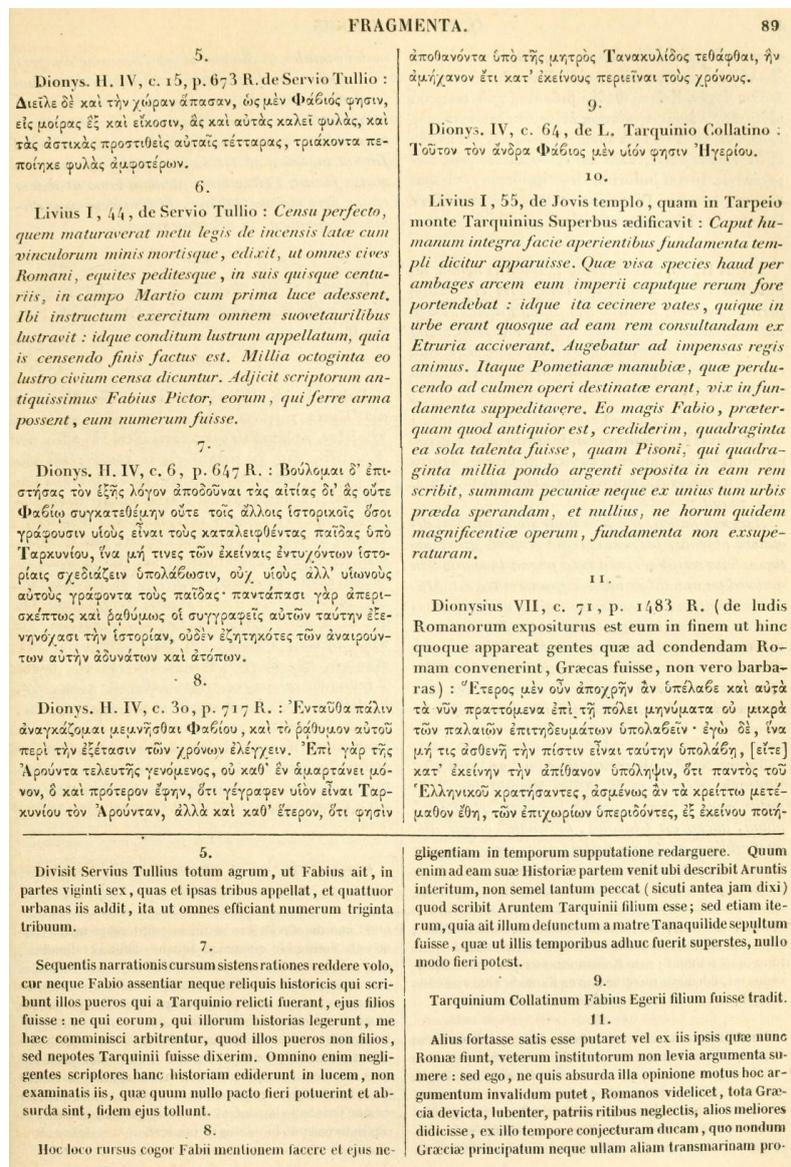


Abb. 1: Müller, S. 89.

31 **ref** wurde nur dann gesetzt, wenn auch ein Link zur Verfügung stand.

32 <http://www.tei-c.org/guidelines/>

33 <https://archive.org/stream/fragmentahistori03mueluoft#page/88/mode/2up>

34 dhfg3FabPic_la.xml

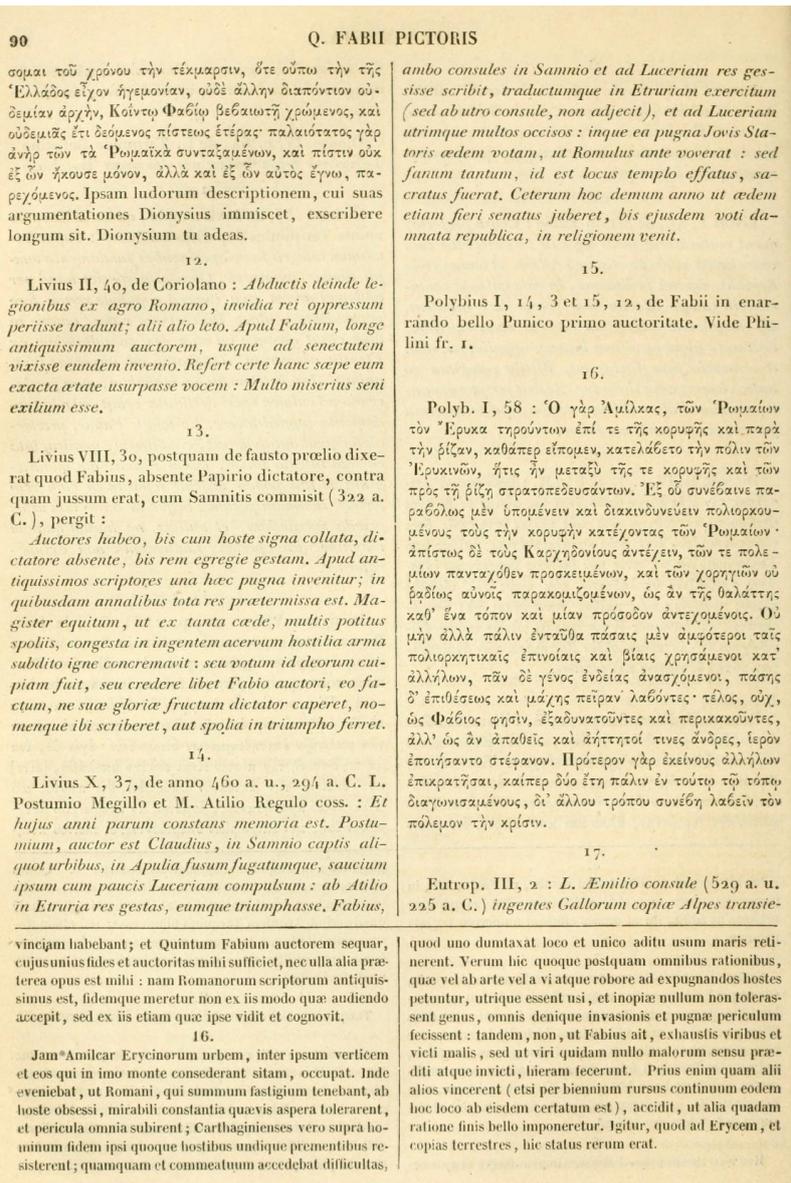


Abb. 2: Müller, S. 90

Jedes Fragment wird von dem Tag **cit** eingeschlossen, erhält eine Nummer und die Information, in welcher Sprache es geschrieben wurde, z.B.: `<cit n="11" xml:lang="grc">` Hier handelt es sich also um das nach Müllers Zählung elfte Fragment, verfasst wurde es in Griechisch – sichtbar durch das Attribut `xml:lang` und den dazugehörigen Wert „grc“.

Daraufhin folgt die Stellenangabe, die der in der Einleitung bereits besprochenen Referenzstelle gleicht:

`<bibl type="primary">`

`<author>Dionysius</author>`

`<biblScope unit="volume">VII</biblScope>`

`<biblScope unit="chapter">71</biblScope>`

`<biblScope unit="page">1483<ref><expan><abbr>R</abbr><ex>eiske</ex></ref></biblScope>`

`</bibl>`

Oftmals stellt Müller bereits eine kleine zusätzliche Information vor das eigentliche Zitat, in diesem Fall:

```
<note>(de ludis Romanorum expositurus est eum in finem ut hinc quoque appareat gentes qu&#230; ad condendam Romam convenerint, Gr&#230;cas fuisse, non vero barbaras)</note>.
```

Es folgt der eigentliche Text:

```
<quote>Ἐτερος μὲν οὖν ἀποχρῆν ἂν ὑπέλαβε καὶ αὐτὰ [...]³⁵</quote>
```

und unter Umständen schließt Müller mit einer weiteren Notiz das Fragment ab:

```
<note>Ipsam ludorum descriptionem, cui suas argumentationes <persName role="author"><-name>Dionysius</name></persName> immiscet, exscribere longum sit. <persName role="author"><name nymRef="Dionysius">Dionysium</name></persName> tu adeas.</note></cit>.
```

```
951 <cit n="11" xml:lang="grc">
952 <bibl type="primary"><author>Dionysius</author> <biblScope unit="volume">VII</biblScope> <biblScope unit="chapter">71</biblScope> <bib
953 Romanorum expositurus est eum in finem ut hinc
954 quoque appareat gentes qu&#230; ad condendam Romam
955 convenerint, Gr&#230;cas fuisse, non vero barbaras)</note>
956 <quote>Ἐτερος μὲν οὖν ἀποχρῆν ἂν ὑπέλαβε καὶ αὐτὰ
957 τὰ νῦν πραττόμενα ἐπὶ τῇ πόλει μηνύματα οὐ μικρὰ
958 τῶν παλαιῶν ἐπιτηδευμάτων ὑπολαβεῖν· ἐγὼ δέ, ἵνα
959 μὴ τις ἀσθενῆ τὴν πίστιν εἶναι ταύτην ὑπολάβῃ, [εἶτε]
960 κατ' ἐκείνην τὴν ἀπίθανον ὑπόληψιν, ὅτι παντὸς τοῦ
961 <placeName ref="http://pleiades.stoa.org/places/1001896">Ἑλληνικοῦ</placeName> κρατήσαντες, ἀσμένως ἂν τὰ κρείττω μετέμαθον
962 ἔθῃ, τῶν ἐπιχωρίων ὑπεριδόντες, ἐξ ἐκείνου ποιή-
963 <pb n="90"/>
964 σσαι τοῦ χρόνου τὴν τέκμαρσιν, ὅτε οὐπω τὴν τῆς
965 Ἑλλάδος εἶχον ἡγεμονίαν, οὐδὲ ἄλλην διαπόντιον οὐδεμίαν
966 ἀρχὴν, <persName role="author"><name type="praenomen" nymRef="Quintus">Κοίντω</name> <name type="gentilicium" nymRef="Fabius">Φα&#976;ῖω</i
967 οὐδεμιᾶς ἔτι δεόμενος πίστεως ἐτέρας· παλαιότατος γὰρ
968 ἀνὴρ τῶν τὰ Ρωμαϊκὰ συνταξαμένων, καὶ πίστιν οὐκ
969 ἐξ ᾧν ἤκουσε μόνον, ἀλλὰ καὶ ἐξ ᾧν αὐτὸς ἔγνω, παρεχόμενος.</quote>
970 <note>Ipsam ludorum descriptionem, cui suas
971 argumentationes <persName role="author"><name>Dionysius</name></persName> immiscet, exscribere
972 longum sit. <persName role="author"><name nymRef="Dionysius">Dionysium</name></persName> tu adeas.</note></cit>
973
```

Abb. 3: Müller Frg. 11 in XML (TEI)

35 Die eckigen Klammern dienen in diesem Beispielfall nur dazu anzuzeigen, dass der Text an dieser Stelle abgekürzt wurde, um Platz zu sparen. Die Klammer wird nicht nach dem Leidener Klammersystem verwendet, um eine Lücke im antiken Text anzuzeigen und wird deshalb zur Unterscheidung fett geschrieben.

Zum Vergleich die Darstellung des gleichen Fragments bei Peter unter Nr. 16:

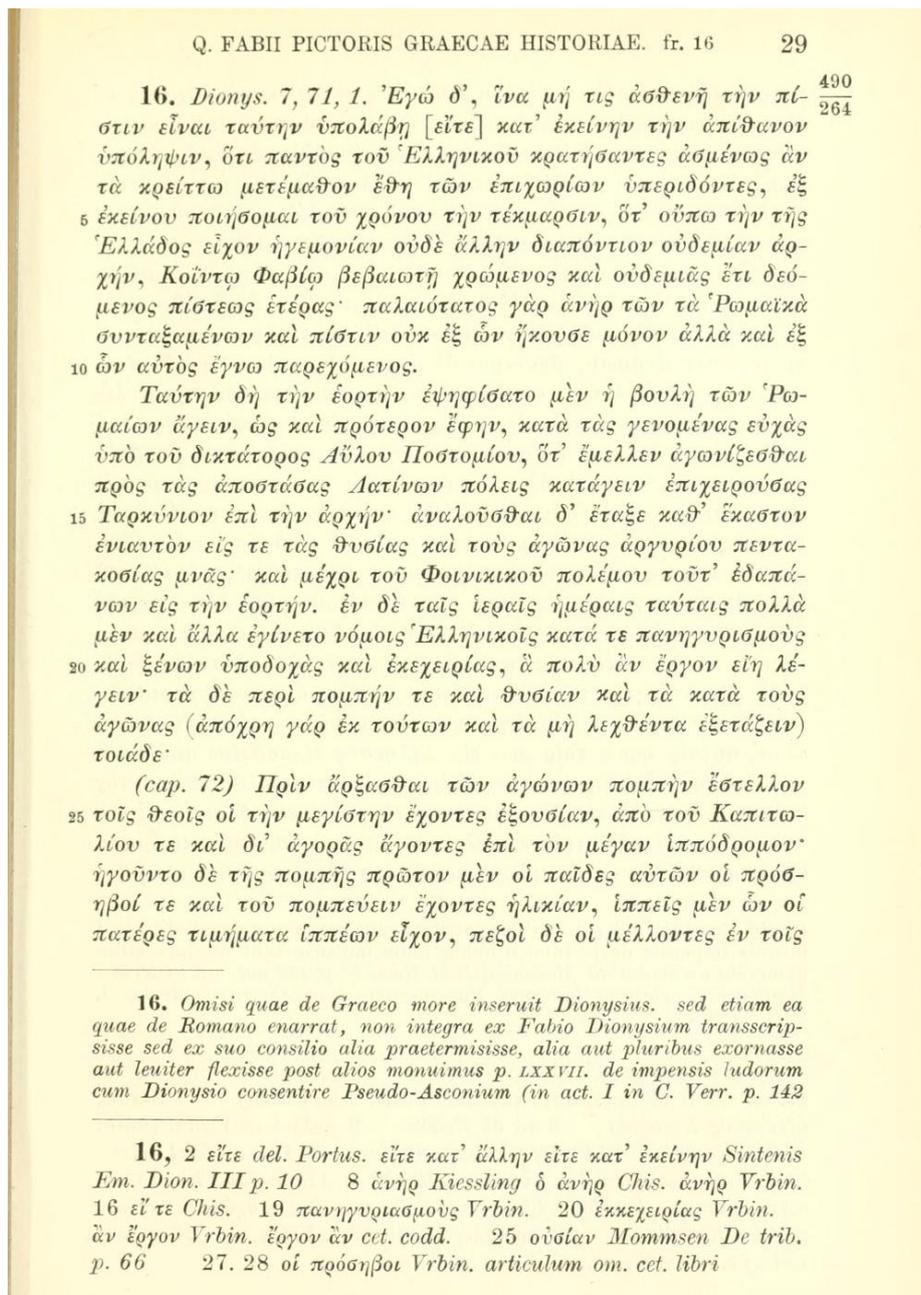


Abb. 4: Peter, Frg. 16, Seite 29

Der Anfang ist identisch: <cit n="16" xml:lang="grc"><bibl type="primary"><author>Dionys.</author> <biblScope unit="volume">7</biblScope>, <biblScope unit="chapter">71</biblScope>, <biblScope unit="line">1</biblScope></bibl>

Jedoch setzt Peter keine Informationen vor das folgende Fragment. Es folgt:

<quote>

<lb n="1">Ἐγὼ δ', ἵνα μή τις ἀσθενῆ τὴν πί-

<lb n="2">στιν εἶναι ταύτην ὑπολάβῃ [εἴτε] κατ' ἐκείνην τὴν ἀπίθανον[...]³⁶</quote>

36 Vgl. Fußnote 35.

Anders als bei Müller müssen die Zeilen separat als solche kenntlich gemacht werden, da Peter einen textkritischen Apparat nutzt und sich auf seine eigenen Zeilenangaben bezieht.³⁷ Zuerst lässt auch Peter eine Notiz folgen, in der Parallelstellen u.a. angegeben werden:

```
<listApp n="16" type="locip">
  <app>
    <note>Omisi quae de Graeco more inseruit Dionysius. sed etiam ea[...] </note></app></
listApp>
```

listApp macht deutlich, dass ein kritischer Apparat folgen wird. **app** ist der Beginn des kritischen Apparates, **n** ist die Nummer, auf die sich der Apparat bezieht und **type** mit dem Wert **locip** verdeutlicht, dass es sich hier um Angaben zu Referenzen handelt. Der eigentliche textkritische Apparat, in dem Peter verschiedene Lesarten angibt und Fehler korrigiert, benötigt größere Detailliertheit, da die Angaben innerhalb des Dokuments sonst nicht gelesen werden könnten:

```
<listApp n="16">
  <app loc="16.29.2">
    <note>εἶτε del. Portus. εἶτε κατ' ἄλλην εἶτε κατ' ἐκείνην Sintenis Em. Dion. III p. 10</note>
  </app>
  [...]
  <app loc="16.31.21">
    <note>Φερσεφόνης Chis. Vrbin.</note>
  </app>
  [...]
  <app loc="16.32.23">
    <note>τινὰς οἷς ἐτίμων libri corr. Reiske</note>
  </app></listApp>38
```

Mit **listApp** folgt die Andeutung eines kritischen Apparates, eine Nummerierung ist ebenso notwendig. Es schließt sich ein **app** an, das mittels Attribut **loc** eine genaue Platzierung im Werk erhält. Wir standen hier vor verschiedenen Problemen, denn zum einen gab es keine Beispiele für einen kritischen Apparat in Textdokumenten und zum anderen ging es um die Zuordnung. Wir entschieden uns für die oben genannte Variante, auch wenn dies bedeutete – anders als bei Müller – jede Zeile mit dem Tag **lb** auszustatten.³⁹ Zu sehen sind die Angaben 16.29.2, 16.31.21 und 16.32.23. Die erste Zahl ist die Nummer des Fragments in der Edition; da es sich um das gleiche Fragment handelt, besteht die erste Stelle immer aus der 16. Die 29, 31 und 32 an zweiter Stelle sind die Seitenangaben in Peters Edition, die ebenfalls im Dokument codiert wurden (<pb n="29"/> usw.), die letzte Zahl bezieht sich auf die Zeilenangaben in Peters Werk auf der zuvor genannten Seite. Im Dokument Peter1FabPic.xml nimmt dieses Fragment mit allen Notizen und dem kritischen Apparat die Zeilen von 2726–2983 ein. Hier ist ein Blick in das XML-Dokument sinnvoll, um die hierarchische Struktur nachzuvollziehen, die es ermöglicht, jede Information festzuhalten.

³⁷ Würde man die von ihm vorgenommen Worttrennungen korrigieren, könnte man seine Zeilenangaben verfälschen.

³⁸ Im 16. Fragment findet sich eine exemplarische Darstellung eines von Peter edierten Fragments, in dem auch die beiden Apparate zur Darstellung kommen.

³⁹ In Peter1FabPic.xml noch per Hand, demnächst sicher auch mittels Software möglich.

```

2824 <lb n="3"/>έσπενδον. οἶνον κατὰ τῶν ἀγνιζομένων. ἕκαστον δ' ὅτι κατὰ
2825 <lb n="4"/>νόμους ἐγένετο τοὺς ἀμφὶ θυσίαν ὑφ' Ἑλλήνων κατασταθέντας
2826 <lb n="5"/>ἐκ τῆς Ὀμήριου ποιήσεως γινῶναι ῥῥῆδον. - (cap. 73.) Λουτὸν δέ
2827 <lb n="6"/>ἔτι μοι καὶ περὶ τῶν ἀγῶνων οὐς μετὰ τὴν πομπὴν ἐπετέλουν
2828 <lb n="7"/>ὀλίγα διελαθῆν. πρῶτος ὁ τῶν τεθρίπων τε καὶ συναριδίων
2829 <lb n="8"/>καὶ τῶν ἀζευκτῶν ἵππων ἐγένετο δρόμος. - ἐν δέ ταῖς ἵππι-
2830 <lb n="9"/>καῖς ἀμίλλαις ἐπιτηδεύματα δύο τῶν πάνυ παλαιῶν ὡς ἐξ ἀρχῆς
2831 <lb n="10"/>ἐνομοθετήθη φυλαττόμενα ὑπὸ Ῥωμαίων μέχρι τῶν κατ' ἐμὲ
2832 <lb n="11"/>διάκειται χρόνων, τό τε περὶ τὰ τρίπῳλα τῶν ἀρμάτων. -
2833 <lb n="12"/>δυσὶ γὰρ ἵπποις ἐζευγμένοις, ὃν τρόπον ζευγνύται συναρίς,
2834 <lb n="13"/>τρίπτος παρέπιετο σερατός ἵππος ῥυτῆρι συνεχόμενος, ὃν ἀπὸ
2835 <lb n="14"/>τοῦ παρηρηθῆσθαι τε καὶ μὴ συνεζευχθαι παρηρον ἐκάλον οἱ
2836 <lb n="15"/>παλαιοί. ἕτερον δέ - ὁ τῶν παρεμβεβηκότων τοῖς ἀρμασι
2837 <lb n="16"/>δρόμος· ὅταν γὰρ τέλος αἱ τῶν ἵππέων ἀμίλλαι λάβωνται,
2838 <lb n="17"/>ἀποσηθῶντες ἀπὸ τῶν ἀρμάτων οἱ παροχούμενοι τοῖς ἡνίο-
2839 <lb n="18"/>χοις - τὸν σταδίατον ἀμιλλῶνται δρόμον αὐτοὶ πρὸς ἀλλή-
2840 <lb n="19"/>λους. τελεοέντων δέ τῶν ἵπτικῶν δρόμων οἱ τοῖς ἑαυτῶν
2841 <lb n="20"/>σώμασιν ἀγνιζόμενοι τότε εἰσήσαν δριμεῖς τε καὶ πύκται καὶ
2842 <lb n="21"/>παλαιστοί. - ἐν δέ τοῖς διὰ μέσου τῶν Ἑλληνικῶν χρόνοις
2843 <lb n="22"/>Ἑλληνικώτατον καὶ κράτιστον ἀπάντων ἔθῶν ἀπεδείκνυντο,
2844 <lb n="23"/>στεφανώσεις καὶ ἀναρρήσεις ποιούμενοι τιμῶν αἷς ἐτίμων τοὺς
2845 <lb n="24"/>ἑαυτῶν εὐεργλετας. - καὶ σκύλων ὄσων ἐκ πολέμων λάβοιεν
2846 <lb n="25"/>ἐπίδειξις τοῖς εἰς θέαν συνεληλυθόσιν. ἀλλὰ γὰρ ὑπὲρ μὲν
2847 <lb n="26"/>τούτων οὔτε μὴθένα ποιήσασθαι λόγον ἀπαιτούσης τῆς ὑπο-
2848 <lb n="27"/>θέσεως καλῶς εἶχεν οὔτε μηκύνειν πέρα τοῦ δέοντος ἤρμοττε.</quote></cit>
2849
2850 <listApp n="16" type="locip">
2851 <app>
2852 <note>Omisi quae de Graeco more inseruit Dionysius. sed etiam ea
2853 quae de Romano enarrat, non integra ex Fabio Dionysium transscripsisse
2854 sed ex suo consilio alia praetermisisse, alia aut pluribus exornasse
2855 aut leuiter flexisse post alios monuimus p. lxxvii. de impensis hidorum
2856 cum Dionysio consentire Pseudo-Asconium (in act. I in C. Verr. p. 142
2857 Orell.) Romani ludi sub regibus instituti sunt magnique appellati, quod
2858 magnis impensis dati. tunc primum ludis impensa sunt ducenta milia
2859 nummum animaduertit Mommsen Hist. Rom. I2 p. 458 not. ad uictoriam
2860 Regillensem originem ludorum Romanorum etiam Cicero de diuin. 1,
2861 26, 55 rettulit.</note>
2862 </app>
2863 </listApp>
2864
2865 <listApp n="16">
2866 <app loc="16.29.2">
2867 <note>εἶτε del. Portus. εἶτε κατ' ἄλλην εἶτε κατ' ἐκείνην Sintenis Em. Dion. III p. 10</note>
2868 </app>
2869 <app loc="16.29.8">
2870 <note>8 ἀνήρ Kiessung ὁ ἀνήρ Chis. ἀνήρ Vrbin.</note>
2871 </app>
2872 <app loc="16.29.16">
2873 <note>εἶτε Chis</note>
2874 </app>
2875 <app loc="16.29.19">
2876 <note>πανηγυρισμοὺς Vrbin.</note>
2877 </app>
2878 <app loc="16.29.20">
2879 <note>ἐκκεχειρίας Vrbin. ἂν ἔργον Vrbin. ἔργον ἂν cet. codd.</note>
2880 </app>

```

Abb. 5: Peter Frg. 16 in XML (Auszug)

Das Potential der XML-Methode ist noch längst nicht ausgereizt. Es ist theoretisch möglich, jedes Wort, jeden Buchstaben bzw. überhaupt jedes Zeichen mit einer Bedeutung zu versehen. Dies war hier jedoch nicht das Ziel, sondern, da die Möglichkeiten vielfältig sind, und so gut wie nie identisch übertragbar, sollte hier lediglich anhand des Beispiels von Fragmentausgaben gezeigt werden, wie die Printeditionen digital übertragen werden können.

Eine besondere Herausforderung war natürlich, überhaupt zu erkennen und zu unterscheiden, welche Informationen existieren und welche davon für das historische Arbeiten Priorität haben. Personen, Schauplätze, Referenz- und Parallelstellen anderer Autoren liegen nahe, doch auch Metadaten sind ein wichtiger Bestandteil von Texten, besonders wenn Editoren sich – wie im kritischen Apparat von Peter – auf ihre eigenen Versionen stützen. Weiterhin sind Personen natürlich nicht von gleicher Art, es gibt fiktive Charaktere, Autoren oder Akteure. Ist der göttliche Olymp eine Ortsangabe oder eine mythische Metapher? Diese Entscheidungen fallen

hier in die Verantwortung des Editors und machen auch eine digitale Edition abhängig vom „Humaneditor“. Das damit verbundene Protokollieren der Entscheidungen allerdings bietet – neben einer guten Nachvollziehbarkeit – die Möglichkeit, divergierende Meinungen schnell und einfach in XML-Dokumenten zu platzieren und einzufügen, von sonstiger Fehlerkorrektur ganz zu schweigen, die ohne Zeitverzögerung stattfinden kann. Ein großer Vorteil der digitalen Edition ist es auch, dass sie jederzeit leicht durch eine Kommentarfunktion mit Hinweisen ausgestattet werden kann, die für jedermann sichtbar sind und im Dokument die Vorgehensweise näher erläutern, damit der Leser in die Gedankengänge des Editors eingebunden werden kann, was bisher aus Kostengründen im normalen Druckverfahren nur minimal möglich war. Eine der größten wissenschaftlichen Prämissen ist die Argumentation und deren Nachvollziehbarkeit bzw. die Auseinandersetzung mit dieser. Hier ist die digitale Edition von unvergleichlichem Vorteil. Alles steht protokollartig zur Verfügung und kann kritischen sowohl ablehnenden wie auch unterstützenden Stimmen eine Grundlage geben. Das bereits angesprochene Potenzial erschöpft sich derzeit noch in der für den menschlichen Editor oftmals ermüdenden Arbeitsweise. Durch die fortschreitende Entwicklung und Automatisierung wird sich jedoch ein erheblicher Teil der Arbeit von einer Software erledigen lassen, die in der Lage ist, Metadaten zu erkennen und diese mit den entsprechenden Tags auszustatten, so dass Historikerinnen und Historiker sich im Wesentlichen auf die wissenschaftliche Arbeit konzentrieren können: das Interpretieren des Textinhalts mit Hilfe von XML, so dass noch mehr Informationen leichter zugänglich gemacht werden können – auch wenn durch maschinelle Bearbeitung sicher neue Fragen zu klären sein werden.

Fazit

Das Übertragen einer Printedition auf eine digitale Ebene weist zahlreiche Möglichkeiten auf. Diese können sich von dem Versuch, die Intention des ursprünglichen Editors so genau wie möglich wiederzugeben, bis zur völligen Verselbständigung bzw. zur Realisierung einer eigenen Idee erstrecken. Mit der Möglichkeit quasi *alle* Informationen auf engstem Raum zur Verfügung zu stellen und diese transparent sowie nachvollziehbar zu gestalten, ist eine der wichtigsten Anforderungen an die Edition eines antiken Textes erfüllt. Dies gilt sowohl für die Edition einer mittelalterlichen Handschrift, eines Papyrus oder einer Inschrift wie auch die Überführung einer Printedition in die Digitalität.

Das hier präsentierte Vorgehen hat anhand zweier Printeditionen mit sehr unterschiedlichen editorischen Konzepten (Müller und Peter) exemplarisch für ein Fragment des Fabius Pictor erläutert, nach welchen editorischen Prinzipien eine digitale Edition mit der Extensible Markup Language (XML) mit und ohne Einbindung des kritischen Apparates erarbeitet werden kann. Denkbar ist natürlich die Erstellung einer eigenen, unabhängigen Edition mit Hilfe von XML und der Entwicklung eigener Standards und einer besonderen Vorgehensweise.⁴⁰ Ebenso ist eine Verlagerung des kritischen Apparates in den Text vorstellbar. Beispielsweise ist es denkbar, weitere Anzeigefenster zu öffnen, die die entsprechenden Informationen der annotierten Wörter/Wortgruppen enthalten wie Referenzstellen, Koordinaten für Ortsangaben, andere Überlieferungsversionen. Im Mittelpunkt des Beitrages stand jedoch weniger die Metadiskussion im Bereich Edition und kritischer Apparat, sondern es wurden die Probleme und Fragen angesprochen, die sich ergeben, wenn man mit Hilfe von XML die Überführung einer Printedition in die digitale Edition durchführen möchte. Diese reichten von verschiedenen Zeichen, die die gleiche semantische Bedeutung hatten, bis hin zu Folgefehlern, die sich aus der

⁴⁰ Beispielsweise eine andere Hierarchisierung der Informationen vorzunehmen oder individuelle Schwerpunkte zu setzen.

Vereinheitlichung des Textes ergaben und nach wie vor ergeben. Nicht für alles konnte hier ad hoc eine Lösung gefunden werden, aber es wurde deutlich, wie detailliert die Arbeit unter Umständen sein kann und muss – auf ganz unterschiedlichen Ebenen. Vom einzelnen Zeichen, über einzelne Wörter, zu Wortgruppen oder Sätzen, bis hin zu ganzen Abschnitten und den Metadaten können Informationen sachgerecht übermittelt und zur Verfügung gestellt werden. Viele der Probleme sind bekannt, ohne dass dafür bis heute eine Lösung gefunden worden wäre.⁴¹ Dass die Diskussionen darüber schon länger andauern, lässt erahnen, wie komplex das Thema und die dazugehörige Diskussion⁴² noch immer sind.

41 Beispielsweise dürfen sich annotierte Elemente nicht überlappen, dazu Witt (2004), S. 42 und Schmidt (2012), S. 129. Gerade für das Overlap-Problem könnte sich CTS (Canonical Text Services) als hilfreich erweisen, vgl. Blackwell, Schubert (2016), S. 94f.

42 Jannidis (2017), S. 248 zeigt einen Überblick über die noch immer aktuellen Fragen in Bereich Digitale Editionen.

Literaturverzeichnis

Quellen

Fragmenta historicorum Graecorum, Bd. 3, ed. Karl Müller, Paris 1849.

Historicorum romanorum reliquiae, Bd. 1, ed. Hermann Peter, Leipzig 2¹⁹¹⁴.

Die frühen römischen Historiker, hrsg., übers. und komm. von Hans Beck und Uwe Walther, Darmstadt 2. vollständig überarb. Aufl. 2005.

The Fragments of the Roman Historians, ed. Tim J. Cornell, Oxford 2013.

Literatur

Blackwell, Schubert (2012): C. W. Blackwell und Ch. Schubert, “Annotating and Editing with Canonical Text Services (CTS), Project founded by the Andrew W. Mellon Foundation: 2016–2017”, DCO 2.1, S. 94–99.

Damon (2016): C. Damon, “Beyond Variants: Some Digital Desiderata for the Critical Apparatus of Ancient Greek and Latin Texts”, in: M. J. Driscoll (Hg.), Digital Scholar Editing: Theories and Practices, Cambridge.

Meins (2016): F. Meins, „Digitale Editionen in den Altertumswissenschaften: Eine Auseinandersetzung mit P. Sahles Kritik kritischer Texteditionen“, DCO 2.1, S. 52–57.

Jannidis / Kohle / Rehhbein (2017): F. Jannidis, H. Kohle, M. Rehbein, Digital Humanities, Eine Einführung, Stuttgart.

Pierazza (2015): E. Pierazza, Digital Scholar Editing: Theories, Models and Methods, Farnham.

Sahle (2013): P. Sahle, Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels, Norderstedt.

Schmidt (2012): D. Schmidt, The Role of Markup in the Digital Humanities, Historische Sozialforschung 37.3, S. 125–146.

Timpe (1972): D. Timpe, Fabius Pictor und die Anfänge der römischen Historiographie, in: H. Temporini (Hg.), ANRW I.2, Berlin–New York, S. 928–969.

Witt (2004): A. Witt, Linguistische Informationsmodellierung mit XML, in: A. Mehler und H. Lobin (Hgg.), Automatische Textanalyse: Systeme und Methoden zur Annotation und Analyse natürlichsprachlicher Texte, Wiesbaden S. 39–54.

Links

DFHG-Projekt:

<http://www.dfhg-project.org/> (17.07.2018).

XML-Dateien:

https://github.com/kevinstrassburger/dfhg-dev/tree/master/Volume_3 (17.07.2018).

DFHG Guidelines:

https://docs.google.com/document/d/1T6_2QXx1NVRE5irIQtYKtgkZhJTcq07_wjrWhoRpO1k/edit?pli=1# (17.07.2018).

Edition H. Peter:

<https://archive.org/details/historicorumroma01peteuoft> (17.07.2018)

Edition K. Müller:

<https://archive.org/details/fragmentahistori03mueluoft> (17.07.2018)

TEI-Guidelines:

<http://www.tei-c.org/index.xml> (17.07.2018).

Autorenkontakt⁴³

Kevin Straßburger, M.A.

Universität Leipzig
Historisches Seminar
Lehrstuhl für Alte Geschichte
Beethovenstr. 15
04107 Leipzig

Email: kevin.strassburger@uni-leipzig.de

⁴³ Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren. Alle Inhalte dieses Beitrages unterstehen, soweit nicht anders gekennzeichnet, der Lizenz CC BY 4.0.

Romans 1by1: Overview of a Research Project

Rada Varga, Annamária-Izabella Pázsint, Imola Boda, Dan Augustin Deac

Abstract: The current article presents the main scientific results extracted from the [Romans 1by1](#) platform. While the database has been technically documented before, the analyses, based on it, is introduced here for the first time. After gathering and structuring all the prosopographical information on all the people, attested epigraphically in Roman Moesia Inferior, Moesia Superior and Dacia, we are able now to present exhaustive statistics and a comprehensive overview, as well as to get relevant conclusions regarding the epigraphic habits of each province.

The database and its purpose

The paper is based on the results of a research project, which was carried out between October 2015 and September 2017 at the Babeş-Bolyai University from Cluj-Napoca (Romania). The goal of the project was to create a database for ancient population, in which all the known individuals dating from the Roman period (exhaustively up to the end of the 3rd c. AD), from the provinces of Dacia and Moesia are recorded. The project aimed at employing new techniques and methodologies that come from other fields (i.e. computer science), in order to approach the study of ancient population in an innovative way, to ease the research, and to create an open access tool, available for the academic community.

Being a research database based on a continuous work in progress, we have opted for two versions of the database: a) one which has a restricted access (login via personalized accounts), and which is used only by the research team, due to the fact, that it is continuously updated and reviewed; and b) the other one which is open access and represents the cleansed, standardized and user-friendly version of the first one.¹

The database has been documented before,² thus we will only schematically present its architecture and the metadata. The purpose of this article is to present the macro-analyses, resulted from the gathered data – as for the first time all the data, on all the people of these provinces, are structured and available in one place.

No doubt, one of the most important advantages, in this case, is the creation of a prosopographical *corpus*. The *corpus* can be freely used for scientific purposes by anyone interested in the field, as well as for educational purposes or for disseminating aspects related to ancient history in a wider sector of the public, in an accessible and friendly manner.

1 <http://romans1by1.com/rpeople/people>.

2 [Varga \(2017\) 333–341](#) and far more detailed, [Varga \(2017b\) 44–59](#), published in Digital Classics Online, in 2017.

Another advantage is given by the fact that the database was built in order to be very user friendly and adaptable to the research question, which anyone has: more precisely we speak about the search filters in each main category which allow the user to request a specific set of information from the database. In this manner, at a click distance, we can find out, for example, who were all members of private associations, or who were the attested governors of the provinces. Moreover, all the information of our choice can be easily downloaded in an excel file and scientifically processed later.

Database architecture

Romans 1by1 is a relational database, built in MySQL and following the best practice models for population databases.³ The metadata are structured into four major tables (*Inscriptions*, *Bibliography*, *Personal Data* and *Personal Relationships* – summing up to more than 100 separate attributes).⁴

The first table, to be filled in, is the file of the source – *Inscriptions*. To begin with, each inscription gets an identification code, formed of 5 digits and an acronym of the province's name (D for Dacia, MS for Moesia Superior, MI for Moesia Inferior) – so we have, for example, 00001MS. In this section certain fields are linked to other existing resources, in order to obtain maximum efficiency for the users. The texts of the inscriptions are linked to the EDH,⁵ Clauss-Slaby⁶ and/or Packhum⁷ databases, while the places have been referenced to Pleiades⁸ or Trismegistos.⁹ The *Bibliography* section has been conceived in a way so that the extracting of complete or selective bibliographical lists would be possible. Thus, a normalization table includes all bibliographical titles referred to and being quoted; with the help of a value list, one can choose one *Bibliography abbreviation* for which the full reference, detailed information and comments are then displayed. Of course, all data are linked to the *Inscription code*, selected as well from a value list.

The core of the database is a table, used for recording data about individual persons (labelled *Personal data*), the entire network of relations is built around. Each new entry represents a singular epigraphic attestation of an individual and a unique ID is generated, which will help linking the character within the various components of the database and with other database entries. The person is also manually linked to the source using a value list of the inscriptions' codes. In the case of one person being attested by multiple epigraphic sources, each attestation will be represented by a new entry, to which a new unique ID will be assigned and which will be doubled during linkage procedures by a common ID for all instances of the same person.

Expectedly, this metadata is the most volatile one, being in a continuous process of modification and enlargement. As a matter of principle, we are trying to remain faithful to the source and to record during the first phase only the minimum of deduced information. The first section of fields¹⁰ from this table includes personal identification information about the person:

3 Mandemakers, Dillon (2004).

4 Varga (2017b) 45, for a schema of the metadata.

5 <http://edh-www.adw.uni-heidelberg.de/home>.

6 <http://www.manfredclaus.de/>.

7 <https://epigraphy.packhum.org/>.

8 <https://pleiades.stoa.org/>.

9 <http://www.trismegistos.org/>.

10 All of the terms match the field labels in the database.

Praenomen, Nomen, Cognomen/Personal name, Father/Master name, Agnomen, Signum, Nazione, Ethnicity, Origo, Domus, Local citizenship, Gender, Juridical status, Occupation, Deities, Age (at death), Details of life/death and Observations. For *Collegium/Association* and *Religious position* we decided to use check boxes, which open a series of fields. For the associative forms, we have three free text boxes: *Type of association, Position within the association, Activities within the association*, which allow certain flexibility and the possibility to accommodate information and realities from both Roman and Greek inscriptions. In the case of *Religious position*, the situation is a bit more complex: when checking this option, a drop down list of *Sacerdotal office* opens whose values are (at the moment) *Augur, Flamen* and *Pontifex* and a series of further check-boxes: *Coloniae / Municipii sacerdos, Military sacerdos, Imperial priest, Divinity priest, Laurentium Lavinatium*. All of them, with the exception of the last one, open a *Details* text field for description.¹¹

The second half of the *Personal information* table is dedicated to political and social status. This part of the metadata consists of a series of check boxes, each opening different categories of specialized attributes: *Ordo senatorius, Ordo equester, Provincial Governor, Procurator, Local magistrate, Decurionate*. The next information items cover *Imperial slave, Imperial freedman* and the *Military personnel*, all equipped with *Details* attributes.

Based on the personal ID, given to each individual, the *Relationship* table will solely name the relationship between individuals (A to B and B to A),¹² choosing from a drop-down menu. The relationship values have been encoded and we have tried to adjust the concepts to the Snap-Drgn ontology.¹³

Very important for us was the search interface of the database. Built with Ruby-on-rails, it was designed to respond to the most manifold and complex search options. Every component of the database has search filters for every particular field, as well as a general search. For administrators and data entry operators there is also an *Author* search field but it is not accessible for external guest users.

The limits of the research

Unavoidably, when working with such large data, which can be very complex and at the same time scarce, the research is faced with some difficulties, or with some anomalies. In our case the difficulties were connected to the nature of the data: more precisely, due to the historical inability of identifying, ‘isolating’, and linking the characters from the inscriptions, the database certainly contains cases of overlapped individuals.

The project did not have as a goal to focus on the onomastic of the individuals, as we tried to remain as faithful as possible to the written source and to avoid registering deductive information. Another aspect, which could not be dealt with in a more constructive manner, is that of the juridical status of the individuals. Once again, due to the quality of the sources, the juridical status of many individuals rests unknown or unclear, as in many cases it cannot be established. Regarding citizenship, a particular challenge has risen: how can one distinguish

¹¹ Varga (2017b) 48–49.

¹² Varga (2017b) 50.

¹³ <https://snapdrgn.net/ontology>.

between Roman and local citizenship in the database? Our choice was the following: we ticked the box "citizen" only in those cases when the Roman citizenship of the individual is clearly mentioned. The most telling case is that of the individuals from the Greek cities of the (future) province of Moesia Inferior, where the local citizenship applies in parallel with the Roman citizenship until Caracalla's edict of 212 AD (*Constitutio Antoniniana*). One telling parallel comes from the most documented province of them all, Egypt, where local citizenship was in some cases more important than the Roman one.¹⁴

Statistics on Dacia

The first obvious type of statistics involves the sheer numbers of sources and attested people. Thus, in Dacia there are 4.354 attested **individuals**, coming from 2.645 **inscriptions**. The **types of inscriptions** are diverse: 1.226 votive monuments, 758 epitaphs, 387 *instrumenta*, 117 honorific stones, 66 construction inscriptions (in various forms: dedications, plaques, even benches), 62 military diplomas, 19 wax tablets, 10 lists (*alba*) and 2 unknown monuments (the stones are so badly deteriorated that one cannot establish their original purpose). It is very interesting to note the proportion between funerary and votive monuments – besides hazard, the percentages show an increased popularity of the votive dedications.¹⁵ The rich urban centres of Dacia Superior mostly generate the situation and it reveals an urban culture of almost ostentatious public display, very interested in the meta-message of a monument, public opinion and indirect self-promotion.

The **language** of the inscriptions is overwhelmingly Latin – only 37 epigraphs from Dacia are written in Greek (mainly votive, but also funerary and *instrumenta* – especially on *amphorae*) and 8 are bilingual: 5 written in Greek and Latin (votive and *instrumenta*) and 3 in Palmyrean and Latin (all funerary, all coming from Tibiscum).¹⁶

Regarding the distribution on the three administrative circumscriptions of Dacia, there are 234 persons in Dacia Inferior (DI in the database), 3.031 in Dacia Superior (DS in the database) and 1.039 in Dacia Porolissensis (DP in the database); the rest of 43 persons come from inscriptions (mainly diplomas) from the Trajanic period, when the province was not yet administratively divided. The predominance of the data from Dacia Superior is completely understandable, as here the two greatest urban centres of the province are located: Sarmizegetusa, *colonia* from the founding of the province, capital and seat of the financial procurator and the conurbation from Apulum, where the Legio XIII Gemina¹⁷ was stationed and the provincial governor resided from 168 AD on. The north is represented by the highly militarized area of the *limes* (Porolissum being the most important centre from the military area) and by Potaissa, an urban settlement which grew larger and richer after the Legio V Macedonica was stationed here, in 168 AD.¹⁸

14 Alston (2003) 61–62.

15 The phenomenon appears as constant in time and the impossibility to narrowly date most of the private monuments doesn't allow us to undertake diachronic analyses on these data.

16 On the language of the inscriptions from Dacia, see Beu-Dachin (2014) and Piso *et. al.* (2016).

17 Piso (1993) 37–38, 82–93. For the seat of the governor see: Varga, Rusu-Bolindeț (2016) 115–126.

18 Bărbulescu (1987); Bărbulescu (1994); Piso (2013b); Nemeti u. Nemeti (2014).

Among the attested characters, there are 682 women; 122 come from Sarmizegetusa and 152 from Apulum, while from the whole northern part of Dacia there are only 164 women. These figures prove the openness of the better-urbanized, civilian area of central Dacia, while the martial character of the northern parts of the province is highly obvious. The *Unknown gender* category was used for badly deteriorated inscriptions, their state makes it impossible to assume if we are dealing with a man or a woman. The team considered it as necessary to include these inscriptions because they can provide information such as age, home areas, relationships, etc.

Getting to the **juridical status**, the most under-represented category is that of the slaves. In the whole province, there are only 76 certain non-imperial slaves (private slaves and *servi publici*) and 35 imperial slaves (two thirds of them coming from Sarmizegetusa). Regarding the first group, it is important to note that roughly half of them (33 individuals) have recorded their professional status and mostly operated in administrative fields: *actor*, *vilicus*, *adiutor tabularii*, etc.¹⁹ The only female slave, registered with the professional status, is also part of the administrative personnel, being a *ministra* from Potaissa.²⁰ The *liberti* are present in Dacia in a number of 176 individuals, most of them associated with the *ordo Augustalium*. The peregrines, attested in Dacia²¹ before the Constitutio Antoniniana, are generally under represented, but have a better ratio than the slaves and freedmen: there is attested a total of 277 certain non-citizens men and women.

While the juridical status of the registered persons is more or less identifiable, the social status is more difficult to quantify, as with this matter we emerge into the discussion concerning where we draw the line between the upper classes and the rest of Roman society. As well, we get to the ‘controversy’ of the provincial middle classes, their definition, social and economic boundaries – and even their mere nomenclature and the justification of using the anachronistic expression of ‘middle class’. As these controversies are not the object of our current project, we will only look at the undisputable **elite categories**: decurions and local magistrates, procurators, *equites* and senators. Thus, in Dacia there are attested 245 decurions, among whom 107 also fulfil magistratures (*duumvir*, *aedilis*, etc.). Only 15 of them have multiple decurional functions, 8 of them are *decuriones* of both Sarmizegetusa and Apulum.

As well, there are 56 *procuratores* recorded, among whom 26 are clearly financial procurators and 18, *praesidial*.²² 158 Roman knights have been identified, while the representatives of the *ordo senatorius* only list 75 individuals (among which, 50 have been provincial governors). The rest of them are *legati legionis* or *decuriones militum legionis*. The only attested woman of senatorial rank is Pullaiena Caeliana, wife of a governor (Quintus Marcius Victor Felix Maximilianus, *legatus* of Dacia between 198–209 AD), known from a family dedication in Apulum.²³

Besides these elite categories, other social particular groups are the **military** personnel and the veterans. In Dacia, there are 863 active military men, with great geographic preponderance in Apulum (292 individuals), followed by Potaissa (143 men) – expectedly, as here the seats

19 Mihăilescu-Bîrliba (2004).

20 CIL III 907 = CIL III 7693 (<http://www.romans1by1.com/rpeople/7936>). Unlike in EDH, the name of the first named person is female and thus *filia* refers to that person and not to the later named Rufina, which is denoted *ministra* (inscription *menesteriis*). For the reading/interpretation of the occupation as *ministra*, see Gorea (2013), with bibliography.

21 Varga (2014) for an overview of the peregrines.

22 For their *fasti* see Piso (2013a).

23 CIL III 1118 (<http://www.romans1by1.com/rpeople/9852>).

of the province's legions were located. Surprisingly enough, we only have 180 veterans – a situation that makes it difficult to acknowledge whether these individuals left Dacia after their discharge, or simply chose not to mention their status of veteran on the monuments they erected.²⁴

158 individuals are listed as members of various **collegia and associations**, excluding the *Augustales*. Worth mentioning is the fact that they all come from a total of 59 inscriptions – generally collective votive dedications. In this context, we also have to mention the variety of *collegia* attested in Dacia, as there are registered 17 different associative organizations. Of course, in most cases they only have one or two attestations, with small numbers of members mentioned.²⁵ We won't undertake a detailed analysis here, but we will mention a couple of relevant data: only 29 of the *collegia* members are females. The most prominent associations are the *spira Asianorum* from Napoca²⁶ (45 members, 18 of the females are attested here), the *cultores Dei Soli Malagbeli* (27 persons, from the same inscription from Sarmizegetusa), the *collegium Pomarensium* (again, singular attestation from Sarmizegetusa), and the *collegium fabrum* from Sarmizegetusa. Some of these *collegia* reveal important information regarding their organization and functioning, such as the *collegium Isidis* from Potaissa, where the head of the association is mentioned and also the financial treasurer (*quaestor collegii*).²⁷

Another category to be mentioned is the **sacerdotal body**.²⁸ First of all, there is a supra-municipal or provincial group: the *sacerdotes provinciae / sacerdotes Daciae / sacerdotes Arae Augusti / sacerdotes arae Augusti nostri coronatus Daciarum trium* (8 people for the entire province, for the approximately 160 years of Roman dominion); their main attribution was to fulfil the obligations of the living emperor's cult, whose core was at Sarmizegetusa. Then we have official functions at the municipal level: *pontifices* (14 persons), *auguri* (7 persons), *flamines* (26 persons), *antistes* (3 persons) and a *haruspex* (1 person). Also on the municipal level, there are the priests of the city: 3 attestations mention the function of *sacerdos colonia Apulensis*. 3 cases from Drobeta and 4 from Porolissum list military *sacerdotes*. Obviously, priests of certain deities are attested: a total of 28 people, with the most numerous sample of those who served Jupiter Dolichenus (18 individuals), followed by priests of local Palmyrene deities (5 men).

The next set of data, on which we have focused, are the monuments that register the **age** of the deceased. The number of monuments listing this type of data is surprisingly high, as we have identified 604 ages derived from the total of 758 funerary monuments. Of course, not all ages are completely readable, but nonetheless the figures are highly illustrative for the provincial epigraphic habit. As well pertaining to manifest epigraphic habits, but perhaps also illustrative for the state of knowledge regarding exact dates of birth, one encounters the phenomenon of rounding the ages: from the 552 known ages, 176 are round ages.

²⁴ The veterans' actual involvement in social activities is a matter which has lately raised interest (for a recent overview, see Martemyanov (2017)). Nonetheless, actual figures are hard to find in bibliography, as they generally relate exclusively to the veterans involved in local administration (Królczyk (2009)).

²⁵ *Collegium Fabrum, collegium Pomarensium, cultores Dei Soli Malagbeli, cultores Dei Herculi, spira Asianorum, collegium Galatarum, collegium Aurariarum, cultores Iovis, cultores Herculanis et Cervabus, collegi Kastellum Baridustarum, collegium Sardiatarum, collegium Utriculariorum, collegium Isidis, collegium duplariorum, collegium Pontobithynorum, collegium centonariorum and collegi Iovi Cerneni.*

²⁶ Piso (2013b).

²⁷ Deac (2013) 65–73.

²⁸ For an analysis of the priests of Dacia, see Szabó (2007).

In the table below we will present the age groups distribution of all clearly readable ages:

Age group	No. of persons
Under 5 yrs.	43
6 – 10 yrs.	39
11 – 20 yrs.	76
21 – 40 yrs.	185
41 – 60 yrs.	136
61 – 70 yrs.	18
71 – 80 yrs.	27
81 – 90 yrs.	24
91 – 100 yrs.	4

Table 1. Age categories attested in Dacia

The causes of death are rarely mentioned and only when we are dealing with violent death: a man²⁹ and a young woman³⁰ were *interfecti a latronibus* (killed by robbers); the man was a *decurio* of Drobeta, and his death had been avenged by living relatives. Explicitly stating on the funerary stone that the manslaughter was avenged was most probably a way of appeasing the spirits of the dead, as well as a social statement. If considered exceptional, the place of death can also be recorded sometimes: Aurelia Flora³¹ died in Poetovio, but her tomb lied in Germisara.

Another topic worth investigating is **mobility**. We have registered in the database the people who were only temporarily part of Dacia’s social fabric, arriving in from other parts of the Empire: Palmyra (6x), Rome (4x), Sirmium (3x), Savaria (2x), Siscia (2x), etc.³² These people mention their provenience explicitly; one can assume, in certain other cases as well, that we are dealing with external mobility examples, but the database tries not register assumptions or implicit conclusions.

The last aspect, we want to highlight, are the **relationships** between individuals. The most commonly expressed relationship is that between parents and children – 620 instances. The next ties are those between spouses (287) and siblings (267). The large number of attested brothers is due, besides the normal dedications for close family, to a social phenomenon and expression pattern as well, namely the military dedications for ‘brothers-in-arms’. Other family relations are scarcely recorded: grandparent – grandchildren (9), uncle/aunt – nephew/niece (14).

29 IDR III/1, 71 (<http://www.romans1by1.com/rpeople/6791>).

30 IDR II 134 (<http://www.romans1by1.com/rpeople/7836>).

31 IDR III/3, 248 (<http://www.romans1by1.com/rpeople/5573>).

32 Sebastopol (6x), Asia (4x), Caesarea (3x), Moesia Superior (3x), Treverus (3x), Bithynus (2x), Hadrumentum (2x), , Aequum (2x), and respectively 1 person from Khara, Pergam, Viminacium, Beneventum, Tarsos, Achaera, Arretio, Ancyra, Colonia Agrippinensium, Delmata, Aequo, Gallia Cisalpina, Bithus, Sentino, Hierapolis, Batava, Amasia, Macedonia, Mediolanum, Tipasa, Mauretania Caesariana, Zigalli, Augusta Victrix, Philadelphia, Edessa, Aspendus, Falerio, Epiphania, Isauria, Aquileia, Clunia, Celeia, Flavia Solva, Theveste,, Dertona, Virunum, Iader, Marciniesis, Sclaietis, Cornovium, Serdica, Chalcis, Stobi, Cilicia, Hispellum, Carthagina, Capua, Picenum, Noviomagus.

The people of Moesia Superior

For Moesia Superior (abbreviated MS) there are 2.356 recorded **persons**, coming from 1002 **inscriptions**. While the total number of surviving inscriptions from Moesia Superior is larger, one must note that many of them are either badly deteriorated, or do not contain names. Thus, the *corpus* of the working material comprises 611 funerary dedications, 263 votive monuments, 52 *instrumenta*, 32 honorific inscriptions, 7 lists (*alba*), 5 *milliaria*, 8 construction dedications and 2 undeterminable pieces. Unlike the Dacian state of the materials, here there is a clear-cut predominance of epitaphs, which represent more than half of the inscriptions. The predominance of **Latin** is once again obvious, as only 28 inscriptions are in Greek, 4 monuments (all funerary) have bilingual, Greek-Latin texts, and 1 funerary stone from Singidunum is in Palmyrenean. On the geographic distribution, almost a quarter of the inscriptions comes from Viminacium (capital and seat of the Legio VII Claudia), followed by Scupi (early, Flavian age *colonia*) and by the cities and metropolitan area of Singidunum (city conjoined with the fort of Legio IIII Flavia Felix) and Ratiaria, with roughly 60 – 70 inscriptions each.

Regarding the **gender** ratio, the percentage of women is larger than in Dacia, as there are 527 individual attestations. We notice a predominance of Scupi (135 attestations) and Viminacium (88), but the same cannot be said about the other urban centres of the province. Anyway, the phenomenon, which links female epigraphic manifestations (especially direct ones, when the woman is the dedicator of the monument, not the object of the dedication) with civilian urbanism, is noticeable in Moesia Superior as well. Also worth noting is the fact that only 20 attestations come from votive monuments. The disproportion was to be expected in the epigraphic context of this province, but nonetheless the ratio of male votive epigraphy is overwhelming, stressing upon the manifestation of women in almost exclusively familial context – or being the object of dedications made by family members, as roughly half of the funerary dedications are for women.

The **slaves** are very much underrepresented again, as we only have 57 private and public slaves and 5 imperial ones. The difference of ratios of imperial slaves, as compared to Dacia, is remarkable and even more, all three inscriptions, which record them, are votive dedications, coming from ‘obscure’ areas (thus not the large urban and administrative centres of the province) and only one man provides an associated occupational status, that of a *vilicus stat(ionis) Aquar(um) Bas(sianarum?)*.³³ The occupational status of the non-imperial slaves is also more rarely registered than in Dacia, as there are only a couple of *vilici* and *dispensatores* recorded and, as an exception, a *statuarius*.³⁴ The *liberti* are much better represented, 124 in number. Almost half of the private freedmen are connected to the *ordo Augustalium*, stating their affiliation in private dedications, or erecting collective monuments.

33 IMS IV 104 (<http://www.romans1by1.com/rpeople/13195>).

34 IMS VI 115 (<http://www.romans1by1.com/rpeople/13600>).

The **elite attestations** contain a number of 103 decurions (with a balanced percentage provenience from the main cities), among whom 52 also fulfilled local magistrate functions (*duumvir*, *aedilis*, *quinquennalis*, etc.). Unlike Dacia, in Moesia Superior we have only one individual who was *decurio* in two cities, serving in the ‘local Senate’ of Drobeta and Viminacium alike.³⁵

Twenty-six representatives of the *ordo equester* are attested, 12 of them being associated with military positions. 20 senators were registered, among whom a single woman, the mother of a *consularis*,³⁶ and two financial procurators. We have identified 11 provincial governors, among whom Marcus Stadius Priscus,³⁷ who chronologically held his office in Moesia Superior between his governing of Dacia and Britannia. Worth mentioning is the dedication of Egnatius Marinianus from Kostolac,³⁸ governor of the province sometimes after 230 AD and – more remarkable – the father or brother³⁹ of Egnatia Mariniana – Valerian’s wife and Gallienus’s mother.

Expectedly, the **military** personnel is quite numerous – the data provides 548 individual attestations. In almost half of the cases, the unit in which the individual served is known: from a quantitative perspective, the best represented is the Legio VII Claudia of Viminacium (partially due to the existence of an *album* attesting almost one hundred men from the legion),⁴⁰ followed by Legio IV Flavia Felix of Singidunum and – interestingly enough – by Cohors II Aurelia Dardanorum from Timacum Minus. Unlike Dacia, where even the legionary veterans are heavily underrepresented, in Moesia Superior there are 365 individuals attested – among them, 321 had explicitly served in a legion and 305 of these in the locally prominent Legio VII Claudia. Although one can assume the hazard of discoveries and the post-Roman historical conditions, comparing the epigraphic manifestations of the provinces’ main legions, we can definitely see patterns of manifestation and epigraphic habits, which go beyond of any hazard.

Unlike Dacia, only a couple of individuals list their **association** membership, if we exclude the *Augustales* – only 4 types of *collegia* are registered,⁴¹ with only 2 women as members.⁴² A total of 16 *cultores* are mentioned by the dedication of the *collegium* of Heros from Singidunum.⁴³ Compared to the variety of sacerdotal functions from Dacia, in Moesia Superior we encountered only 31 men. These are municipal priests (6 *pontifices*, 2 *auguri* and 3 *flamines*) and 18 divinities’ priests. What lacks are the provincial priests, as well as – for example – the *antistes* and *haruspices*, although their actual existence in the cities cannot be doubted.

35 IMS II 75 (<http://www.romans1by1.com/rpeople/11871>).

36 IMS VI 27.

37 PIR² S 880, Piso 1993, 66 – 73, no. 16; Birley (2005) 151 – 155, no. 29 (<http://www.romans1by1.com/rpeople/596>).

38 IMS II 68 (<http://www.romans1by1.com/rpeople/11855>).

39 PIR² E 25.

40 CIL III 8110.

41 *Cultores* of Heros, *collegium Fabrum*, *collegium centonariorum*, *collegium Herculis*.

42 AÉ 2009, 1190 (*mater* of the *collegium Herculis* <http://www.romans1by1.com/rpeople/14150>); Feraudi-Gruénais 2014, 254 (*mater* of an unknown *collegium*, <http://www.romans1by1.com/rpeople/14780>).

43 IMS I 2 (<http://www.romans1by1.com/rinscriptions/3903>); the exact term on the inscription is *collitores*.

The **ages** of death are registered on 522 monuments, with a total of 473 fully readable figures. Among them, 196 are round ages. Compared to Dacia, we noticed a greater representativeness of the 71 – 80 age group, but smaller figures for the superior ones.

Age group	No. of persons
Under 5 yrs.	23
6 – 10 yrs.	34
11 – 20 yrs.	60
21 – 40 yrs.	139
41 – 60 yrs.	141
61 – 70 yrs.	45
71 – 80 yrs.	19
81 – 90 yrs.	9
91 – 100 yrs.	3
Above 100 yrs.	3

Table 2. Age categories attested in Moesia Superior

The causes of death are recorded even more rarely than in Dacia. We have attestations of violent deaths, as in the case of Serenia Quarta⁴⁴ from Singidunum, who was *interfecta ab hoste*, but we do not know if she was killed by robbers or by raiders, nor do we have references of a possible retribution. Another example that can be brought forward, is that of a veteran,⁴⁵ *ex beneficiarius consularis*, decurion of Viminacium, who is recorded to have died in Britannia, without further details on circumstances.

Compared to Dacia, we must mention two interesting facts regarding extra-provincial **mobility**: there are more attestations, but fewer areas of origin: There are numerous examples of external mobility, represented by people coming from Sirmium (9x), Salona (6x), Nicopolis (5x), Aquileia (3x), Pautalia (3x), Trimontium (3x), Mursa (3x), Vindobona (2x), Virunum (2x), Antiochia (2x), Thyatirenus (2x), Isaura (2x), Germanicia (2x), etc.⁴⁶

The most commonly expressed **relationship** is that between parents and children (240), but the number and percentage of attestations alike are clearly much smaller than in Dacia. The next ties are those between spouses (222) and siblings (167). Other relationships, such as grandparent – grandchildren (4), uncle/aunt – nephew/niece (2), are severely underrepresented.

44 IMS II 110 (<http://www.romans1by1.com/rpeople/9638>).

45 IMS II 110 (<http://www.romans1by1.com/rpeople/12114>).

46 Mentioned only once as provenience places: Kybira, Augusta Treverorum, Liburnum, Lucus, Iader, Traianopolis, Nicomedia, Pergam, Heraclea, Philippi, Cyrbera, Margum, Ancyra, Bracara, Brixia, Caesarea Augusta, Varvaria, Carthago, Hierapytna, Verona, Flanona, Pelagonia, Roma, Sagalasso and Hierapolis as provenience places.

Statistics for Moesia Inferior

The data coming from the province of Moesia Inferior (abbreviated MI) has been the subject of some exceptions, which we will present first, before tackling it. Even though the timeframe, we focus on, comprises the period beginning with the creation of the Roman provinces and up to the third century, in this case some exceptions have been made. One of it concerns specifically the dating of the introduced data: in the case of the Greek colonies Histria, Tomis and Kallatis the exception consists in including a wider timespan, more precisely we covered also the data from before the creation of the province Moesia Inferior,⁴⁷ introducing the Classical and Hellenistic inscriptions and this due to the fact that for these cities, which provide an epigraphic continuity, we considered it useful to make available the entire sample. However, these pre-Roman inscriptions are not included here in the evaluation, so that the result is not distorted.

Another exception, that we made, regards the anachronistic inclusion of all the inscriptions which come from the not yet organised province of Moesia Inferior under its name (MI). This choice is supported not only by the sometimes imprecise dating of the inscriptions (i.e. an inscription from the first century A.D. could come, from an administrative point of view, either from the province Moesia, either from Moesia Inferior), but also by the need of providing a certain degree of uniformity to the sample. However, in the following lines the interpretation will take into consideration only the data which comes from our timeframe.

Moesia Inferior is a militarised province,⁴⁸ but also benefitting from a very strong urbanization. The urban settlements are numerous (besides the Greek cities, which have a different juridical status in relation with the Roman power, we add Marcianopolis⁴⁹ and Nicopolis ad Istrum,⁵⁰ established by Trajan as well as urban settlements which emerged adjacent to camps such as Cius, Halmyris, Montana, Novae, Noviodunum, Oescus, Sexaginta Prista, Transmarisca and Troesmis,⁵¹ but also settlements such as *castella*, *canabae*, *civitates peregrinae*: *civitas Ausdecensium*, *civitas Dianensium*, *Libida*, *Melta*,⁵² or the Roman city of Tropaeum Traiani).⁵³

From a statistical point of view, we have here the largest number of attested **individuals** – 6.359, but not the largest number of **inscriptions** – 2.561 (those from Dacia are more numerous). Besides these inscriptions there are also 38, which might have come from the Roman period, but due to the uncertain dating, we have not taken them into account for the present study. Compared to the other provinces of our focus the large number of attested individuals can be explained through the type of inscriptions we have: we are dealing with a large number

47 At first the Greek cities started to be controlled by the Romans (the *terminus post quem* for this historical landmark is considered as being the expedition of Crassus, while the *terminus ante quem* is Ovid's exile at Tomis: Ov. Tr. 2.197 – 200; Pipidi (1974) 256 – 260. Avram (1999) 49 – 51 offers a more precise date for this: P. Vinicius' mission, sometimes between 3-2 B.C.), next the province Moesia was created, which in 86 A.D. was divided by Domitian into the provinces Moesia Inferior and Superior (Matei-Popescu (2010) 25), and these were further reorganised under Diocletian and Constantine into two new provinces: Moesia Secunda and Scythia Minor (Suceveanu u. Barnea (1991) 154 – 155).

48 Most of the units are situated along the Danube, namely the two and then three legions (Matei-Popescu (2010) 27): Legio V Macedonica stationed at Troesmis, Legio I Italica stationed at Novae, Legio XI Claudia stationed at Durostorum, additional the *auxilia*, and the Classis Flavia Moesica; the areas where units were not situated along the Danube are those of Montana, Tomis and Tropaeum Traiani.

49 Gerov (1988) 124 – 127.

50 Gerov (1988) 118-122; Boteva (2014) 195 – 204.

51 Bottez (2009) 26.

52 Bottez (2009) 26.

53 Bottez (2009) 27; Matei-Popescu (2014) 205 – 223.

of *alba* (109 – large compared to the other provinces), which obviously attest a higher number of individuals, but which seldom provide consistent prosopographical data.

Regarding the most frequently attested **type of monuments**, besides *alba* we have a high number of funerary inscriptions (1.078), followed by votive monuments (925), honorific ones (183), construction inscriptions (42), military diplomas (44), regulations/ decrees (14). Phantom data is unfortunately also present at this level: we deal not only with unknown inscription types (75), but also with unknown dating (out of the overall number of inscriptions, 215 could not be dated, representing however less than 10%). When analysing the types of inscriptions, one can see from this statistics that the individuals from Moesia Inferior were keener on exhibiting not their religiosity or their ‘contracts’ with the gods, but rather their lifetime events, seldom mirroring identity or status (i.e. Titia Matrina’s funerary monument⁵⁴ mentions the place of her death, *ad vila sua* (sic!)).⁵⁵

Somehow surprising is the relatively presence of **Greek** (1.147) and Latin (1.371) inscriptions, to which we add 43 bilingual inscriptions. However, the Greek inscriptions tend to be wordier, and often record a higher number of individuals. In the bilingual inscriptions the phenomenon of code switching is present,⁵⁶ involving either the usage of both languages or, among others, the usage of transliterations from one language to another.

From a **gender** perspective we have an overwhelming number of males (5.305), only 809 females,⁵⁷ and a number of 245 individuals whose gender is unknown due to the fragmentary state of the inscriptions. The low representativeness of women is not surprising; however, we do have some atypical cases, where women took the stance of benefactors of their city and community, such as in the case of Ἄβα,⁵⁸ daughter of Ἐκαταίος, wife of Ἡράκων, or other in which they are even attested as land owners, such as Messia Pudentilla.⁵⁹ Besides few exceptions, women are attested in the inscriptions most frequently as wives, mothers, sisters, and daughters, being epigraphically present especially when they were deceased, and were remembered for their virtue and piety (*dulcissima, bene merens, carissima, dextrata, eusebeia, nobilissima, optima, pientissima*).⁶⁰

All **social and juridical categories** are represented, however all of them are underrepresented: the peregrines (1.124), the imperial slaves (3), the private slaves (26), the freedmen (98), but also the Roman citizens (3.323 out of which only 356 are women). This can be explained by the poor prosopographical data provided by some of the inscriptions, by the fragmentary state of other inscriptions, or due to their uncertain dating. Despite all these, it results that the population of the province was quite heterogenic, comprising all social categories and juridical statuses, and including individuals whose places of origin were varied, from the hearth of the Greek world, Athens, and up to the closer Greek cities such as Nicomedia.

54 ISM V 29 = CIL III 14214(2) (<http://romans1by1.com/rpeople/9475>).

55 ISM V 29 (<http://romans1by1.com/rinscriptions/4018>).

56 For code-switching in Moesia Inferior see as well Curcă (2011) 71 – 80.

57 For an analysis on the women coming from Histria, Tomis, and Kallatis see Pázsint (2017) 49 – 70.

58 ISM I 57 = SEG 18 293 = SEG 24 1112 (<http://romans1by1.com/rpeople/413>).

59 ISM I 359 = CIL III 14447, ISM I 360 (<http://romans1by1.com/rpeople/3074>).

60 Pázsint (2017) 59 – 60.

As expected, one of the most prolific urban centres, that records most of the individuals, is Tomis, the city was the metropolis of the *laevus Pontus* probably at beginning with the reign of Trajan,⁶¹ and until recently,⁶² it was believed to be the seat of the governor.⁶³ Among the most notable individuals we find here several governors (20), pontarchs (10)⁶⁴ and other individuals involved in the local administration. Overall, in Moesia Inferior we have 108 provincial governors,⁶⁵ among whom Caius Ovinus Tertullus⁶⁶ appears most frequently (in 26 inscriptions), despite the fact that not much is known about his life or his career. While a governor of Moesia Inferior, we do know that he made a gift (*τῶν ἡμεῖν τὴν δωρεὰν δεδωκότων*) to a *θίασος* dedicated to Cybele at Tomis.⁶⁷ To these members of the elite, who also possess the Roman citizenship, we should add the members of the local elite, who might not be Roman citizens, but local citizens.

From the **administration** we have a reasonable number of local magistrates (485). In the case of the Greek cities, even in the imperial period, the type of organisation was maintained, the two most important bodies were the Council and the People (*βουλή καὶ δῆμος*).⁶⁸ Among the local magistratures we find the archonship, to which we add the position of *ἀγορανόμος*, that of *ἀστυνόμοι*, *ταμίας* and *γυμνασίαρχος*. In case of the Roman cities, such as Oescus, Durostorum, Troesmis and Tropaeum Traiani, we have a duumviral organization, with *duumviri*, *aediles*, and *quaestores*.⁶⁹

An important category is that of the **military personnel** (726) and the **veterans** (464).⁷⁰ Most originate from the Greek East, which is in accordance with the wider reality regarding the origin of the individuals attested in Moesia Inferior,⁷¹ but evidence for western origin is also present.⁷² Among the veterans some (26) decided upon discharge to settle in different centres than the ones they came from. Their number is relatively small, and in some cases (5), they even come from other cities of the same province,⁷³ which leads not to a truly speaking mobility.

Military personnel and veterans are also present in great numbers at Tomis (57 military men, 19 veterans), probably due to its harbour – where it seems that the personnel was in charge of policing and surveillance,⁷⁴ but also on the Danube limes, where we have the headquarters of the legions: Oescus (25 military men, 11 veterans), Novae (89 military men, 22 veterans), Durostorum (18 military men, 11 veterans), Troesmis (25 military men, but 253 veterans), to which we add the inner areas, like that of Montana (123 military, and 2 veterans), and Tropaeum (117 military men, 4 veterans). In the case of Montana and its surrounding area, which represents together with Tomis and Tropaeum Traiani an exception to the geographical position

61 Deininger (1965) 120.

62 Piso (2014) 489 – 504.

63 Stoian (1962) 28 – 39; Haensch (1997) 332 – 335; Matei-Popescu (2010) 27; Buzoianu u. Bărbulescu (2012) 147.

64 Maurer (2014) 141 – 188.

65 Stein (1940) 18 – 34, 58-111; Boteva (1996) 239 – 247.

66 PIR² O 191; Stein (1940) 84; Boteva (1996) 242 (<http://romans1by1.com/rpeople/3073>).

67 ISM II 83 = SEG 27 399 (<http://romans1by1.com/rinscriptions/1804>).

68 For Tomis see Buzoianu u. Bărbulescu (2007), 7 – 48.

69 Aparaschivei (2009) 161 – 194; Matei-Popescu (2014) 212.

70 The individuals attested in the military diplomas have been recorded as veterans.

71 Curcă u. Zugravu (2005) 313 – 329; Mihăilescu-Bîrliba (2009) 153 – 160; Mihăilescu-Bîrliba (2012) 125 – 132; Mihăilescu-Bîrliba (2015) 141 – 146; Mihăilescu-Bîrliba u. Piftor (2016) 125 – 136; Matei-Popescu (2017) 139 – 159.

72 Mihăilescu-Bîrliba u. Dumitrache (2015) 465 – 469.

73 Regarding the mobility inside the province see Mihăilescu-Bîrliba (2013) 185 – 188.

74 Matei-Popescu (2010) 27.

of the units,⁷⁵ the placing of units has been previously explained through its mining activity.⁷⁶ In the case of Tropaeum, the high number is due to the inscription, which mentions the list of the Roman soldiers, who died following Trajan’s war with the Dacians.⁷⁷

Much better represented than in Moesia Superior and Dacia are the **private associations**, which count among their members 1108 individuals; the associative phenomenon is attested in the area from the Hellenistic period on, but the evidence we have points to its bloom during the Roman period, when we encounter not only a wide variety of cult associations, but also professional associations. The discrepancy between the provinces in what concerns the proliferation of the associations is not necessarily related to the discoveries, but rather to the evolution of the provinces and to their historical background.

Compared to the evidence coming from Dacia and Moesia Superior, the **sacerdotal body** of Moesia Inferior seems to be much narrower, with 139 individuals mentioned as divinity priest; however, especially in the urban areas, attestation of the Imperial cult is frequent, as we have 42 imperial priests/priestesses, some of whom are not from the respective cities, but they established there (such as Αὐρήλιος Πρίσκοις Ἰσιδωροῦ⁷⁸ and Αὐρήλιος Πρίσκοις Ἀννιανοῦ⁷⁹ – probably his brother –whom may have come from Flavia Neapolis, and whom were also *ποντάρχης*).⁸⁰

Even though the **age** sample is quite large (584 ages are mentioned, out of which 102 are indecipherable or uncertain, to which we add some uncommon forms, or the result of an error – e.g. XLIIIX, IIIIX –) we will not approach age related interpretations, such as life expectancy or age structure, because they have been already tackled.⁸¹ The distribution of age groups can be seen in the following table:

Age group	No. of persons
Under 5 yrs.	40
6 – 10 yrs.	27
11 – 20 yrs.	64
21 – 40 yrs.	131
41 – 60 yrs.	144
61 – 70 yrs.	48
71 – 80 yrs.	19
81 – 90 yrs.	5
91 – 100 yrs.	2
Above 100 yrs.	2

Table 3. Age categories attested in Moesia Inferior

75 Matei-Popescu (2010) 27; Matei-Popescu (2014) 214.

76 Velkov (1980) 85 – 101; Gerov (1988) 101 – 107; Matei-Popescu (2014) 214.

77 CIL III 14214 = ILS 9107 = AÉ 1980 794 = ISM IV 8 (<http://romans1by1.com/rinscriptions/3512>).

78 ISM II 96 (<http://romans1by1.com/rinscriptions/2460>).

79 ISM II 97 = SEG 36 690 (<http://romans1by1.com/rinscriptions/2461>).

80 For the Imperial cult in Moesia Inferior see Bottez (2009).

81 Piftor (2007 – 2008) 46 – 57, Mihăilescu-Bîrliba u. Curcă (2010) 139 – 153, Mihăilescu-Bîrliba u. Piftor (2011) 419 – 451.

Compared to the overall number of inscriptions, and to the other provinces, the data coming from Moesia Inferior which records the age of the deceased is not so large, even though it forms the largest of the three samples. As one can see from the table above, in Moesia Inferior there are attestations of individuals who supposedly lived over 100 years, more precisely the eldest character, [---] Marcianus,⁸² supposedly died at CIII years old, followed (in age) by another one, named [---]a Publi[---]⁸³ who lived CIII years; however these two individuals are certainly not the eldest of all characters because in Moesia Superior an individual named Caius Catus qui et Felix,⁸⁴ surprisingly and supposedly died at CL years, which is probably a way of saying the deceased man was very old, a metaphor for his age, especially keeping in mind that seldom they did have accurate information on the year they were born in.

The causes of death are mentioned when they are the result of extreme medical conditions, or situations such as: perhaps epidemics/accident (the siblings, an *ignotus*, son of Ἰσαγόρας⁸⁵ and Ἡρακλέω, son of Ἰσαγόρας,⁸⁶ are both mentioned on an inscription as having died at the age of 5 years), death in childbirth,⁸⁷ manslaughter.⁸⁸ Among similar cases of tragic deaths there is a case where all the children (Decius,⁸⁹ Saturninus,⁹⁰ Varrus⁹¹) of a family from Ulmetum died within seven months.⁹² To this we add some examples which point to the dangers of the era: Aelius? Mucianus⁹³ had a grim fate, being *deceptus a barbaris*, while Valerius Marcus⁹⁴ died at a relatively young age (XVIII) being killed by thieves; an *ignotus*⁹⁵ died in an ambush in Rome, while Timo, son of Dassius,⁹⁶ was *interfectus a Costobocis* (inscription: *interfec/ti (!) a Costobocos (!)*).

‘Foreigners’ are also attested in Moesia Inferior,⁹⁷ some of whom appear to have settled in for a longer period, or for life.⁹⁸ The **relocation** of some individuals is connected to their profession: we have attestations of military personnel and of other professionals, but there is also a category of individuals whose presence cannot be explained due to the poor prosopographical data.⁹⁹ Their origin is diverse, these persons came not only from the Greek East, but also from the Roman West (even though this place of origin is not so common). Besides the military personnel, one of the best represented categories is that of them involved in transport and commerce, some of the professionals were organised in private associations (such as the *Οἶκος τῶν*

82 ILBulg 168 = CIL III 6131 = CIL III 7454 (<http://romans1by1.com/rpeople/12459>).

83 AÉ 1984, 744 = AÉ 1985, 750 (<http://romans1by1.com/rpeople/14817>).

84 IMS VI 68 (<http://romans1by1.com/rpeople/13363>).

85 ISM III 86 (<http://romans1by1.com/rpeople/8468>).

86 ISM III 86 (<http://romans1by1.com/rpeople/8468>).

87 ISM III 134 = SEG 25 752 (<http://romans1by1.com/rpeople/8395>).

88 See the following footnotes, n. 93, 94, 95, and 96.

89 IGLR 207 (<http://romans1by1.com/rpeople/13868>).

90 IGLR 207 (<http://romans1by1.com/rpeople/13869>).

91 IGLR 207 (<http://romans1by1.com/rpeople/13870>).

92 IGLR 207 = AÉ 1976 632 (<http://romans1by1.com/rinscriptions/5930>).

93 ILBulg 155 = CIL III 12392 (<http://romans1by1.com/rpeople/12433>).

94 AÉ 1901, 19 = CIL III 14587 = ILS 8504 (<http://romans1by1.com/rpeople/13828>).

95 AÉ 1901, 20 = CIL III 14588 = IMS III/2 106 = ILS 8150 (<http://romans1by1.com/rpeople/13842>).

96 AÉ 2005, 1315 = AÉ 2006, 1190 (<http://romans1by1.com/rpeople/13981>).

97 Mihăilescu-Bîrliba (2009) 153 – 160; Iacob (2013) 214 – 222.

98 ISM I 356 = SEG 49 1009 (<http://romans1by1.com/rinscriptions/1520>).

99 Iacob 2013, 219 – 220.

Ἀλεξανδρέων).¹⁰⁰ As their places of origin we mention: Alexandria,¹⁰¹ Byzantion,¹⁰² Corinth,¹⁰³ Hermione,¹⁰⁴ Nicomedia,¹⁰⁵ or Prusias ad Hypium.¹⁰⁶ On the other way around, there are also individuals coming from the province of Moesia Inferior whom are attested in other parts of the ancient world,¹⁰⁷ from the nearer provinces, up to the more distant western provinces.

The **relationships** between individuals are also a strong component of the dataset. The connexions, which have been made, based not only on the specific mentioning in the text, but sometimes it was deduced based on the onomastics (especially in the brother to brother relations, or that of brothers and fathers). As expected, most relationship types are those between parents and children, followed by spouses and siblings, grandparents and grandchildren, and those between uncles/ aunts and nephews/ nieces.

To the rarer examples we add those which invoke step-brothers or step-parents,¹⁰⁸ and those which indicate adoption, as in the case of Χρυσάων,¹⁰⁹ son of Γαῦκος, who was adopted by Ἀχιλλᾶς, and as a result was also the brother of Ἀχιλλεύς,¹¹⁰ son of Ἀχιλλᾶς. Both brothers are attested at Tomis as part of an association of Δενδροφόροι dedicated to Cybele, and Ἀχιλλεύς was also a member in an association dedicated to Dionysos at Histria.¹¹¹ Such another example comes from Histria, from an inscription which is dated during the 2nd/1st century BC, where an *ignotus*,¹¹² might have been the adopted son of Μενεχάρμος.¹¹³

From the overall number of attested inscriptions it seems that those recording military personnel and veterans, tend to have a higher occurrence rate of family members, especially children, but this fact might be misleading because they were more likely to afford both, having a larger family and erecting monuments.

100 ISM II 153 = IGR I 604 = RICIS 618/1005 = SEG 47 1040 (<http://romans1by1.com/rinscriptions/1812>).

101 ISM II 153 = IGR I 604 = RICIS 618/1005 = SEG 47 1040 (<http://romans1by1.com/rinscriptions/1812>); ISM II 463 = IGLR 28 (<http://romans1by1.com/rinscriptions/21>).

102 ISM II 308 (<http://romans1by1.com/rinscriptions/6594>).

103 ISM I 271 (<http://romans1by1.com/rinscriptions/1451>).

104 ISM II 375 = SEG 24 1081 (<http://romans1by1.com/rinscriptions/33>).

105 ISM II 256 (<http://romans1by1.com/rinscriptions/3047>).

106 ISM II 248 = SEG 30 845 (<http://romans1by1.com/rinscriptions/19>); ISM II 462 (<http://romans1by1.com/rinscriptions/20>).

107 See Avram (2013) for the prosopography of all the individuals coming from the Black Sea poleis and externally attested.

108 Aurelius Sabinus (ILBulg 216: <http://romans1by1.com/rpeople/12522>) step-brother of Aurelia (ILBulg 216 <http://romans1by1.com/rpeople/12523>), Aurelius Silvanus (ILBulg 216 <http://romans1by1.com/rpeople/12520>) step-father of Aurelius Sabinus (ILBulg 216 <http://romans1by1.com/rpeople/12522>); Secunda (IGLNovae 107 = ILBulg 325 = IGBulg V 5262 = SEG 47 1153 <http://romans1by1.com/rpeople/10828>) step-mother of Basileus (IGLNovae 107 = ILBulg 325 = IGBulg V 5262 = SEG 47 1153 <http://romans1by1.com/rpeople/10825>), Titius [--]r (IGLNovae 104bis <http://romans1by1.com/rpeople/10819>) step-son of Suriana (IGLNovae 104bis <http://romans1by1.com/rpeople/10817>) and step brother of Claudius Surianus (IGLNovae 104bis <http://romans1by1.com/rpeople/10818>), Titius Heraclius (IGLNovae 104bis <http://romans1by1.com/rpeople/10820>) step-son of Claudius Surianus (IGLNovae 104bis <http://romans1by1.com/rpeople/10818>), and the step-son of Suriana (IGLNovae 104bis <http://romans1by1.com/rpeople/10817>), [--er] inus (IGLNovae 104bis <http://romans1by1.com/rpeople/10816>) the step-father of Claudius Surianus (IGLNovae 104bis <http://romans1by1.com/rpeople/10818>); Aurelia Dusia (ISM I 337 <http://romans1by1.com/rpeople/3029>) step-mother of Aurelius Erculanus (ISM I 337 <http://romans1by1.com/rpeople/3028>).

109 ISM II 83 = SEG 27 399 (<http://romans1by1.com/rpeople/3663>).

110 ISM I 99 = SEG 19 477; ISM I 100 = SEG 17 342; ISM II 83 = SEG 27 399 (<http://romans1by1.com/rpeople/431>).

111 For the involvement of individuals in associations coming from both, Histria and Tomis, see Ruscu (2014) 139 – 152.

112 ISM I 123 = SEG 24 1134 (<http://romans1by1.com/rpeople/467>).

113 The word *φουσει* is in the lacuna of the text, but D.M. Pippidi, the editor of the inscription, considers that the word is the one missing.

Future prospects

After having attained this phase with the database we expect to develop it even further, by: a) introducing new data coming from other provinces of the Roman Empire (Pannonia Inferior and Superior), b) updating the database with new information corresponding to the first provinces we focused on (Dacia, Moesia Inferior, Moesia Superior), c) scientifically processing the information we obtained, both from a quantitative and qualitative point of view.

Overall conclusions

The scientific contributions of this research were made evident throughout the article. Nonetheless, we need to highlight once again that their value lies in the ‘infrastructure’ of the research, in the fact that we have analysed all prosopographic data provided by the epigraphy of the area in a standardized and exhaustive manner. Some of the most striking results are connected to the ways in which epigraphy was employed for communicating in society; thus, Dacia massively brings forth a type of urban epigraphy, intertwined with social promotion (the predominance of votive monuments illustrates that), while Moesia Superior is massively represented by epitaphs. Another important aspect is the difference between Latin language epigraphs, more matter of fact and poorer in biographical details, and Greek language ones, more descriptive and often offering minor life and death details as well. Feminine epigraphy appears highly connected to urbanization, but at the same time less represented in the very urbanized Moesia Inferior. In this case, we are facing two factors: the first are the female traditional roles in Greek culture and the latter is the nature of the sources, as in Moesia Inferior a consistent part of the names come from lists, where members of different public bodies, councils, etc. – thus males – were inscribed. The last large issue we want to underline is the representativeness of veterans – and here we are definitely facing a feature of epigraphic habit: while veterans are rather well represented in the two Moesias, in Dacia they are highly underrepresented. For sure, in such a militarized province we cannot assume that the actual number of veterans integrated in civil society was low, but rather that they were not accustomed to inscribe their ex-military status on monuments.

The advancement, of the research techniques from the last decades has led to an increased inter-disciplinarily approach in humanistic studies, which started to use the techniques and methodologies coming from other fields, in order to provide a more accurate view on the long-gone worlds. As already proven by the growing number of databases used in ancient history, its advantages are considerable, and while they do not always replace the effort and the old methodologies, they do offer an easier and more flexible way of organising, filtering, researching and sharing the information. The examples, we have provided, try to point out the fact that through the use of specific databases certain patterns and epigraphic habits can be clearly and easily distinguished, joining together more fragments of information on ancient population. Considering all these benefits, our initiative is certainly a *plaidoyer* to databases.

Acknowledgement

This work was supported by a grant of the Romanian Ministry of Research and Innovation, through UEFISCDI, project no. PN-III-P4-IDPCE-2016-0255.

Caption of figures

	Dacia	Moesia Superior	Moesia Inferior
Alba	0	7	109
Funerary	758	611	1076
Votive	1226	263	923
Honorific	117	32	181
Unknown	10	2	75
Regulations/Decrees	0	0	10
Construction	66	8	42
Agonistic	0	0	3
Instrumenta	387	52	41
Military diploma	62	14	45
Defixiones	0	0	1
Letters	0	1	7
Milliaria	0	5	39
Contracts	19	0	0
Mosaic	0	7	0

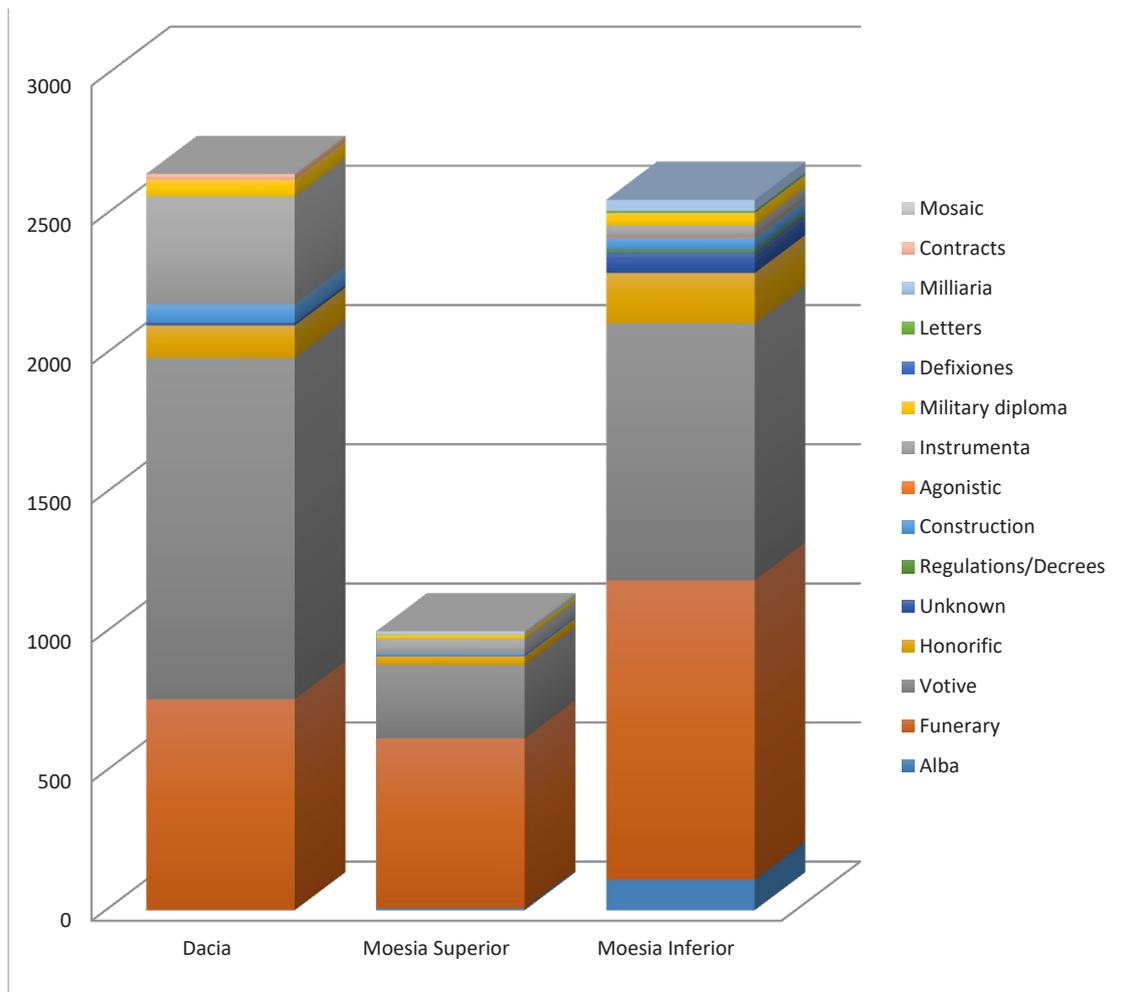


Fig. 1. Comparative statistics on the types of sources

	Dacia	Moesia Superior	Moesia Inferior
Greek	37	28	1142
Latin	2600	969	1368
Greek-Latin	5	4	42
Palmyrean-Latin	3	1	0

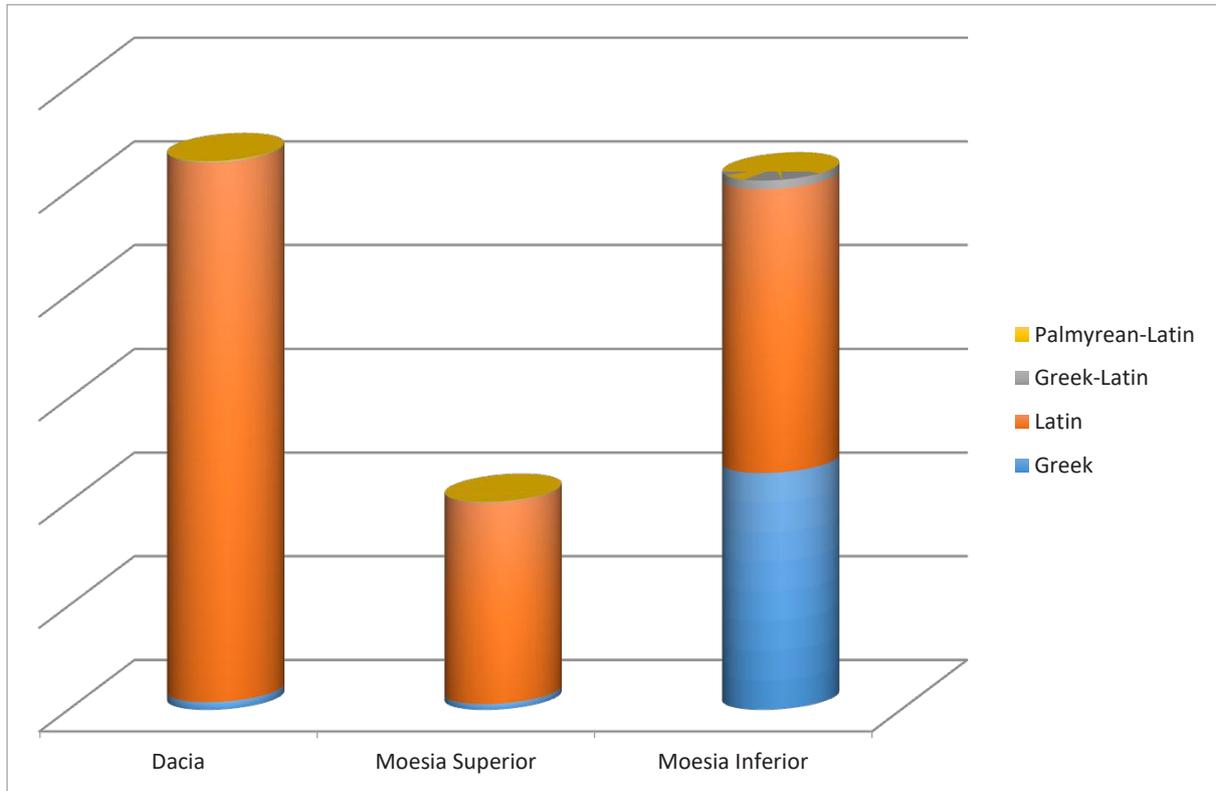


Fig. 2. Comparative statistics on the language of the sources

	Dacia	Moesia Superior	Moesia Inferior
Female	682	527	860
Male	3530	1794	5820
Unknown	142	35	275

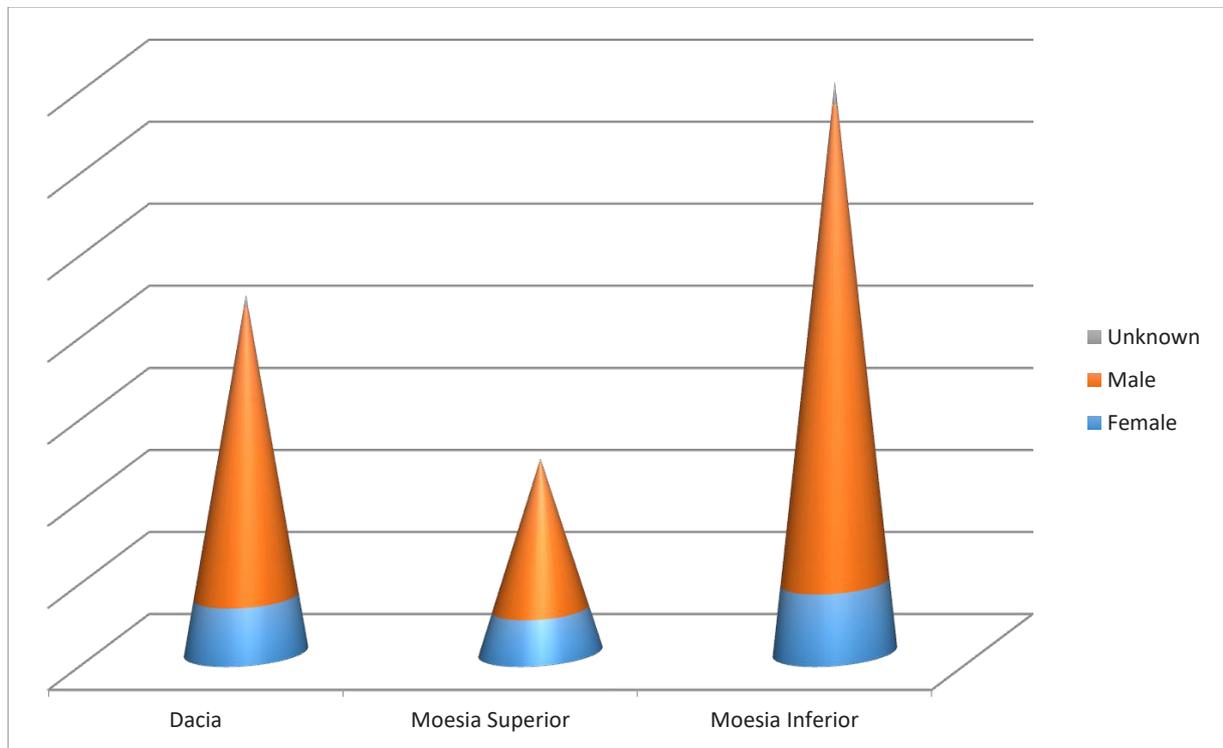


Fig. 3. Comparative statistics on the gender of the attested persons

	Dacia	Moesia Superior	Moesia Inferior
Citizens male	2848	1516	2922
Citizens female	556	425	357
Freedmen	147	92	71
Freedwomen	29	32	26
Peregrines male	235	95	940
Peregrines female	42	38	216
Slaves male	58	39	21
Slaves female	18	16	4
Imperial slaves male	32	5	3
Imperial slaves female	3	0	0
Imperial freedmen	26	2	0
Imperial freedwomen	0	0	0
Unknown	360	96	2395

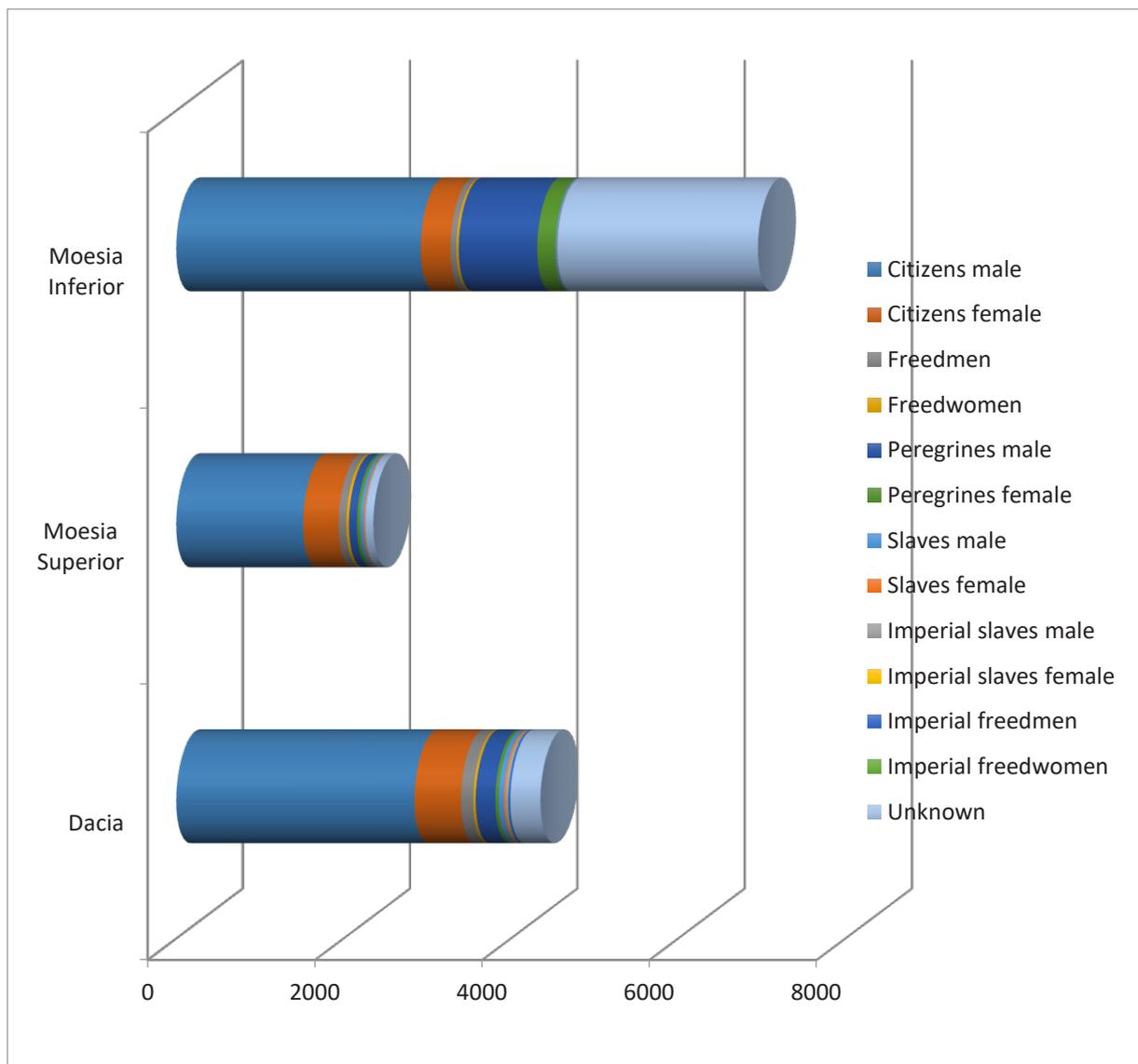


Fig. 4. Comparative statistics on the juridical status/gender of the attested persons



Fig. 5. Map of Dacia, Moesia Inferior & Moesia Superior (from the Digital Atlas of the Roman Empire - <http://dare.ht.lu.se/>)

Bibliography

Alston (2003): Alston, Richard, *Soldier and Society in Roman Egypt. A Social History*, London and New-York.

Aparaschivei (1990): Aparaschivei, Dan, “Instituții și elite în orașele romane ale provinciei Moesia Inferior. Studiu statistic”, in: L. Mihăilescu-Bîrliba (ed.), *Structuri etno-demografice la Dunărea de Jos (sec. I – VII p. Chr.)*, Iași, 161 – 194.

Avram (1999): Avram, Alexandru, *Inscriptions grecques et latines de Scythie Mineure. Callatis et son territoire*, Paris-București.

Avram (2013): Avram, Alexandru, *Prosopographia Ponti Euxini Externa*, Louvain.

Bărbulescu (1994): Bărbulescu, Mihai, Potaissa. Studiu monografic, Turda.

Bărbulescu (1997): Bărbulescu, Miha, *Das Legionslager von Potaissa (Turda). Castrul legionar de la Potaissa (Turda)*, Zalău.

Beu-Dachin (2014): Beu-Dachin, Eugenia, *The Latin Language in the inscriptions of Roman Dacia*, Cluj-Napoca.

Birley (2005): Birley, Anthony R., *The Roman Government of Britain*, Oxford.

Boteva (1996): Boteva, Dilyana, “Legati Augusti Pro Praetore Moesiae Inferioris AD 193 – 217/218”, *ZPE* 110, 239 – 247.

Boteva (2014): Boteva, Dilyana, “Trajan and his Cities in Thrace. Focusing on the Two Nicipoleis”, in: I. Piso, R. Varga (eds.), *Trajan und seine Städte*, 195 – 204.

Bottez (2009): Bottez, Valentin, *Cultul imperial în provincia Moesia Inferior (sec. I – III p. Chr.)*, București.

Buzoianu u. Bărbulescu (2007): Buzoianu, Livia u. Bărbulescu, Maria, “Tomis”, in: D.V. Grammenos, E.K. Petropoulos (eds.), *Ancient Greek Colonies in the Black Sea 2. Volume I*, Oxford, 287 – 336.

Buzoianu u. Bărbulescu (2012): Buzoianu, Livia u. Bărbulescu, Maria, *Tomis. Comentariu istoric și arheologic. Historical and archaeological commentary*, Constanța.

Curcă (2011): Curcă, Roxana, “The Bilingual Inscriptions of Moesia Inferior. The Historiographic Framework”, *Classica et Christiana* 6/1, 71 – 80.

Curcă u. Zugravu (2005): Curcă, Roxana u. Zugravu, Nelu, “Orientaux dans la Dobroudja romaine. Une approche onomastique”, in: V. Cojocaru (ed.), *Ethnic Contacts and Cultural Exchanges North and West of the Black Sea from the Greek Colonization to the Ottoman Conquest*, Iași, 313 – 329.

Deac (2013): Deac, Dan A. “Being an Isiac in Potaissa. Short Remarks on RICIS 616/0102 (= CIL III 882)”, *ReDIVA* I, 65 – 73.

Deiniger (1965): Deiniger, J., *Die Provinziallandtage der römischen Kaiserzeit von Augustus bis zur Ende des dritten Jahrhunderts n. Chr.*, München.

Feraudi-Gruénais (2014): Feraudi-Gruénais, Franziska, “Corrigendum ad *ZPE* 188, 2014, 275 – 283 (Nomina mulierum in Ulpiana)”, *ZPE* 189, 254.

Gerov (1988): Gerov, Boris, *Landownership in Roman Thracia and Moesia (1st – 3rd century)*, Amsterdam.

Gorea (2013): Gorea, Maria, "Stèle funéraire à banquet d'un Palmyrénien vétéran à Potaissa (Dacia Superior)", *Semitica et Classica* 6, 2013, p. 291 – 296.

Haensch (1997): Haensch, Rudolph, *Capita provinciarum. Statthaltersitze und Provinzialverwaltung in der römischen Kaiserzeit*, Mainz am Rhein.

Iacob (2013): Iacob, Mihaela, *Moesia Inferior între Orient și Occident. Identitatea culturală și economică a provinciei în contextul lumii romane*, București.

Królczyk (2009): Królczyk, Krzysztof, *Veteranen in der Donauprovinzen des Römischen Reiches (I. – III. Jn. N. Chr.)*, Poznań.

Mandemakers u. Dillon (2004): Mandemakers, Kees u. Dillon, Lisa, "Best Practices with Large Datasets on Historical Populations", *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 37.1, 34 – 38.

Martemyanov (2017): Martemyanov, Alexei, "On the method of Determination of Level of Social Activity of Veterans in Provinces of the Roman Empire (The Case of Lower Moesia)", *Studia Europaea Gnesniana* 16, 415 – 429.

Matei-Popescu (2010): Matei-Popescu, Florian, *The Roman Army in Moesia Inferior*, București.

Matei-Popescu (2014): Matei-Popescu, Florian, "Tropaeum Traiani", in: I. Piso, R. Varga (eds.), *Trajan und seine Städte*, Cluj-Napoca, 205 – 224.

Matei-Popescu (2017): Matei-Popescu, Florian, "The Dacians from Moesia Inferior", in: L. Mihăilescu-Bîrliba (ed.), *Migration, Kolonisierung, Akkulturation im Balkanraum und im Osten des Mittelmeerraumes*, Konstanz, 139 – 159.

Maurer (2014): Maurer, Karin, "Der Pontarch des westpontischen Koinons", *Dacia N.S.* LVIII, 141 – 188.

Mihăilescu-Bîrliba (2004): Mihăilescu-Bîrliba, Lucrețiu, *Sclavi și liberti imperiali in provinciile romane din Illyricum*, Iași.

Mihăilescu-Bîrliba (2009): Mihăilescu-Bîrliba, Lucrețiu, "Străinii în Moesia Inferior. Posibilități de identificare", in: L. Mihăilescu-Bîrliba (ed.), *Structuri etno-demografice la Dunărea de Jos (sec. I – VII p. Chr.)*, Iași, 153 – 160.

Mihăilescu-Bîrliba (2012): Mihăilescu-Bîrliba, Lucrețiu, "Les Pontobithyniens à Troesmis", in: D. Boteva-Boyanova, L. Mihăilescu-Bîrliba, O. Bounegru (eds.), *Pax Romana. Kulturaustausch und Wirtschaftsbeziehungen in den Donauprovinzen des römischen Reichs*, Kaiserslautern, 125 – 132.

Mihăilescu-Bîrliba (2013): Mihăilescu-Bîrliba, Lucrețiu, "La mobilité interne en Mésie Inférieure", in: G. Tsetskhladze, S. Atasoy, A. Avram, Ş. Dönmez, J. Hargrave (eds.), *The Bosphorus: Gateway between the Ancient West and East (1st Millennium BC – 5th Century AD). Proceedings of the Fourth International Congress on Black Sea Antiquities*, Istanbul, 14th – 18th September 2009, Oxford, 185 – 188.

Mihăilescu-Bîrliba (2015): Mihăilescu-Bîrliba, Lucrețiu, "Prosopographic Remarks on the Population of Troesmis (Lower Moesia)", in: A. Tomas (ed.), *Ad fines Imperii Romani. Studia Thaddaeo Sarnowski septuagenario ab amicis, collegis discipulisque dedicata*, Warsaw, 141 – 146.

Mihăilescu-Bîrliba u. Curcă (2010): Mihăilescu-Bîrliba, Lucrețiu u. Curcă, Roxana, "New Perspectives on the Demography of the Roman Province of Moesia Inferior", *The Mankind Quarterly* 51.2, 139 – 153.

Mihăilescu-Bîrliba u. Dumitrache (2015): Mihăilescu-Bîrliba, Lucrețiu u. Dumitrache, Iulia, “Les militaires originaires d’Italie à Troesmis. Notes prosopographiques”, in: L. Vagalinski, N. Sharankov (eds.), *Limes XXII. Proceedings of the 22nd International Congress of Roman Frontier Studies*, Ruse, Bulgaria, September 2012, Sofia, 465-469.

Mihăilescu-Bîrliba u. Piftor (2011): Mihăilescu-Bîrliba, Lucrețiu u. Piftor, Valentin, “L’espérance de vie, la structure d’âge et la mortalité en Mésie Inférieure et en Scythie Mineure. Un regard comparatif”, in: V. Nikolov, K. Bacvarov, H. Popov (eds.), *Interdiziplinäre Forschung zum Kulturerbe auf der Balkanhalbinsel*, Sofia, 419-451.

Mihăilescu-Bîrliba u. Piftor (2016): Mihăilescu-Bîrliba, Lucrețiu u. Piftor, Valentin, “Die Fremden in der lokalen Elite aus Troesmis (Moesia Inferior)”, in: L. Mihăilescu-Bîrliba (ed.), *Migration, Kolonisierung, Akkulturation im Balkanraum und im Osten des Mittelmeerraumes*, Konstanz, 125 – 136.

Nemeti u. Nemeti (2014): Nemeti, Sorin u. Nemeti, Irina, “Civic Space and Municipal Statutes in Potaissa”, in: S. Cociș (ed.), *Archäologische Beiträge Gedenkschrift zum hundertsten Geburtstag von Kurt Horedt*, PAT 7, Cluj-Napoca, 85 – 98.

Pázsint (2017): Pázsint, Annamária – I., “Between oikos and polis. The Life of Women in Istros, Tomis, and Callatis (6th century BC – 4th century AD)”, *Ephemeris Napocensis XXVII*, 49-70.

Piftor (2007-2008): Piftor, Valentin, “L’espérance de vie et la structure d’âge de la population féminine en Mésie Inférieure (Ier – IIIe siècles ap. J.-Chr.)”, *SAA* 13 – 14, 46 – 57.

Pippidi (1974): Pippidi, Dumitru M., “În jurul alipirii orașelor vest-pontice la Imperiul Roman”, *Studii Clasice* 16, 256 – 260.

Piso (1993): Piso, Ioan, *Fasti provinciae Daciae I. Die senatorischen Amtsträger*, Bonn.

Piso (2013a): Piso, Ioan, *Fasti provinciae Daciae II. Die ritterlichen Amtsträger*, Bonn.

Piso (2013b): Piso, Ioan, “*Studia Porolissensia (III)*”, *Acta Musei Porolissensis* 25, 159 – 176.

Piso (2014): Piso, Ioan, “Le siège du gouverneur de Mésie Inférieure”, in: V. Cojocaru, A. Coșkun, M. Dana (eds.), *Interconnectivity in the Mediterranean and Pontic World during the Hellenistic and Roman Periods*, Cluj-Napoca, 489-504.

Piso et al. (2016): Piso, Ioan, Ardevan, Radu, Fenechiu, Carmen, Beu-Dachin, Eugenia u. Lalu, Ștefania, *Lexicon Epigraphicum Daciae*, Cluj Napoca.

Ruscu (2014): Ruscu, Ligia, “On Cult Associations at Istros and Tomis”, *Ephemeris Napocensis XXIV*, 139 – 152.

Stein (1940): Stein, Arthur, *Die Legaten von Moesien*, Leipzig.

Stoian (1962): Stoian, Iorgu, Tomitana. *Contribuții epigrafice la istoria cetății Tomis*, București.

Suceveanu u. Barnea (1991): Suceveanu, Alexandru u. Barnea, Alexandru, *La Dobroudja romaine*, Bucarest.

Szabó (2007): Szabó, Ádám, *Daciai papság*, Budapest.

Varga (2014): Varga, Rada, *The Peregrini of Roman Dacia*, Cluj-Napoca.

Varga (2017): Varga, Rada, “Romans 1by1. Documenting a Population Database for the Roman world”, in: S. Orlandi, R. Santucci, F. Mambrini, P. M. Liuzzo (eds.), *Digital and traditional Epigraphy in Context. Proceedings of the EAGLE 2016 international Conference*, Roma, 333 – 341.

Varga (2017b): Varga, Rada, “Romans1by1 v.1.1. New Developments in the Study of Roman Population”, *Digital Classics Online*, 3, 2, 44 – 50.

Varga u. Rusu-Bolindeț (2016): Varga, Rada u. Rusu-Bolindeț, Viorica, “Provincial Landmarks of the official Power. The Praetorium Consularis of Apulum”, in: R. Varga, V. Rusu-Bolindeț (eds.), *Official Power and local Elites in the Roman Provinces*, London – New York, 115 – 126.

Velkov (1977): Velkov, Velizar I., *Cities in Thrace and Dacia in Late Antiquity*, Amsterdam.

Velkov (1980): Velkov, Velizar I., *Roman Cities in Bulgaria. Collected Studies*, Amsterdam.

Abbreviations

AÉ = *Année Épigraphique*, Paris.

CIL = *Corpus Inscriptionum Latinarum*, Berlin.

IDR = *Inscriptiones Daciae Romanae*, București-Paris.

ILBulg = B. Gerov, *Inscriptiones Latinae in Bulgaria repertae*, Serdica, 1989.

ILS = Hermann Dessau, *Inscriptiones Latinae selectae*. 3 vols. In 5 parts, Berlin, 1892 – 1916.

IMS = *Inscriptions de la Mésie Supérieure*, Beograd.

ISM I = Dionisie M. Pippidi, *Inscriptiones Scythiae Minoris graecae et latinae*. Vol. 1. *Inscriptiones Histriae et vicinia*, Bucharest, 1983.

ISM II = Iorgu Stoian, *Inscripțiile din Scythia Minor*. Vol. 2. *Tomis et territorium*, Bucharest, 1987.

ISM III = Alexandru Avram, *Inscriptions de Scythie Mineure*. Vol. 3. *Callatis et territorium*, Bucharest, 1999.

ISM IV = Emilian Popescu, *Inscriptions de Scythie Mineure*. Vol. 4. *Tropaeum – Durostorum – Axiopolis*, Bucharest, 2015.

ISM V = Emilia Doruțiu-Boila, *Inscriptiones Scythiae Minoris graecae et latinae*. Vol. 5. *Capidava, Troesmis, Noviodunum*, Bucharest, 1980.

IGBulg = G. Mihailov (ed.), *Inscriptiones Graecae in Bulgaria Repertae*, Sofia, 1958 – 1970.

IGLR = E. Popescu, *Inscripțiile grecești și latine din sec. IV – XIII descoperite în România*, Bucharest, 1976.

IGLNovae = Kolendo, Jerzy, Violeta Božilova, *Inscriptions Grecques et Latines de Novae (Mésie Inférieure)*, Paris, 1997.

SEG = *Supplementum Epigraphicum Graecum*, Amsterdam – Leiden.

Authors¹¹⁴

Dr. Rada Varga

Babeş-Bolyai University
Cluj-Napoca

Email: radavarga@gmail.com

Annamária-Izabella Pázsint

Babeş-Bolyai University
Cluj-Napoca

Email: a.i.pazsint@gmail.com

Imola Boda

Babeş-Bolyai University
Cluj-Napoca

Email: boda_imola@yahoo.com

Dan Augustin Deac

Babeş-Bolyai University
Cluj-Napoca

Email: dandea12@gmail.com

114 The rights pertaining to content, text, graphics, and images, unless otherwise noted, are reserved by the author. This contribution is licensed under CC BY 4.0 International.

κριθή oder πυρός? - Eine systematische Annäherung im klassischen Athen mit Hilfe der Kookkurrenzanalyse von eAQUA und den Word Frequency Statistics der Perseus Digital Library

Sven-Philipp Brandt

Abstract: This article deals with the long-lasting discussion about grain (e.g. demand, crop growing, trade and consumption of wheat or barley) in Classical Athens. The author of this article tries to augment this discussion with the help of some tools of the Digital Humanities, especially the analysis of cooccurrences of eAQUA and the word frequency study tool of the Perseus Digital Library. The Blended Reading will be the main method, so the quantitative and the qualitative analysis of the ancient literary sources will be combined in order to get new information about the distribution of barley and wheat in classical times.

1. Einleitung

In diesem Artikel¹ soll die anhaltende Forschungsdiskussion über Getreide (Bedarf, Anbau, Import, Konsum) im klassischen Attika mit Hilfe einiger Tools der Digital Humanities um einige Gesichtspunkte erweitert werden.² Denn genauso, wie die Berechnungen des tatsächlichen Getreideanbaus im klassischen Attika im Allgemeinen recht spekulativ erscheinen,³ lässt sich dies auch beim Anbauverhältnis von Weizen zu Gerste feststellen.⁴

1 Die Idee für diesen Artikel entstand im Rahmen des von ‚Niedersächsisches Vorab‘ geförderten Projekts ‚Nachhaltigkeit als Argument‘ im Rahmen der Projektreihe ‚Wissenschaft für nachhaltige Entwicklung‘. Ein ganz besonderer Dank gilt Prof. Reinhold Scholl und Prof. Charlotte Schubert für die zahlreichen Hinweise und Anregungen.

2 Vgl. hierzu die gegensätzlichen Positionen von Garnsey 1988, Gallant 1991 und Sallares 1991, die den Import von Getreide und dessen Auswirkung auf die politischen Handlungen Athens eher gering einschätzen, während Whitby 1998 und Morley 2007 wieder Jardé 1925 folgen und den Getreidehandel als politisch handlungsleitend für das Athen der klassischen Zeit betrachten. Vgl. hierzu neuerdings auch Moreno 2007, S. 11–33 und Bissa 2009, S.169–191, die in ihren ausführlichen Monographien auf Grundlage der Auswertung von Inschriften und bisheriger Forschungsliteratur ein differenzierteres Bild geliefert haben, grundlegend aber ebenfalls von einer grossen Bedeutung der Getreideimporte ausgehen.

3 Vgl. Moreno 2007, Table 1, S. 10, der die Forschungsdiskussion aufgreift, tabellarisch visualisiert und damit ersichtlich macht, welche unterschiedlichen Faktoren die jeweiligen Forschungspositionen zu Grunde gelegt haben. Ebenso ausführlich unter Einbeziehung moderner Anbaudaten Bissa 2009, S. 172–176.

4 Die einzig verlässliche Quelle IG II2 1672 – eine Inschrift, die in spätklassischer Zeit ein Verhältnis von 9,3 zu 1 zu Gunsten der Gerste nahelegt – wurde bereits mit Hilfe von klimatischen Krisenszenarios ‚erklärt‘ und stattdessen ein Mittel von 4 zu 1 bis 3 zu 1 vermutet, vgl. Ruffing 2012, S. 60f. Einen weiteren Anhaltspunkt liefert das Getreidegesetz des Agyrrhios von 374/3, in dem es um die Reform einer Naturalsteuer auf Gerste- und Weizenträge der Klerucheninseln Lemnos, Imbros und Skyros geht, vgl. Engels 2000, S. 97–124.

Zunächst wird daher ein kurzer Einblick in die Problematik gegeben, die mit der Frage nach dem Getreidebedarf im klassischen Athen einhergeht. Diese Problematik geht von dem Aspekt aus, in welchem Ausmaß Getreide importiert werden musste und führt schließlich zur Überlegung, inwieweit die Notwendigkeit des kontinuierlichen Getreideimports die athenische Politik der klassischen Zeit beeinflusst hat. Hierzu gibt es einige Punkte zu beachten, die von den verschiedenen Forschungspositionen mit unterschiedlicher Gewichtung einbezogen wurden: 1. Wie viele Menschen lebten im klassischen Athen? Bereits diese Zahl ist umstritten und offenbart in der Forschungsliteratur eine Bandbreite von 150.000 bis über 400.000 Einwohnern.⁵ Diese Unterschiede ergeben sich aus unterschiedlichen Hochrechnungen, die auf der Grundlage von gesicherten Angaben zu den Athener Bürgern beruhen. Als *communis opinio* gelten meist die Zahlen von Mogens H. Hansen, der kurz vor Ausbruch des Peloponnesischen Krieges von 300.000 Einwohnern⁶ und im 4. Jahrhundert von ca. 200.000 Einwohnern ausgeht.⁷ Josiah Ober wiederum geht von durchschnittlich ca. 250.000 Einwohnern in der Region Attika aus⁸ und schätzt diese auch für die Zeit unmittelbar vor dem Peloponnesischen Krieg nicht viel höher ein, da er ca. 50.000 Einwohner außerhalb Attikas als Kleruchen oder Soldaten in außerattischen Garnisonen annimmt und auch die Koloniegründungen wie Thurioi und Amphipolis in seine Berechnungen mit einbezieht.⁹ So kommt er auch für das Ende des 4. Jahrhunderts noch auf ca. 240.000 Einwohner in Attika.¹⁰ 2. Wie viel Getreide benötigte ein Einwohner? Diese Frage lässt sich ebenfalls nur schwer beantworten, da es einen Unterschied macht, wie viel von welcher Getreideart konsumiert wurde. Der Versuch, sich dem Wert mit Hilfe antiker Quellen und neuzeitlicher Ergebnisse zu nähern, hat ein weiteres breites Zahlenspektrum von 4,8 Medimnoi (≈252 l) bis 8,7 Medimnoi (≈457 l) Jahresverbrauch pro Person hervorgebracht – ohne dabei aber zwischen Gerste und Weizen oder entsprechenden Unterarten zu unterscheiden.¹¹ Hinzu kommt die von Josiah Ober aufgeworfene Frage, wie groß der Anteil der Bevölkerung war, der am Existenzminimum lebte.¹² 3. Wieviel des Getreidebedarfs konnte durch Getreideanbau in Attika abgedeckt werden? Diese Frage baut nicht nur auf den vorherigen, unsicheren Zahlen auf, sondern birgt selbst noch zwei unsichere Faktoren. Zunächst die Frage, wie viel der attischen Fläche überhaupt zum Getreideanbau genutzt werden konnte. Denn dies schwankt zwischen 17 – 50% der 2.400 km² Attikas.¹³ Die zweite Unsicherheit ergibt sich aus der Getreideart und dem möglichen Ertrag. Wie im Folgenden noch zu zeigen ist, machte es aufgrund der klimatischen Gegebenheiten einen enormen Unterschied, wo in Attika Getreide angebaut wurde, und im Hinblick auf den tatsächlichen Ertrag, ob es sich

5 Vgl. Bissa 2009, S. 172.

6 Vgl. Hansen 1988, S. 26.

7 Vgl. Hansen 1988, S. 12, ähnlich auch Garnsey 1988, S. 90. Vgl. auch die weiteren Ausführungen von Hansen, 2006.

8 Vgl. Ober 2016, S. S. 239, Abb. 7.1.; 444. Vgl. zu Josiah Obers Methode auch S. 123, Abb. 4.1.

9 Vgl. Ober 2016, S. 292. Vgl. hierzu auch Ober 2008, S. 74, fig. 2.5; S. 293, fig. E.1/E.2

10 Vgl. Ober 2016, S. 144, Tab. 4.4.

11 Vgl. Bissa 2009, S. 173.

12 Josiah Ober berechnet hier je nach Auslegung der verfügbaren Daten, dass 41,1%–56,6% der Einwohner Attikas am Existenzminimum lebten, vgl. Ober 2016, S. 144, Tab. 4.4. Somit würde sich bereits auf Grundlage seiner Methode ein erheblicher Unterschied im Konsum ergeben, da es auch auf den Getreidebedarf eine große Auswirkung hat, ob die Differenz von 15,5% ,immerhin ca. 36.300 Einwohner, nun nur das Nötigste an Getreide konsumierten oder, im Sinne einer Mittelschicht, mehr verbrauchen konnten.

13 Vgl. Bissa 2009, S. 173, während Morenos Synthese ausgewählter Forschungspositionen 20–40% ergibt, woraufhin er von 35% ausgeht, vgl. Moreno 2007, S. 10.

hierbei um Gerste oder eine Weizenart handelte.¹⁴ Insbesondere der klimatische Aspekt wurde bei den Hochrechnungen nicht beachtet¹⁵ und auch die Ergebnisse neuerer Pollenproben, die gerade in klassischer Zeit einen geringeren Getreideanbau als beispielsweise in mykenischer oder byzantinischer Zeit nahelegen,¹⁶ konnten noch nicht mit einbezogen werden.

Trotz dieser zahlreichen unsicheren Faktoren etablierte sich in den Altertumswissenschaften die Annahme, dass Gerste zunächst überwiegend in archaischer Zeit angebaut wurde, aufgrund seiner geringeren Nahrhaftigkeit in klassischer Zeit jedoch immer mehr durch Weizenanbau zurückgedrängt wurde,¹⁷ wie selbst im Neuen Pauly an prominenter Stelle zu lesen ist.¹⁸

Der vorliegende Artikel möchte sich mit dieser Annahme auseinandersetzen und mit Hilfe des Textminingtools des Greek Word Study Tools und dessen Funktion der Word Frequency Statistics sowie der Kookkurrenzanalyse von eAQUA auf der Textbasis der Perseus Digital Library des Perseus Projects¹⁹ überprüfen,²⁰ ob sich diese Vermutung auch anhand der überlieferten literarischen Quellen für die klassische Zeit Athens stützen lässt und damit die verschiedenen Hochrechnungen um eine bisher vernachlässigte Komponente ergänzen. Denn wenn sich diese Entwicklung bestätigen lassen sollte, sollte sich das auch in der Häufigkeit der Verwendung der Wörter κριθή (Gerste) und πῦρος (Weizen) in literarischen Quellen niederschlagen, also Weizen in spätklassischer Zeit deutlich häufiger erwähnt werden. Hierzu sollen zunächst alle schriftlich-literarischen Quellen der klassischen Zeit, in denen eine deklinierte Form von κριθή

14 Osborne 2004 unterscheidet bei seinen Hochrechnungen beispielsweise gar nicht zwischen Getreide und Weizen, vgl. hierzu auch Moreno 2007, S. 10.

15 So geht Sallares 1991, S. 311f. zwar auf Attika als semiaride Zone ein, vernachlässigt aber das sehr trockene Südostattika.

16 Vgl. hierzu insbesondere Kouli 2012, fig. 3, S. 272; ferner auch Triantaphyllou 2010, fig. 5, S. 19 und Kouli 2009. Die Getreidesorten werden bei Pollenproben nicht gesondert angeführt, da sie kaum verlässlich zu unterscheiden sind und daher nur als cerealia angegeben werden. Vgl. zu den Pollenproben der Marathonebene auch Kouli 2009.

17 Vgl. Morley 2007, S. 24, der zudem den Geschmack von Gebäck aus Gerste als Grund für die Geringschätzung für Gerste anführt, was Ruffing 2012, S. 60 jedoch explizit ablehnt. Schmitz 2004, S. 468 und 2007, S. 21 differenziert diese Tendenz dahingehend, dass die enorme Steigerung der Handelsaktivitäten in klassischer Zeit einen Import von Weizen in solcher Menge nach Athen ermöglicht habe, dass dieser Gerste als wichtigstes Getreide abgelöst habe und Gerste nur noch zur Ernährung von Sklaven und Tiere eingesetzt worden sei.

18 „Die wichtigsten Entwicklungen in der Gesch[ichte] des ant[iken] Getreideanbaus waren zunächst der Rückgang der Gerste in vielen Regionen, während gleichzeitig zunehmend Weizen angebaut wurde, und die Tendenz, Spelzweizen durch Nacktweizen zu ersetzen.“ Robert Sallares: s. v. Getreide, in: Der Neue Pauly. Enzyklopädie der Antike. Band 4 (Epo-Gro), herausgegeben von Hubert Cancik und Helmuth Schneider, Stuttgart/Weimar 1998, Sp. 1031.

19 <http://www.perseus.tufts.edu/> (abgerufen am 08.07.2018).

20 Der Grund für diese Entscheidung liegt darin, dass die Texte der Perseus Digital Library Open Access verfügbar sind und die Ergebnisse dadurch transparent sind. Zwar ist das Textcorpus des TLG deutlich umfangreicher, doch gilt dies vor allem für ‚periphere‘ Autoren, die für die Forschungsfrage ohnehin keine Bedeutung haben. Ein weiterer Aspekt, der für die Perseus Digital Library spricht, ist das Greek Word Study Tool, das die Möglichkeit bietet, Wortfelder wie eben κριθή so anzeigen zu lassen, dass man eine nach Autoren sortierte Trefferliste zur Verfügung gestellt bekommt, und die wiederum sämtliche Deklinationen mit einbezieht: <http://www.perseus.tufts.edu/hopper/wordfreq?lang=greek&lookup=kriqh%2F> (abgerufen am 08.07.2018). Lediglich die Autoren, die wie die Atthidographen nicht in der Perseus Digital Library verzeichnet sind, werden zusätzlich über den TLG gesucht und gesondert aufgeführt.

und *πυρός* vorkommt und die im Kontext der Polis Athen anzusehen sind, mit einbezogen werden. Dies umfasst alle Autoren, die nachweislich in Athen in klassischer Zeit wirkten.²¹

Die Verwendung und Kombination beider Textminingtools ist für die Überprüfung der These besonders hilfreich, da sie dem methodischen Ansatz des *Blended Readings* folgt,²² der eine Verknüpfung verschiedener Analyseverfahren vorsieht.²³ Dabei ist das Greek Word Study Tool mit der Funktion der Word Frequency Statistics das grundlegende Verfahren, das die Datenbestände durch eine Frequenzanalyse zunächst strukturiert und einen Überblick über die Häufigkeit der jeweiligen Begriffe gibt. Durch die bei diesem Tool automatisch erfolgende Zuteilung der Ergebnisse zu den jeweiligen Autoren kann zudem ein schneller quantitativer Einblick in den für die These relevanten Textbestand gewonnen werden. Als vertiefendes Analyseinstrument bietet sich anschließend die Kookkurrenzanalyse von eAQUA an, da sie Bedeutungs- und Sinnzusammenhänge offenlegt. So ermöglicht die Verwendung des Tools eine Analyse der jeweiligen Suchwörter, die sowohl deren Häufigkeit im Textcorpus als auch signifikante Kookkurrenzen mit Nachbarn nach Position anzeigt und visualisiert.²⁴ Die Verwendung und die Kombination dieser beiden Tools funktioniert dabei als eine Art modularer Prozess, der im Sinne des *Blended Readings* einen „epistemologisch angemessenen Umgang mit großen Textdatenmengen“²⁵ gewährleistet. So kann sowohl das quantitative als auch das semantische Vorkommen der Begriffe *πυρός*, *κριθή* und *σίτος* analysiert werden.²⁶

Darauf aufbauend werden abschließend alle Beispiele auch in ihrem Kontext betrachtet und mit Hilfe der gängigen Methoden der Altertumswissenschaften auch qualitativ ausgewertet. So wird geprüft, ob Weizen in den Quellen der klassischen Zeit tatsächlich vermehrt genannt und ein Umschwenken der attischen Landwirte auf Weizen in klassischer Zeit nachvollziehbar wird.

Es stellt sich freilich die Frage, ob nicht eine gründliche Auswertung aller vorhandenen Pollenproben für Attika eine zielführendere Methode wäre, um sich dieser Diskussion zu nähern. Diese zweifelsfrei sehr sinnvolle Methode der Geobotanik ist jedoch für Attika aus zweierlei Gründen wenig hilfreich. Zunächst ist das Einzugsgebiet, über das eine Pollenprobe Auskunft geben könnte, sehr gering. Da das antike Attika eine Fläche von ca. 2400 – 2750 km² besaß, bedarf es hierfür also sehr vieler Pollenproben, die aber aufgrund der großflächigen Bebauung Attikas mit der Metropole Athen schwer zu nehmen sind. Hinzu kommt, dass der südliche Teil Attikas zur subariden Zone gezählt werden kann und damit eine Feuchtigkeit benötigende

21 Welche Autoren das betrifft, lässt sich an der Trefferliste Tabelle 2 nachvollziehen, in der auch die negativen Ergebnisse, also die Autoren aufgeführt werden, die weder *κριθή*, noch *πυρός* verwenden. Als Autorengrundlage wurden auch Autoren wie Herodot hinzugezogen, der zwar nicht Attika geboren wurde, aber zumindest nachweislich dort wirkte, vgl. Forsdyke 2005, 243. Bei Plut. an seni 3, Mor. 785b ist sogar eine Ode des Sophokles an Herodot erwähnt, die zwar nicht überliefert ist, jedoch die enge Verbindung der beiden offenbart. Vgl. zudem die auffälligen Parallelen von Hdt. 3,119,6 und Soph. Antig. 909-912.

22 Vgl. hierzu grundlegend Lemke / Stulpe 2016, S. 43–54.

23 Vgl. <https://sozmethode.hypotheses.org/139> (abgerufen am 30.07.2018).

24 Vgl. hierzu ausführlich Schubert 2011, S. 33–53.

25 Vgl. <https://sozmethode.hypotheses.org/139> (abgerufen am 30.07.2018).

26 Es findet zwar spätestens beim Höchstpreisedikt des Diokletian ein Wandel von *πυρός* zu *σίτος* als Begriff für Weizen statt, doch lassen sich für das klassische Athen hierfür keine Anhaltspunkte finden. Im Gegenteil scheint die Gleichsetzung von Weizen und *σίτος* eine Konsequenz der Ausbreitung des hochwertigeren Weizens in römischer Zeit zu sein, obwohl auch Plin. NH. 18,48,1 noch klar zwischen den verschiedenen Getreidesorten trennt.

Pollenprobe kaum zu nehmen ist.²⁷ Ein gesondertes Problem dieser Fragestellung ist zudem, dass die Unterscheidung verschiedener Getreidepollen sehr schwierig ist und nur indirekt über die Bestimmung von spezifischen Parasiten geschehen kann. Daher findet sich bei Auswertungen von Pollenproben stets der allgemeine Begriff *Cereals*.²⁸ Aus diesem Grund ist eine textmining-basierte Analyse der schriftlichen Quellen im Sinne des *Blended Readings* mit anschließender qualitativer Auswertung die zielführendere Methode zur Annäherung an die Forschungsfrage, auch wenn dies nicht ausschließt, die vorhandenen Ergebnisse der Pollenproben in die qualitative Auswertung mit einzubeziehen.

2. Hinführung - Eigenschaften von Weizen und Gerste

Bereits in der klassischen Antike waren verschiedene Getreideformen für den gezielten Ackerbau bekannt, sodass im klassischen Athen eine gewisse Varianz sowohl beim Getreideanbau als auch -verzehr bestand. Anhand der Quellen zeichnen sich für Attika zwei Getreidearten ab, die in der Region angebaut wurden: einerseits Weizen, andererseits Gerste. Dies ist vor allem mit den Wuchseigenschaften der jeweiligen Pflanzenarten sowie den klimatischen Bedingungen zu erklären, die in Attika durchaus komplex und lokal unterschiedlich sind. Der Bereich nördlich des Hymettos kann, wie der Großteil des mediterranen Gebiets, als semiaride Zone kategorisiert werden, der südliche Teil Attikas hingegen eher als aride Zone.²⁹ Da die klimatischen Bedingungen in der klassischen Zeit den heutigen sehr ähneln,³⁰ kann davon ausgegangen werden, dass diese klimatische Trennung schon in klassischer Zeit vorherrschte. Dies bedeutet aber, dass auch beide Zonen in etwaige Hochrechnungen des Getreideanbaus mit einbezogen werden müssen.³¹

Im Folgenden sollen kurz die Wuchseigenschaften von Gerste und Weizen erläutert werden. Gerste, κριθή, gilt als die älteste kultivierte Getreideart, die im griechischen Mutterland angebaut wurde.³² Der Vorteil beim Anbau von Gerste liegt darin, dass diese deutlich weniger Niederschlag als viele andere Getreidearten benötigt und auch auf kargen Böden wie in Attika

27 Dass das Gebiet südlich des Hymettos nicht nur allgemein landwirtschaftlich, sondern speziell auch für Getreideanbau massiv genutzt wurde, zeigen die Ausführungen zur Siedlungsstruktur der Region von Lohmann 1993, S.136-139 sowie jene zur so genannten Prinzessinnenfarm mit ihrem Dreschplatz von Goette 2000, S. 78 – 83. Hinsichtlich der Niederschläge und der Grundwassersituation, vgl. Van Liefferinge et al. 2012.

28 Vgl. Kouli 2012, S. 270.

29 Sofern man die Empfehlung des ‚United Nations Environment Programme‘ zu Grunde legt, der folgenden Ariditätsindex definiert: Der AI ergibt sich aus dem Quotienten des Werts des jährlichen Niederschlags zum Wert der jährlichen potenziellen Evapotranspiration und legt für die semiaride Zone einen AI von 0,5–0,2 zu Grunde und für die aride Zone einen AI von 0,2–0,05, vgl. hierzu Feng / Fu 2013, S. 10082f. Im Hinblick auf Attika und den Osten des griechischen Festlands, vgl. *ibid.*, S. 10084, fig. 1.

30 Vgl. Sallares 2007, S. 15–37.

31 Vgl. hierzu den Überblick von Moreno 2007, Table 1, S. 10.

32 Vgl. Philochoros FGtH 328 F73, Athen. 6, 27 p. 235 CD: διὸ καὶ ἐν τῷ τοῦ βασιλέως νόμῳ γέγραπται ταυτί (Krates 362 F 7) «ἐπιμελεῖσθαι δὲ τὸν βασιλεύοντα τῶν τε ἀρχόντων, ὅπως ἂν καθιστῶνται, καὶ τοὺς παρασίτους ἐκ τῶν δήμων αἰρῶνται κατὰ τὰ γεγραμμένα. τοὺς δὲ παρασίτους ἐκ τῆς βουκολίας ἐκλέγειν ἐκ τοῦ μέρους τοῦ ἑαυτῶν ἕκαστον ἐκτέα κριθῶν δαίνυσθαι τε τοὺς ὄντας Ἀθηναίῳ ἐν τῷ ἱερῷ κατὰ τὰ πάτρια τὸν δ' ἐκτέα παρέχειν εἰς τὰ ἀρχεῖα τῷ Ἀπόλλωνι τοὺς Ἀχαρνέων παρασίτους ἀπὸ τῆς ἐκλογῆς τῶν κριθῶν». Später auch Plin. NH 18, 72. Der Brauch, dass die Braut beim Hochzeitszug ein Gerät zum Rösten von Gerste trug, spricht ebenfalls für eine sehr alte Anbautradition in Attika, vgl. Schmitz 2004, S. 218; 253.

gedeihen kann.³³ Ein Nachteil ist jedoch, dass Gerste im Gegensatz zum Weizen kein Gluten beinhaltet und damit das Gerstenmehl zum Brotbacken ungeeignet war, da Gluten für das Aufgehen des Teigs beim Backvorgang notwendig ist.³⁴

Die in der Antike kultivierten Weizenarten besaßen wiederum ausreichend Gluten, um sich gut zum Backen zu eignen. Doch dafür hatten die Weizensorten – in unterschiedlicher Ausprägung – deutlich höhere Ansprüche an Bodenbeschaffenheit und Niederschlag.

Zunächst einmal lässt sich Weizen – unabhängig von der genetischen Einteilung der Biologie – in Nackt- und Spelzweizen unterscheiden. Ersterer lässt sich dabei nach der Ernte durch gewöhnliches Dreschen gewinnen, während Spelzweizen zu fest an der Pflanze sitzt und nur nach einem Röstvorgang abgetrennt werden kann. Daher ging die Tendenz bereits früh dahin, den leichter zu verarbeitenden Nacktweizen anzubauen.³⁵ Hier waren besonders die Getreidesorten *triticum durum*, *turgidum* und *aestivum* gefragt, wobei auch diese sich durchaus unterschieden. So eignete sich *aestivum* durch seine hohe Feuchtigkeit ideal zur Mehlerverarbeitung, benötigte jedoch auch deutlich mehr Niederschlag und war somit in Attika besonders schwer anzubauen. Mit weniger Niederschlag konnten *durum* und *turgidum* angebaut werden, wobei sich *durum*, wie der Name vermuten lässt, nur schwer zermahlen und eher zu Grieß verarbeiten ließ, während *turgidum* deutlich weniger Mehlertrag brachte als *aestivum*.

Wenn man nun diese Grundlagen auf die klimatischen Gegebenheiten sowie das Relief Attikas überträgt, liegt nahe, dass Gerste diejenige Getreidepflanze war, die am einfachsten in Attika anzubauen war. *Triticum durum* und *turgidum* dürften zwar in den feuchteren Gebieten im Norden wie der Marathonebene oder dem Gebiet um Dekeleia angebaut worden sein, doch zeigen die Pollenproben aus der Marathonebene eine sehr geringe Anzahl an Getreidepollen, was für einen geringen Anbau in diesem Gebiet spricht.³⁶ Auch dieser Umstand weist bei gleichzeitig recht hoher landwirtschaftlicher Aktivität³⁷ im südlichen, arideren Raum darauf hin, dass es einige unbekannte Faktoren gibt, die eine Annäherung mit Hilfe des Textminings als weiteren Indikator für das Anbauverhältnis sinnvoll erscheinen lassen.

3. Auswertung der Ergebnisse

Eine rein quantitative Auswertung der Treffer ist nicht möglich, da bei der Digitalisierung der antiken Corpora zwar alle Wörter und Formen mit eigenen Identifiern versehen wurden, aber dabei lediglich die Form an sich betrachtet und getaggt wurde, nicht jedoch die jeweilige Bedeutung, wie beispielsweise bei der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek (DNB). Für die Fragestellung dieser Arbeit ist das jedoch problematisch, da der Nominativ (ὁ) πῦρος (der Weizen), von den Datenbanken nicht von (τοῦ) πῦρος (des

33 Vgl. Schmitz 2004, S. 468.

34 Vgl. Ruffing 2012, S. 60.

35 Vgl. Robert Sallares: s. v. Getreide, in: Der Neue Pauly. Enzyklopädie der Antike. Band 4 (Epo–Gro), herausgegeben von Hubert Cancik und Helmuth Schneider, Stuttgart/Weimar 1998, Sp. 1031–1033.

36 Vgl. Kouli et al. 2009, S. 43–51. Ähnliches lässt sich auch für das Gebiet der Brauronebene sagen, da die dortigen Pollenproben grössere Getreideanbauaktivitäten lediglich in Mykenischer Zeit und dann wieder in Byzantinischer Zeit zeigen, vgl. Kouli 2012, S. 271f.

37 In klassischer Zeit steigt die Zahl der Olivenpollen und in Pistazienpollen deutlich an, vgl. Kouli 2012, S. 271, fig. 2 und S. 272, fig. 3 sowie insbesondere ihre Ausführungen zu VRA 3 auf S. 272.

Feuers) unterschieden wird und denselben Identifier trägt.³⁸ So kann man sich zwar mit Hilfe des Greek Word Study Tools³⁹ die Treffer für ὁ πυρός anzeigen lassen, indem in das Suchfeld ‚Puros‘ eingegeben wird⁴⁰ und dann im Kästchen für ebenjenes ὁ πυρός die Word Frequency Statistics ausgeführt wird.⁴¹ Doch wird aufgrund des gleichen Identifier in der Trefferspalte (Max) wiederum eine Zahl angegeben, die die Summe der möglichen Lexeme ist, jedoch eben nicht die Summe der Textstellen, in denen sich eine Variante von πυρός verbirgt.⁴² So findet sich bei Herodot beispielsweise die Zahl 61, die aber lediglich 13 Textstellen abbildet, von denen dann wiederum nur 7 tatsächlich auf eine deklinierte Form von ὁ πυρός entfallen.⁴³ Was für eine große Auswirkung diese Identifier-Problematik auf die quantitative Auswertung hat, zeigt Platons Dialog Timaios, da die 56 Treffer für πυρός dort ausschließlich dem Genitiv von πῦρ zuzuordnen sind und schlussendlich von den insgesamt 80 Treffern, die für ὁ πυρός in allen Platondialogen angezeigt werden, lediglich 6 der Bedeutung Weizen zugeordnet werden können.

Somit ist bereits für die quantitative Auswertung eine genaue Betrachtung jeder einzelnen Fundstelle nötig, um zu prüfen, ob es sich tatsächlich um eine Textstelle mit der Bedeutung Weizen oder einer aus dem Wortfeld Feuer handelt.

Nach Durchsicht aller Treffer ergibt sich daher folgendes Gesamtergebnis:

<i>πυρός</i>		<i>κριθή</i>		<i>σῖτος</i>	
Perseus	TLG	Perseus	TLG	Perseus	TLG
47	125	57	83	405	65
	172		140		470

Tab. 1: Gesamttrefferzahl für die jeweiligen Deklinationen der Begriffe, von fehlerhaften Treffern bereits bereinigt.

Die Auszählung wirft zwei Fragen auf. Zunächst, weshalb bei πυρός und κριθή die Autorensuche über den TLG, die ja nur als Ergänzung zu den bei Perseus digitalisierten Autoren hinzugezogen wurde, deutlich mehr Treffer liefert als die Perseus Digital Library, die den Großteil der Autoren und damit auch eine erheblich größere Textmasse abbildet. Dieser Umstand ist damit zu erklären, dass die beiden Werke von Theophrast *Περὶ φυτικῶν αἰτιῶν* und *Περὶ φυτῶν ἱστορία* bei Perseus nicht digitalisiert sind, sich aber intensiv mit Pflanzen und dementsprechend mit Weizen und Gerste auseinandersetzen. So resultieren 122 der 125 TLG-Treffer für πυρός und 79 der 83 TLG Treffer für κριθή aus den Schriften Theophrasts.⁴⁴

38 Dies ist freilich nicht die einzige Form, bei der es zu Überschneidungen kommt. Ein weiteres problematisches Beispiel sind die Überschneidungen der Deklination von κριθή und der Konjugation von κριτάω.

39 <http://www.perseus.tufts.edu/hopper/morph> (abgerufen am 08.07.2018).

40 <http://www.perseus.tufts.edu/hopper/morph?l=puros&la=greek> (abgerufen am 08.07.2018).

41 <http://www.perseus.tufts.edu/hopper/wordfreq?lang=greek&lookup=puro%2Fs> (abgerufen am 08.07.2018).

42 <http://www.perseus.tufts.edu/hopper/searchresults?q=puro/s&target=greek&doc=Perseus:text:1999.01.0125&expand=lemma&sort=docorder> (abgerufen am 08.07.2018).

43 Das ist darauf zurückzuführen, dass es eben nicht nur diese beiden Bedeutungsvarianten für πυρός gibt, sondern sich in der Deklination von ὁ πυρός noch weitere Überschneidungen verstecken, häufig mit einer konjugierten Form von πυρόω, anzünden.

44 Mit Ausnahme der *Ἡθικοὶ Χαρακτήρες*, da dieses Werk bei Perseus digitalisiert ist und in die Spalten der jeweiligen Perseustreffer Eingang fand.

Die zweite Frage bezieht sich auf $\sigma\acute{\iota}\tau\omicron\varsigma$, da hier die Trefferverteilung exakt umgekehrt ist, also $\sigma\acute{\iota}\tau\omicron\varsigma$ deutlich seltener vorkommt als $\pi\upsilon\rho\acute{o}\varsigma$ und $\kappa\rho\iota\theta\acute{\eta}$. Dieses Ergebnis ist darauf zurückzuführen, dass Theophrast klar zwischen beiden Getreidesorten trennt und den allgemeinen Begriff daher seltener verwendet. Doch ist dies nicht die einzige Erklärung, sondern führt direkt zur ersten Interpretation der Ergebnisse auf Grundlage der Gesamtzahlen. Es ist auffallend, dass bei den gängigen klassischen Autoren, die über die Perseus-Spalten erfasst sind, die Treffer für den allgemeinen Begriff $\sigma\acute{\iota}\tau\omicron\varsigma$ deutlich höher sind. Damit kann bereits festgehalten werden, dass die Corpora der Klassiker wie Platon und Aristoteles oder Thukydides und Herodot offensichtlich überwiegend die allgemeine Bezeichnung vorgezogen haben beziehungsweise diese für ihren Rezipienten als ausreichend einschätzten; sie also keinen besonderen Wert auf die genaue Unterscheidung der Getreidesorten legten.

Im Hinblick auf die Differenzierung von Gerste und Weizen fällt auf, dass Gerste im Verhältnis zu Weizen zwar im Corpus der klassischen Autoren der Perseus Digital Library häufiger genannt wird, jedoch lediglich mit einem Verhältnis von 55% zu 45%.

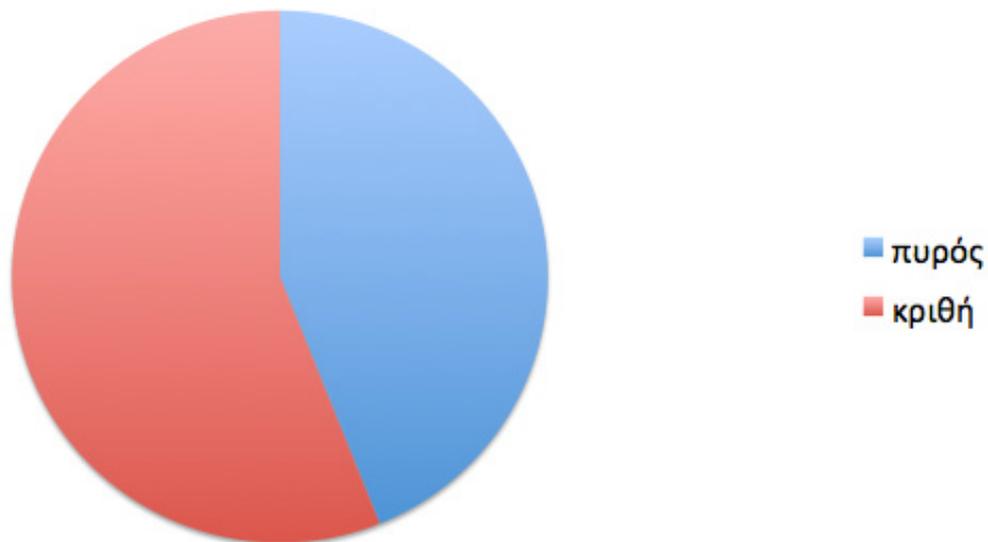


Abb. 1: Ein Tortendiagramm zum Verhältnis von $\pi\upsilon\rho\acute{o}\varsigma$ und $\kappa\rho\iota\theta\acute{\eta}$ bei den klassischen Autoren der Perseus Digital Library

Damit kann gerade im Hinblick auf die Überlieferungssituation bestenfalls von einer Tendenz in Richtung Gerste gesprochen werden, sodass sich das Verhältnis von 9,3 zu 1 aus IG II² 1672 zwischen Gerste und Weizen nicht in den Ergebnissen abbildet. Im Gegenteil sorgt die Einbeziehung der ergänzenden Autoren aus dem Corpus des TLG dafür, dass sich das Verhältnis zu Gunsten von Weizen auf 174 zu 143 verschiebt. Das lässt sich allerdings ausschließlich mit den Werken Theophrasts erklären, dessen botanische Werke an 43 Stellen mehr Weizen als Gerste behandeln, was im Hinblick auf die Vielfalt der Weizensorten aber nicht verwundert. Es lässt sich also nach Sichtung der Gesamttreffer bereits festhalten, dass die klassischen Autoren mit einem Bezug zu Athen überwiegend den allgemeinen Terminus $\sigma\acute{\iota}\tau\omicron\varsigma$ verwenden und sich bei der Verwendung von Weizen und Gerste – mit Ausnahme von Theophrast – eine leichte Tendenz zur Gerste erkennen lässt.

4. Qualitative Auswertung der Ergebnisse

Die Verteilung der Treffer auf die Autoren ergibt folgendes Bild:⁴⁵

Autor	πυρός		κριθή		σίτος	
	Perseus	TLG	Perseus	TLG	Perseus	TLG
Philosophen						
Anaxagoras		0		0		0
Platon	6		10		29	
Aristoteles	2		2		24	
Theophrast	1	122	1	79	0	56
Hippokrates	8		10		7	
Historiker						
Herodot	7		5		34	
Thukydides	1		1		47	
Xenophon	8		8		156	
Hellanikos von Lesbos		2		1		0
Ion von Chios		0		0		0
Damastes von Sigeion		0		0		0
Philochorus von Athen		1		3		4
Androtion von Athen		0		1		0
Daimachos von Plataiai		0		0		0
Ephoros von Kyme		0		0		1
Pherekydes von Athen		0		0		0
Attische Redner						
Antiphon		0		0		1
Andokides	0		0		3	
Lysias	0		0		11	
Isokrates	0		0		5	
Isaios	0		2		2	
Aischines	0		0		1	
Lykurgos	1		1		1	

⁴⁵ Bei Hdt. 4,128 sowie 5,34 gibt es den unregelmäßigen Plural σῖτα, τὰ,

Demosthe- nes	2		6		72	
Hypereides		0		0		0
Deinarchos	1		0		0	
Dramatiker						
Phrynichos		0		0		0
Aischylos	0		2		0	
Sophokles	0		0		0	
Euripides	0		0		8	
Aristopha- nes	11		9		4	
Alexandros		0		0		0
Dichter						
Artemon von Athen		0		0		0
Philetairos		0		0		0
Anakreon		0		0		0
Solon		0		0		0
Weitere						
Demetrios von Phale- ron		0		0		3
Ps. Xeno- phon	0		0		1	

Tab. 2: Zuordnung der Treffer zu den jeweiligen Autoren nach Bereinigung der Fehlergebnisse

Es fällt bereits auf den ersten Blick auf, dass sich die Verwendung der Begriffe nicht am Umfang der Werke orientiert und auch nur bedingt am Inhalt festzumachen ist. So verwendet beispielsweise Thukydides nur je einmal πυρός und κριθή jeweils im Akkusativ Plural, während Herodot 7 mal eine Form von πυρός und immerhin 4 mal eine Form von κριθή verwendet; Xenophon wiederum in seiner Ἀνάβασις 5 mal πυρός und 7 mal κριθή.⁴⁶ Beim allgemeineren σῖτος findet sich bei Herodot wiederum nur 34 mal eine entsprechende Form, während sie bei Thukydides mit 47 Belegstellen wesentlich häufiger zu finden ist und gerade im Verhältnis zu den spezifischen Begriffen für Gerste und Weizen heraussticht. Die 156 Belegstellen für σῖτος bei Xenophon reduzieren sich wiederum auf lediglich 61, wenn man sich auf die historischen Werke Ἑλληνικά, Ἀνάβασις und Ἀγησίλαος konzentriert. Herodot und Xenophon nutzen also, wenn der Bedarf besteht, durchaus die spezifischen Varianten, während Thukydides diese Differenzierung nur ein einziges Mal vornimmt, wobei hervorzuheben ist, dass beide Begriffe bei Thuydides an gleicher Stelle verwendet werden:

⁴⁶ Die verbliebenen Treffer entfallen auf den Οἰκονομικός und ein weiterer auf Weizen in der Kyropädie.

ὀπλίτας τε οὖν πολλοὺς μοι δοκεῖ χρῆναι ἡμᾶς ἄγειν καὶ ἡμῶν αὐτῶν καὶ τῶν ξυμμάχων, τῶν τε ὑπηκόων καὶ ἦν τινα ἐκ Πελοποννήσου δυνώμεθα ἢ πείσαι ἢ μισθῷ προσαγαγέσθαι, καὶ τοξότας πολλοὺς καὶ σφενδοθήτας, ὅπως πρὸς τὸ ἐκείνων ἰππικὸν ἀντέχῃσι, ναυσί τε καὶ πολὺ περιεῖναι, ἵνα καὶ τὰ ἐπιτήδεια ῥᾶον ἐσκομιζώμεθα, τὸν δὲ καὶ αὐτόθεν σῖτον ἐν ὀλκάσι, **πυροὺς** καὶ πεφρυγμένας **κριθάς**, ἄγειν, [...]

„Schwerbewaffnete müssen wir meines Erachtens in großer Zahl mitführen, aus unserer Mitte und den Bundesgenossen, von unseren Untertanen und wenn wir Leute aus dem Peloponnes entweder dazu überreden oder durch Geld auf unsere Seite ziehen können, dazu noch Bogenschützen in großer Zahl und Schleuderer, um der Reiterei dort das gegengewicht zu halten. An Schiffen müssen wir ihnen weit überlegen sein, um die Lebensmittel leichter herbeischaffen zu können, auch von hier müssen wir Getreide, Weizen und gedörrte Gerste, in Lastschiffen mitführen [...].“

(Thuk. 6,22; Übers. Vretska 2009)

Der Kontext dieser Passage ist die Debatte über die Sizilienexpedition im Sommer 415 v. Chr. Nachdem es Alkibiades gemäß Thukydides geschafft hatte, die Stimmung der Volksversammlung durch seine Rede für eine Sizilienexpedition zu seinen Gunsten zu beeinflussen, sah der expeditionsskeptische Nikias seine letzte Chance darin, durch das Aufzeigen der gewaltigen Investitionen auf verschiedenen Ebenen die Versammlung abzuschrecken.⁴⁷ In diesem Kontext ist die Stelle zu betrachten und einzuordnen. Thukydides lässt also Nikias über diese Differenzierung des Getreides sprechen, die er als Apposition zu σῖτον, also auch grammatisch als eine Spezifizierung verwendet. Es lässt sich hier festhalten, dass Thukydides der Differenzierung der Getreidesorten und damit auch ihrer unterschiedlichen Beschaffungsmöglichkeiten in Attika keine größere Bedeutung beimisst. Bemerkenswert ist jedoch, dass es für Thukydides‘ Nikias offensichtlich kein Argument ist, dass es schwieriger sei, Weizen für die Expedition zu bekommen, da dieser generell in Attika schlechter gedeihe als Gerste und erst importiert werden müsse. Das könnte dafür sprechen, dass der Getreideimport für Attika bereits in klassischer Zeit so umfangreich war, dass ohnehin beide Getreidesorten in recht großer Menge importiert wurden.⁴⁸ In diesem Fall wäre das ein weiteres Indiz für die Position der Gruppe um Jardé, Whitby und Morley, die dem Getreideimport im Gegensatz zu Garnsey, Sallares und Gallant eher eine größere Bedeutung zuspricht.

Wenn man nun einen Blick auf die Treffer bei Herodot wirft, zeigt sich ebenfalls eine deutliche Tendenz. Es lässt sich für die πυρός-Belegstellen zunächst feststellen, dass diese überwiegend in den Exkursen der ersten Werkhälfte zu finden sind: 1,193,3; 2,36,2; 3,22,4 4,33,1; 4; 5; 7,31 und 7,187,2. In den ersten beiden Belegstellen beschreibt Herodot die Getreideanbaugewohnheiten der Babylonier bzw. der Ägypter. In 3,22,4 lässt er wiederum den Perserkönig erklären, dass er nur Brot aus Weizen esse, und die dreifache Belegstelle in Kapitel 4,33 behandelt den Ritus der Hyperboreer, ihre Opfer in Weizenstroh einzuwickeln. Bemerkenswert ist an dieser Stelle der Hinweis in 4,33,5, dass dies auch Thrakerinnen täten, womit deutlich wird, dass es sich hierbei

47 Vgl. Thuk. 6,19. Ergänzend hierzu auch Hornblower, CoT III, 353f.

48 Vgl. hierzu auch Dalby 1992, S. 25, der insbesondere in FN. 66 auf die Preissteigerungen durch Händler und Städte verweist, vgl. auch Aristot. Oik. 2,7.

um eine für Griechen des Mutterlands zwar ungewöhnliche, aber im Norden der griechischen Kontaktzone – aus der auch ein beträchtlicher Teil des Weizens importiert wurde – durchaus übliche Opferhandlung handelte.⁴⁹ Vergleichbare kultische Handlungen mit Weizenstroh sind bezeichnenderweise für Athen nicht belegt, was keine lange Tradition des Weizenanbaus in Attika vermuten lässt.

Die Herodotstellen 7,31 und 7,187,2 befinden sich zwar bereits im Kontext des Feldzugs des Xerxes gegen Griechenland, doch beschreibt die erste Belegstelle lediglich Frauen aus Kallatebos, die Weizensirup ‚geschäftsmäßig‘ herstellten. 7,187,2 behandelt die Vorratskalkulationen seitens Xerxes‘, der für jeden Kämpfer pro Tag eine Choinix Weizen zugeteilt habe, was nach Herodots Rechnungen täglich 110.340 attische Medimnen gewesen sind, also ca. 5.792.850 Liter Weizen pro Tag. Herodot bezweifelt diese Menge jedoch selbst, was gleichzeitig als weiterer Beleg dafür hinzugezogen werden kann, dass die von Herodot kolportierte Zahl an Kämpfern auf persischer Seite deutlich über der tatsächlichen Zahl lag – zumal ihm an dieser Stelle selbst ein Rechenfehler unterläuft.⁵⁰

Auffallend ist jedoch, dass Herodot ausschließlich von Weizen spricht und nicht von Gerste, während bei den vorher genannten Textstellen zu Babylonien (1,193,3) und Ägypten (2,36,2) – beides Kornkammern des persischen Reichs – noch explizit beide Sorten genannt wurden. Somit dürfte sich auch die Belegstelle 7,187,2 in die Reihe derjenigen Passagen einreihen, die in Herodots Werk als Leitmotiv gelten, in denen auf die Hybris stets ein fundamentales Scheitern folgt.⁵¹ Denn nicht nur die Masse an sich, sondern auch die Absicht, nur Weizen für den Feldzug einzuplanen und auf die günstigere Gerste zu verzichten, dürfte das Bild des überheblichen Xerxes mitgezeichnet haben – wo doch sogar für die ebenfalls nicht unüberhebliche Sizilienexpedition durch Nikias gleichermaßen Weizen und Gerste einkalkuliert wurden.⁵²

Von den Belegstellen für κριθή bei Herodot wurden bereits zwei genannt (1,193,3; 2,36,2), da sie sowohl Weizen als auch Gerste erwähnen. Hinzu kommen noch die Stellen 1,160,5 und 2,77,4. In der Stelle 1,160,5 befasst sich Herodot mit der Unterwerfung der Lyder und dem Umstand, dass es in Folge der Unterwerfung und eines Frevels⁵³ sehr lange gedauert habe, bis ein Bürger überhaupt wieder Feldfrüchte und eben Gerstenkörner einem Gott hätte opfern können.⁵⁴ 2,77,4 ist wiederum eine Passage, in der Herodot berichtet, dass die Ägypter Bier aus Gerste brauen würden.

49 Hinsichtlich der kultischen Verbindung der Hyperboreer zu Apollon und der Insel Delos sowie zu den Thrakern, vgl. How / Wells 1989, S. 309 f. sowie Asheri et al. 2007, S. 605.

50 Vgl. ausführlich hierzu How / Wells 2002, S. 213f.

51 Vgl. Raafflaub 2016, 601f.

52 Vgl. Thuk. 6,22.

53 Vgl. Hdt. 1,159,3f.

54 Hdt 1,160,5: ἦν δὲ χρόνος οὗτος οὐκ ὀλίγος γινόμενος, ὅτε Χίων οὐδεὶς ἐκ τοῦ Ἀταρνέος τούτου οὔτε οὐλὰς κριθέων πρόχρυσιν ἐποιέετο θεῶν οὐδενὶ οὔτε πέμματα ἐπέσσετο καρποῦ τοῦ ἐνθεῦτεν, ἀπειχετο τε τῶν πάντων ἰρῶν τὰ πάντα ἐκ τῆς χώρας ταύτης γινόμενα. – „Bei den Chiern aber verging lange Zeit, ehe irgendein Bürger von dem Ertrage der Feldfrüchte in Atarneus einem Gott Gerstenkörner streute oder einen Opferkuchen buk. Alles, was aus diesem Lande stammte, wurde von Opfern ferngehalten.“ (Übers. Feix 1995).

Eine Auffälligkeit, die bereits bei Thukydides zu beobachten war, ist die gemeinsame Nennung von Gerste und Weizen in den jeweils gleichen Textstellen. Es stellt sich also die Frage, ob eine solche Dopplung in den Belegstellen häufiger vorkommt. Hier gibt die Kookkurrenzanalyse von eAQUA eine zielführende Möglichkeit zur Visualisierung. Zwar lässt sich die Kookkurrenzanalyse nicht auf einzelne Autoren beschränken, doch wenn man sich bei der Auswahl auf das Corpus der Perseus Digital Library beschränkt, so bekommt man aufgrund der guten Überlieferungslage der klassischen Autoren aus dem Umfeld Athens auch ohne eine temporale oder regionale Einschränkungsmöglichkeit des Tools einen ersten Überblick über die Kookkurrenzen in klassischer Zeit. Hierfür müssen allerdings die eindeutigen Formen ausgewählt werden, die in möglichst geringem Maße durch die in Kapitel 3 beschriebenen Doppelbedeutungen bei gleicher ID korrumpiert werden. Geeignete Formen sind hier der Genitiv und der Akkusativ Plural, die zwar auch jeweils andere Bedeutungen haben können,⁵⁵ allerdings in den geprüften Ergebnissen nicht vorkamen:



Abb. 2: Signifikante Kookkurrenzen für πυρῶν, Wort-ID 9211 auf Grundlage der Perseus Digital Library

Von den 102 Treffern, die sich für πυρῶν mit der Wort-ID 9211 ergeben, entfallen 26 (28) auf den Genitiv Plural von Gerste, nämlich κριθῶν sowie 4 weitere auf die abgewandelte Form κριθέων. Damit kookkurriert mehr als jede vierte Belegstelle mit dieser KNG-kongruenten Form von κριθῆ, wobei bemerkenswert ist, dass die linken Kookkurrenzen mit 9 deutlich geringer ausfallen als die rechten Kookkurrenzen mit 19:⁵⁶

55 Vgl. hierzu <http://www.perseus.tufts.edu/hopper/morph?l=purwn&la=greek> (zuletzt abgerufen am 08.07.2018) für den Genitiv Plural und <http://www.perseus.tufts.edu/hopper/morph?l=πυρούς&la=greek> (zuletzt abgerufen am 08.07.2018) für den Akkusativ Plural.

56 Die Zahlen 26 bei den signifikanten und 9 zu den linken sowie 19 zu den rechten Kookkurrenzen (zusammen 28) sind darauf zurückzuführen, dass nur die Sätze an sich gezählt werden, in denen Kookkurrenzen vorkommen. Ein Satz mit beiden Arten wird bei den signifikanten Kookkurrenzen nur einmal und einmal sowohl für die linken als auch für die rechten Kookkurrenzen gezählt. Mehrfachkookkurrenzen werden dabei nicht berücksichtigt. Würde man dies tun, käme man auf 10 linke und 22 rechte Kookkurrenzen für κριθῶν, die sich aber auf nur 26 Sätze verteilen.

Signifikante linke Kookkurrenzen:	<p>κριθῶν (9); μεδίμνουσ (7); ἐνδεδεμένα (2); μέδιμνός (2); μυρίουσ (5); μέδιμνον (3); μάκαροσ (2); ἀμῆτηρεσ (2); μέτρημα (2); ἐλαύνωσι (2); Σικελικόσ (2); ἀλφιτα (3); καλάμη (2); μεδίμνου (2); ὄμνον (2); μέτρου (3); ἐλαίου (3); κοίνικα (2); δραχμῆσ (2); δραχμῶν (3); ἀνθρωπιείαν (2); δραχμάσ (3); κτηνῶν (2); ἄροι (2); ἄρουραν (2); ἰματίδιον (1); σιτανίω (1); Πρασιεῦσιν (1); περιφέρωσι (1); νωτοφορίαν (1); φασήλων (1); ἀντηλλάξαντο (1); γεφυρίζω (1); ἱεροφάντιδι (1); σιτομετροῦνται (1); εὐογκότερα (1); συνερρηκῶτα (1); ἐσβάλλουσι (1); Ἀριμασποί (1); σπανίων (1); Ἀρχίμηλον (1); πυρριχίζω (1); θυσίμων (1); κρεοφαγούμεναι (1); ἀρτοπώλαι (1); Πρασιάσ (1); ἡττημέναισ (1); ἀφευε (1); ἐνελημένα (1); ΜαρτυρίαΝόμοσ (1); μιλωθροί (1); ἡμιυφάντιοσ (1); κροτήσειν (1); ἀμφορεῦσι (1); σιτηγούντεσ (1); ξυλοκοπεῖ (1); κκεχειμακότασ (1); κατετρυκέτην (1); λελοιπότασ (1); σκευαζόμενοι (1); καταλάμμεται (1); διαπεφευγία (1); Ἐκπεμφθεῖσ (1); πεφυρημένω (1); σκιαμαχοῦντα (1); ἐξηρέθιζεν (1); χλοίωσ (1); ὑπορρηγῆναι (1); δίνω (1); Λέκοσ (1); διτύλων (1); Ὑπερβορείοσι (1); ἰαμβοποῖον (1); πρόσοδοτε (1); ὄνωϊδασ (1); ῥίπιοσ (1); ἀπίστων (1); συγκομιστόσ (1); ἀνατετακότων (1); σαργνάσιοσ (1); προφανέστατον (1); Παιονίδασ (1); προδιδαξάντων (1); Ἰσηδόσι (1); ὕδρευσόμενον (1); πωλήσουσιν (1); δοσηγερυσ (2); ἀργυρίου (3); ἰρᾶ (2); ὑποδήματα (2); σίτοσ (2); ἐναντίοι (2); ἀλλήλοισιν (2); πυρῶν (2); κτήσεισ (2); ἀποστέλλοντασ (1); ἀρότιό (1); Ἰσειόσι (1); ἀποδέκεσθαι (1); ἀποβαῖεν (1);</p>
Signifikante rechte Kookkurrenzen:	<p>κριθῶν (19); μεδίμνουσ (8); μεδίμνων (4); οἶνου (7); κριθῶν (3); ἔπεσσο (2); μυκαίτατα (2); μετρητήσ (2); Ἀλεξανδρεινῶν (2); ἐξηγητικόσ (2); ἀσπαστων (2); μέδιμνον (3); ἄρτον (4); κατακλείοντεσ (2); μῶνον (2); ἡμικεκτον (2); ἀνέργαστον (2); ὄνωιοσ (2); δῆου (2); ἔριφοσ (2); καρπόν (3); θορόντεσ (2); πενέστεροι (2); κριθῆσ (2); ἄρτουσ (3); μύρτων (2); πλουσιώτεροι (2); λαγῶσ (2); ταρφέα (2); δράγματα (2); τυγχάνοιεν (2); φόβοιο (2); ὀβολοῦ (2); πλησιοχώρουσ (2); κυάμων (2); ὀβολῶν (2); ὄλοοῖο (2); δραχμῆσ (2); κριθᾶσ (2); ἥσθιο (2); διατιθέσθων (1); ἀγγελῆναι (1); πλάσσοντεσ (1); παρορώμεναι (1); ἀλευρίται (1); ἐμφέρηται (1); πολυτροφώτεροι (1); συγκομιστοί (1); βωστρησάτω (1); καταλέσαντεσ (1); ἱεροφάντιδι (1); ἀποκρῶντεσ (1); πυρριχίζω (1); λαγαρά (1); ἀφύρητοσ (1); πράσιοσ (1); προσαιτούση (1); κριθίνουσ (1); διεμετρήσαμεν (1); πεφυρημένοσ (1); ξυνθέτων (1); μάννην (1); κνήκου (1); καταφανέντων (1); φυρᾶν (1); ἀνατεινάντασ (1); ἔνιοι (1); πρασσοκουρίσ (1); ἰσόκριθον (1); σῖραν (1); ξυστίδοσ (1); ἱππαστρίασ (1); ἐγκρυφία (1); θυλακίσι (1); ἡδυσῶμων (1); διαπονούμεβά (1); δεδευμένων (1); προσεξυτιονσ (1); παρακναῦσαι (1); δεκομένουσ (1); ἡμαιο (1); ἐκέρασαν (1); ἀντιδωρησαμένην (1); Δωδωναίουσ (1); συνεκδιδόντεσ (1); ἔνωμοσ (1); κηδεύουσι (1); νωτοφοροῦσιν (1); κοδράτουσ (1); μίλακί (1); ἀνάκωλο (1); εὐκονον (1); πέφαντεσ (1); μεδίμνοισ (1); παχύσ (1); ἀπυρέτω (1); πομπείω (1); χᾶλα (1); σεμιδαλίται (1); ὀρόβω (1);</p>

Abb. 3: Linke und rechte Kookkurrenzen für πυρῶν, Wort-ID 9211 auf Grundlage der Perseus Digital Library

Zunächst ist festzuhalten, dass bei der Nennung beider Getreidesorten in einem Satz πυρός häufiger zuerst genannt wird. Das könnte darauf hindeuten, dass dem Weizen seitens der Autoren indirekt eine größere Bedeutung zugemessen wurde, was die These unterstützt, dass Weizen durchaus ein höher angesehenes Getreide war.⁵⁷ Ebenfalls zu beobachten ist, dass der Akkusativ von μέδιμνος, also die Maßeinheit für Getreide 1921 mal (15 mal im Plural, 6 mal im Singular) in größerer Häufigkeit in näherer Umgebung auftaucht. Auch das ist plausibel, da die Maßeinheit als Akkusativobjekt und die jeweiligen Genitive als deren Attribute fungieren. Dies zeigt sich auch durch die Visualisierungs-Funktion der Kookkurrenzanalyse:

57 Vgl. hierzu auch Athen. 137e, wo beschrieben wird, dass Solon ein Gesetz erlassen habe, dass die aus öffentlichen Geldern finanzierten Speisungen im Prytaneion lediglich an Festtagen Speisen aus Weizen sein sollten. An allen übrigen Tagen wurden Mahlzeiten aus Gerste, vermutlich der verbreitete Gerstenbrei, gereicht.

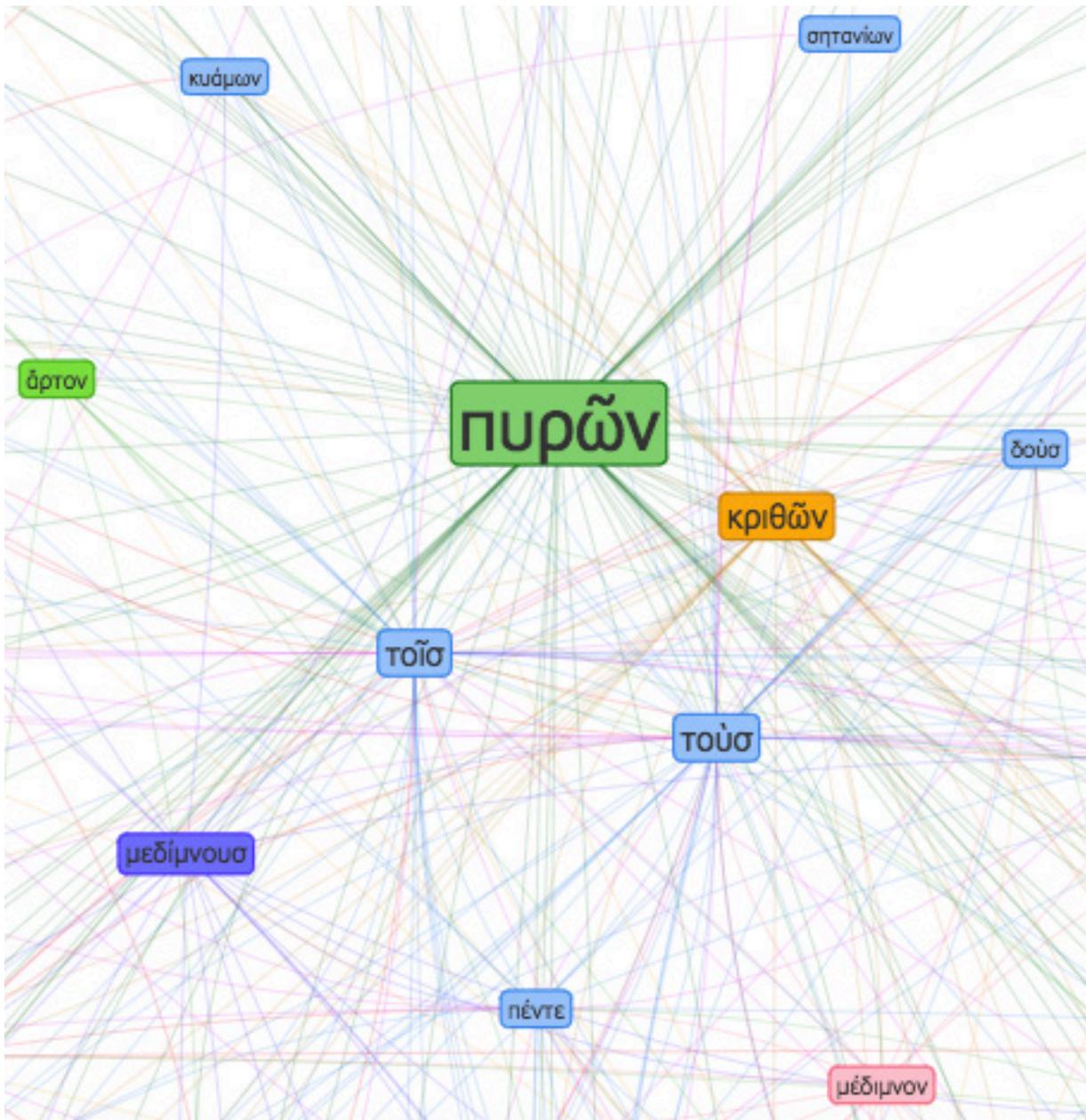


Abb. 4: Visualisierungs-Funktion Network [Log-Likelihood] für die Kookkurrenzen für πυρῶν⁵⁸

Bereits in dieser Abbildung wird ohne jegliche Hervorhebungen der Zusammenhang der genannten Wörter deutlich und bildet die Kookkurrenz auch graphisch ab. Unterstrichen wird dies, wenn man umgekehrt die Form κριθῶν markiert:

58 Die Visualisierungs-Funktion Network [Log-Likelihood] verwendet aus computerlinguistischen Gründen nur ein einheitliches Zeichen für Sigma, sodass in der Visualisierung auch am Wortende das gewöhnliche Sigma auftaucht.

Beim Akkusativ Plural von κριθή besteht mit 11 Kookkurrenzen die mit Abstand größte Verbindung zwischen den KNG-kongruenten Formen von Weizen und Gerste. Und auch die Positionierung innerhalb des Satzes bestätigt sich:

Signifikante rechte
Kookkurrenzen:

κριθάς (5); κριθᾶσ (4); βριθῆσι (2); ὄσπρια (2); ἄλω (2); ἔμπεδα (2); παρέχη (2); πυρούσ (2); δένδρεα (2); πίσον (1); πεφρυγμένασ (1); μυλώνων (1); καρπῶ,τίκτη (1); ἠπεροπεύσασ (1); φυτουργία (1); διακομιζόμενοι (1); δοθῆσιν (1); διαφοραί (1); κεραμίουκαί (1); διακρεόμενοι (1); ἐμμίσθους (1); μεταβάλλοντάσ (1); κηπευόμενα (1); νήττη (1); ἔρηνε (1); φυραθῶσιν (1); μελίνας (1); ἐπιθειῆσ (1); εὐηγείσῃσ (1); ξυσσίτους (1); μασσαάμενοσ (1); ὑπεστήριξε (1); παράδειγμαδύναι (1); ἄρακα (1); φυτώρια (1); προβαλοῦσιν (1); μελιτοῦντασ (1); ἀπολαμβάνωμεθα (1); ἀνθητικᾶ (1); συμπέττοντασ (1); ἐνιστάσιν (1); μεταλαμβάνομένησ (1); φαληρίδι (1); ἀπέποισ (1); λεπτομερεστέρου (1); νασοῦσ (1); κεράμι (1); σῦκα (2); ἄρτοισιν (1); ὄρχιλοσ (1); παγέντεσ (1); ἠδουίνουσ (1); ξυλλεξάμενοσ (1); εὐξόμεθ (1); λιλαιόμενοσ (1); ἔρμαιου (1); συναίρειν (1); ἐσφενδόνησαν (1); κριθᾶσ (1); καθαγίζειν (1); μελίνας (1); σέρφον (1); θερῖση (1); ἐθέριζον (1); ἀνατείνοντεσ (1); ἀρετώσι (1); διαμετρῶν (1); διαπυθόμενοσ (1); ἀγνωσίασ (1); ὀλκή (1); Ἡρακλέει (1); κόπτονται (1); ἐνόρχην (1); ἀντέκωσιν (1); κατερρήγγυντο (1); λάρω (1); ἔρκετ (1); μήλα (2); κρέα (2); ἐπέσασεν (1); πρίαο (1); σφαγιάζειν (1); θερείασ (1); κέγχρουσ (1); βρωτόν (1); αὐξήση (1); προγένοντο (1); ἀμενηνῶσ (1); ἐστερημένουσ (1); σιτοποιοῦσ (1); θάλασσα (2); δανείση (1); διαπέμποντεσ (1); ἠγακασμένουσ (1); ἐσπίπουσιν (1); τίκτη (1); βάτραχοσ (1); εὐραντο (1); γαστέρασ (1); ἀργούσ (1);

Abb. 7: Rechte Kookkurrenzen für πυρούς, Wort-ID 23267 auf Grundlage der Perseus Digital Library

Bei den Akkusativformen im Plural zeichnet sich ebenso deutlich die Tendenz dahingehend ab, dass zuerst πυρούς und erst dann die KNG-kongruente Form κριθᾶσ genannt wird, nämlich bei 9 von 11 Belegstellen.⁵⁹ Wenn man die Kookkurrenzanalyse bereinigt, ergibt sich bei den Perseus-Treffern, also alle relevanten Autoren der klassischen Zeit, sogar eine noch größere Überschneidung. So kommen auf 47 Treffer aller Kasus mit πυρός 22 Treffer, bei denen auch eine Form von κριθή angeführt wird. Somit findet sich bei fast jeder zweiten Belegstelle für πυρός innerhalb des Satzes auch eine Form von κριθή.⁶⁰

Somit lässt sich bereits an dieser Stelle festhalten, dass die Ergebnisse der Kookkurrenzanalyse zeigen, dass der antithetische Ansatz des Artikels, der auf einem ‚Entweder Weizen oder Gerste‘ aufbaut, deutlich eingeschränkt werden muss, da insgesamt in mehr als jedem vierten Fall die entsprechenden Formen von Gerste und Weizen ohnehin zusammen verwendet werden und bei der händischen Auszählung der Beispiele sogar in fast jedem zweiten. Darüber hinaus zeichnet sich aufgrund der Wortstellung eine größere Gewichtung des Weizens gegenüber der Gerste ab, da jener sowohl im Genitiv Plural als auch im Akkusativ Plural in der Regel vor der Gerste genannt wird. Dieser Befund lässt die Vermutung zu, dass sich die größere Nachfrage des schwieriger zu beschaffenden Weizens gegenüber der Gerste auch in der Syntax der antiken Texte abzeichnet.

5. Auswertung der übrigen Ergebnisse

Die Philosophen

Ein ähnlich heterogenes Bild wie bei den Historikern ergibt sich auch bei den Philosophen. Während sich bei Anaxagoras mit Hilfe des TLGs keine Belegstellen finden ließen, war Theophrast mit seinen Ἠθικοὶ Χαρακτῆρες und je einer Erwähnung sowie das Corpus

⁵⁹ Diese teilen sich auf, da es durch den unterschiedlichen Akzent und den Unterschied der Schreibweise beim Sigma an letzter Stelle zwei IDs für dasselbe Wort gibt.

⁶⁰ Vgl. Tab. 3 „Bereinigte Trefferliste“ unter Zusatzdateien in DCO 4,2 (2018).

Aristotelicum mit je 2 Belegstellen bei Weizen und Gerste ausgeglichen.⁶¹ Das Corpus Aristotelicum verzeichnet jedoch beim allgemeinen σῖτος eine große Zahl von 24 Belegstellen, wobei dieser allgemeinere Begriff 5 Mal in der Ἀθηναίων πολιτεία und 13 Mal im Οἰκονομικός verwendet wurde.

Jedoch bietet besonders Theophrast hier bedeutende Textstellen, da er in diesen Gerste und Weizen für die Charakterisierung verschiedener Charaktere nutzt. So verwendet er Weizen für ‚den‘ Redseligen:

εἶτα δὴ προχωροῦντος τοῦ πράγματος λέγειν, ὡς πολὺ πονηρότεροί εἰσιν οἱ νῦν ἄνθρωποι τῶν ἀρχαίων, καὶ ὡς ἄξιοι γεγόνασιν οἱ πυροὶ ἐν τῇ ἀγορᾷ, καὶ ὡς πολλοὶ ἐπιδημοῦσι ξένοι, καὶ τὴν θάλατταν ἐκ Διονυσίων πλόμιον εἶναι.

„Dann spricht er, langsam in Fahrt kommend, davon, wieviel schlechter die Menschen heute sind als früher, wie billig der Weizen auf dem Markt ist, wie viele Fremde in der Stadt sind, und das Meer sei seit den Dionysien schiffbar.“

(Theophr. char. 3,3, Übers. Klose 1988)

Um die Passage und damit den Einsatz des Weizens für eine Charakterisierung deuten zu können, müssen zunächst die anderen Elemente dieser Aufzählung bzw. das ‚in Fahrt kommend‘ näher betrachtet werden. Denn die Aussage, Weizen sei zu jener Zeit auf dem Markt günstig zu haben, lässt sich auf verschiedene Weise interpretieren. Dies kann entweder als Allgemeinplatz gelten, wenn es der Realität entsprach oder als übertreibende Behauptung, da Weizen als Importgut oder als unter den widrigen klimatischen Bedingungen in Attika angebautes und Teuerungen ausgesetztes Getreide selten günstig gewesen sein dürfte. Allerdings ist letzterer Aspekt nicht zu erwarten, da er eine solche Verhaltensweise in char. 7 mit einem eigenen Charakter, dem Schwätzer, behandelt hat. Es dürfte sich daher eher um Behauptungen handeln, die Allgemeinplätze sind.⁶² Dafür sprechen auch die übrigen Glieder der Aufzählung, dass die Menschen schlechter geworden seien – eine bis heute gängige Einschätzung der Zeitgenossen –, viele Fremde in der Stadt seien, was auch Mitte des 4. Jahrhunderts durch Xenophons πόροι (Über die Staatseinkünfte. Möglichkeiten der Geldbeschaffung für den Stadtstaat Athen) nachvollziehbar ist, da dort der Gewinn aus eben jenen Metöken ein zentraler Ansatz des Konzepts ist.⁶³ Und auch die Feststellung, dass die Seefahrt seit den Dionysien wieder möglich sei, also seit Ende März/Anfang April, klingt wegen der Winterruhe in der antiken Schifffahrt nach einer recht banalen Information.⁶⁴ Somit muss zumindest für einen gewissen Zeitraum vor der Abfassung der Ἡθικοὶ Χαρακτῆρες ca. 319 v. Chr.⁶⁵ der Weizenpreis recht günstig gewesen sein.

61 Hierbei entfielen sowohl für Gerste als auch für Weizen die Treffer je einmal auf die Ἀθηναίων πολιτεία sowie den Οἰκονομικός. Bemerkenswert ist jedoch, dass Aristoteles bei der gemeinsamen Nennung die Gerste vor den Weizen zieht: Aristot. oik. 2, 1347a: [...] τῆς ἐν ἀκροπόλει ὑπὲρ τοῦ ἀποθανόντος φέρειν χοίνικα κριθῶν καὶ πυρῶν ἐτέραν καὶ ὀβολόν, [...].

62 Vgl. Diggle 2009, S. 199.

63 Vgl. Xen. vect. 2.

64 Vgl. hierzu auch Ussher 1993, S. 52.

65 Vgl. Ussher 1993, S. 13.

Die Gerste wird wiederum bei der Charakterisierung des Unverschämten erwähnt:

καὶ ἐπὶ τὴν ἀλλοτρίαν οἰκίαν ἐλθὼν δανείζεσθαι κριθάς, ποτὲ δὲ ἄχυρα, καὶ ταῦτα τοὺς χρήσαντας ἀναγκάσαι ἀποφέρειν πρὸς αὐτόν.

„Er geht in ein fremdes Haus, borgt sich Gerste, ein andermal Stroh⁶⁶ und zwingt die, die es ihm geliehen haben, es sich selbst zurückzuholen.“

(Theophr. char. 9,7, Übers. Klose 1988)

Auch bei dieser Textstelle ist die Interpretation nicht so eindeutig, da diejenigen, die die Gerste verleihen, nicht weiter beschrieben werden. Wenn man allerdings die Annahme zu Grunde legt, dass Gerste in der Antike nicht nur wegen der mangelnden treibenden Wirkung beim Backen, sondern auch generell als weniger nährstoffreich angesehen wurde,⁶⁷ so ist die Gerste hier als ein Mittel der Charakterisierung der Personen des ‚fremden Hauses‘ anzusehen.⁶⁸ Hinzu kommt der bereits erläuterte Aspekt, dass zur Zeit der Abfassung der *Ἡθικοὶ Χαρακτῆρες* der Weizenpreis im Piraeus offensichtlich häufiger niedrig war, sodass wohl nur die ärmsten Athener weiterhin auf Gerste zurückgriffen.⁶⁹ Dadurch sind die leihenden Personen als ohnehin eher ärmere Bürger zu betrachten, die dann noch selbst dafür sorgen mussten, das Geliehene zurückzuerhalten, während sich ‚der Unverschämte‘ über die Lage der Beliehenen hinwegsetzt und damit eben jene Unverschämtheit noch vergrößert.⁷⁰

Mit diesen Textstellen gibt Theophrast in seinen Charakterisierungen wichtige Hinweise zur Bedeutung von Gerste und Weizen in spätklassischer oder frühhellenistischer Zeit. Denn die Weizenversorgung in Athen scheint so stabil gewesen zu sein, dass der günstige Preis zwar dem Redseligen noch als Erzählstoff würdig erschien, im Grunde aber schon ein Allgemeinplatz geworden war und die Gerste offensichtlich als indirekte Charakterisierung ärmerer Leute immer noch taugte.

Bei Platon finden sich wiederum mehr Belegstellen für Gerste als für Weizen: 6 Belegstellen für Weizen stehen 10 für Gerste und 29 für Getreide allgemein gegenüber. Dabei stammen die Treffer zum großen Teil aus den *Νόμοι* sowie der *Πολιτεία*. Interessant ist auch hier ein Blick auf die Wortstellung von *πυρός* und *κριθή*. Denn die vier Stellen, an denen beide Bedeutungen vorkommen, teilen sich je zwei Mal auf die *Νόμοι* und die *Πολιτεία* auf. Und in beiden Werken steht je einmal die Form von *πυρός* (nom. 847e und pol. 363c) und je einmal die Form von *κριθή* an erster Stelle (nom. 849c und pol. 372b). Doch fällt bei einem genaueren Blick auf die Textstellen auf, dass bei einer einfachen Aufzählung mit *καὶ* jeweils *πυρός* vorne steht, während die anderen beiden Belegstellen lediglich durch die Konjunktion *ἢ* verbunden werden bzw. sogar ganz in zwei verschiedene Nebensätze gleichen Grades aufgeteilt sind. Somit scheint sich bei genauerer Betrachtung der Ergebnisse der Eindruck weiter zu bestätigen, dass bei gemeinsamer, positiver Nennung meist die Form von *πυρός* zuerst genannt wird.

66 Stroh ist hier etwas irreführend, da es sich bei *ἄχυρα* gemäß LSJ um die Abfallprodukte des Dreschens handelt, also Spreu (chaff), Kleie (bran) und Hülse (husk), vgl. LSJ s.v. *ἄχυρον*.

67 Vgl. Schmitz 2004, S. 467.

68 Diggle 2009, S. 298, sieht hier Gerste analog zu *ἄχυρα* nur als Tierfutter an.

69 Vgl. hierzu auch die Einschätzung von Schmitz 2004, S. 468, dass in spätklassischer Zeit vor allem Sklaven mit Gerste ernährt wurden.

70 Vgl. hierzu Ussher 1993, S. 101.

Eine ähnliche Aufteilung findet sich auch beim letzten Opus der hinzugezogenen Philosophen, dem *Corpus Hippocraticum*. Hier zeigt sich allerdings eine deutliche Abweichung der bisherigen Ergebnisse. So stehen 8 Erwähnungen von Weizen 10 von Gerste gegenüber, während Getreide im Allgemeinen entgegen den bisher behandelten Autoren mit 7 Belegstellen zumindest im Verhältnis zu jenen Autoren stark abfällt. Doch auch hier liegt eine ähnliche Erklärung wie bei Theophrasts botanischen Werken zu Grunde. Da die Schriften des *Corpus Hippocraticum* sich vor allem mit medizinischen Aspekten auseinandersetzen und die Beschreibung, Kategorisierung und Linderung von Krankheiten niedergeschrieben wurde, ist es naheliegend, dass die Bezeichnungen auch wesentlich präziser sind und demnach die konkreten Begriffe πυρός und κριθή verwendet werden – sowohl im Verhältnis aber eben auch in absoluten Zahlen. Auffallend ist zudem, dass Gerste mit einem Verhältnis von 10 zu 8 etwas häufiger als Weizen genannt wird und somit im medizinischen Kontext präsenter war. Dieses Verhältnis dürfte darauf zurückzuführen sein, dass Gerste vor allem in archaischer Zeit in den semiariden Zonen das Hauptgetreide war und so auch in alltäglichen medizinischen Anwendungen eher im Fokus stand als Weizen.

Die Dramatiker und die Redner

Abschließend soll ein Blick auf die Dramatiker und die Redner geworfen werden. Dies ist insbesondere unter dem Aspekt der zeitlichen Verteilung der Treffer von größerer Bedeutung. Denn die großen Corpora der ausgewählten Dramatiker stammen aus dem fünften Jahrhundert, während ein Großteil der Redner wie Demosthenes oder Aischines in der zweiten Hälfte des 4. Jahrhunderts wirkte. Gleichzeitig bewegen sich sowohl die Komödien als auch die Reden nah an der Realität und geben meist den besten Eindruck in Alltagsangelegenheiten, da sie, ähnlich wie bei Theophrasts *Ἠθικοὶ Χαρακτῆρες* ein vielfältig geformtes Abbild der Zeitgenossen und deren Alltags geben. Jedoch sind beiden Gattungen Überzeichnungen als gängiges und wirkungsvolles Stilmittel eigen: Der Komödiendichter wollte beim ἀγών die Zuschauergunst für sich gewinnen. Und auch die Redner versuchten in ihren Reden ihre Hörer von ihrem Anliegen zu überzeugen,⁷¹ sodass auch bei diesen der Aspekt der Autorenintention bei entsprechenden Treffern dahingehend hinterfragt werden muss.

Zunächst soll der Fokus auf die Tragiker gelegt werden. Es finden sich lediglich bei Aischylos zwei Belegstelle für Gerste in den *Ἰκέτιδες*, die das Brauen von Bier aus Gerste beschreibt und als Parallelstelle zur bereits genannten Textstelle Hdt. 2,77 gelten kann. Euripides verwendet lediglich 8 mal die allgemeinere Bezeichnung Getreide. Bei Sophokles finden sich keinerlei Treffer. Doch veranschaulicht dieser Befund, dass immerhin in der früheren Tragödie des Aischylos die Gerste eine Rolle spielt, während Weizen noch keinen Einzug in die Handlung hält.⁷²

Bei den Komödiendichtern Aristophanes und Phrynichos lässt sich Folgendes beobachten: Aristophanes verwendet 11 mal Weizen, 9 mal Gerste und sogar nur 4 mal den allgemeinen Begriff σῖτος, während letzterer – freilich bei den *Poetae Comici Graeci* sehr fragmentarisch

71 Vgl. Harris 2003, S. 8.

72 Vgl. Tabelle 2.

überliefert – keine der drei Bezeichnungen verwendet. Bei Aristophanes ist zunächst einmal auffällig, dass er *σῖτος* seltener einsetzt als die konkreten Beschreibungen *πυρός* und *κριθή*. Das rührt vermutlich daher, dass eben in Komödien gewisse Alltagsbezüge integriert sein müssen und hierfür die genaue Unterscheidung von Weizen und Gerste notwendig erschien. Somit dürften Gerste und Weizen bei Aristophanes ebenfalls als Mittel der Charakterisierung handelnder Personen fungiert haben.

Diese Vermutung erhärtet sich bei einem genaueren Blick auf die Herkunft der Ergebnisse im Corpus des Aristophanes, da sich die meisten Nennungen in jeweils zwei Werken finden. So stammen 5 der 11 Treffer für eine Form von *πυρός* aus den *Ὀρνιθεῖς* und vier der neun Treffer für *κριθή* aus der *Εἰρήνη*. Die Zahl lässt sich also weniger auf Kookkurrenzen zurückführen, sondern auf die Tatsache, dass die Getreidesorten innerhalb der jeweiligen Stücke mit der Handlung korrelieren. In den *Ὀρνιθεῖς* (506, 565, 566, 580, 622, 626), die davon handeln, dass die Vögel sich ihre Position zwischen Himmel und Erde zu Nutze machen und sich so zu den neuen Göttern aufschwingen wollen, ist Vers 506 zunächst eine allgemeinere Nennung von Weizen- und Gerstenanbau bei den Phöniziern. Die folgenden Treffer in Vers 565, 566 und 580 finden sich in dem zentralen Abschnitt, in dem Pisthetairos vorträgt, was von den Menschen zu fordern sei, wenn diese den Göttern opfern würden. So solle beim Opfer eines Schafs für Poseidon auch der Ente stets Weizen als Opfergabe hinzugegeben werden.⁷³ Und sofern die Menschen dies nicht beachteten, sollten Heerscharen von Vögeln die menschlichen Äcker leerfressen und die Menschen anschließend vergeblich auf Weizen von Demeter hoffen.⁷⁴ Daran schließen die beiden letzten Stellen aus der Passage an, in der Pisthetairos seine Vision vorstellt. So erläutert er, dass die Menschen den Vögeln nicht wie den Göttern Tempel zu erbauen hätten, sondern ihren Weizen als Opfergabe direkt aufs Feld streuen sollten, wobei auch an dieser Stelle Gerste ebenfalls – und zwar als erstes – genannt wird.⁷⁵ Das Ziel des Pisthetairos ist jedoch klar: *πυροῦς ὀλίγους* – eine Handvoll Weizen, nicht Gerste. Besonders in der Passage zu den Götteropfern, an die die neuen Opfer für die Vögel angeschlossen werden, wird also die Bedeutung des Weizens für Aristophanes deutlich. Bei diesem religionspraktischen Akt scheint Weizen eine exponierte Stellung zu haben und an die Hybris des Xerxes bei Herodot angelehnt zu sein scheint,⁷⁶ während die Gerste nur als historische Anspielung oder – sofern man Raus Emendation folgt – sexueller Wortwitz dient.

Im Gegensatz dazu spielt die Gerste in der *Εἰρήνη* (eir. 449, 962, 965, 1322) eine andere Rolle. Diese Komödie handelt von einem Winzer namens Trygaeus, der sich aufgrund des um sich greifenden Peloponnesischen Krieges auf einem Mistkäfer zum Olymp aufmacht, um die Ursache für die flächendeckenden kriegerischen Auseinandersetzungen der Gegenwart

73 Vgl. Av. 566. Rau hat in seiner Ausgabe in Vers 565 *πυροῦς* mit *κριθάς* emendiert, woraus sich der Sinn ergibt, dass Gerste als Opferbeigabe für das Blässhuhn beim Opfervorgang für Aphrodite gegeben werden sollte, wobei es sich hier um eine doppelte sexuelle Anspielung handeln würde, da Phaleris und Phalleris beim Hören doppeldeutig sind und die Gerstensamen ebenfalls eine sexuelle Anspielung sind, vgl. Rau 2016, S. 356. Die Emendation ist jedoch nicht durch eine Textüberlieferung gesichert, sondern wird von Rau nur mit einem Verweis auf Pax 960–967 vorgenommen. Allerdings scheint auch Vers 622 diese Emendation zu unterstützen, da dort als Abschluss der verschiedenen Opferungsvorschriften von Weizen und Gerste die Rede ist, was aber nur Sinn macht, wenn Gerste vorher auch mal genannt werden würde.

74 Vgl. Av. 580.

75 Vgl. Av. 622.

76 Vgl. Raaflaub 2016, S. 602f.

zu finden. Nachdem er die Göttin des Friedens, Eirene, befreit und die kriegerischen Auseinandersetzungen beendet hat, schlägt ihm jedoch nicht nur Gunst entgegen. So werfen ihm unter anderem Lanzen- und Schildmacher vor, ihre Geschäfte zerstört zu haben, woraufhin der Chor Trygaeus antwortet, dass diese, wenn sie Krieg wünschten, doch lieber von Piraten ergriffen und nichts außer Gerste zu essen bekommen sollten.⁷⁷ Ein deutlicher Hinweis für die Geringschätzung der Gerste gegenüber anderen Nahrungsquellen und besonders gegenüber Weizen.⁷⁸ Doch zeigt die nächste Stelle, in der zwei der Treffer belegt sind und in der ein Gerstenopfer beschrieben wird, dass Gerste dennoch offensichtlich eine grundlegende Bedeutung für die Götterverehrung hatte, wenn auch aus althergebrachter Tradition.⁷⁹ Dies wird in der letzten Belegstelle deutlich, in der der Chor im Zuge der Hochzeit von Trygaeus mit Opora, der Göttin der Ernte, einstimmt, sie möge unter anderem eine reiche Gerstenernte gewähren.⁸⁰ Auffallend ist, dass in der Aufzählung der Ernteerträge nur die zum Anbau in Attika besser geeignete Gerste, aber nicht Weizen genannt wird. Dies kann – gerade unter Einbeziehung der Belegstellen in den Ὀρνιθεὺς – ein Indiz dafür sein, dass in der zweiten Hälfte des 5. Jahrhunderts noch überwiegend Gerste angebaut wurde, sodass der Fokus des Aristophanes beim Ernteertrag auf der Gerste lag, während durch den Import von Weizen die Opferung in den Ὀρνιθεὺς durchaus mit kostspieligerem Weizen erfolgen sollte.

Die Durchsicht der Belegstellen bei Aristophanes zeigt, dass dieser beide Getreidesorten sehr bewusst einsetzt. Während beispielsweise Pisthetairos in den Ὀρνιθεὺς den hochwertigen Weizen als Getreide-Opfergabe einfordert, wird in der Εἰρήνη Gerste geopfert. Bereits in der ersten Stelle mit der Forderung, den Krieg wünschenden Lanzen- und Schildmachern solle in Piratengefangenschaft nur Gerste als Nahrung dienen, wird die Geringschätzung der Gerste bei Aristophanes' Zeitgenossen deutlich. Gerste war also offensichtlich in Übereinstimmung zu IG II² 1672 tatsächlich das am häufigsten angebaute und damit alltägliche Getreide, was daher auch regelmäßig als Opfergabe verwendet wurde. Es zeichnet sich also anhand der Dramatiker ab, dass es zwar keine größere Präsenz der Gerste in den Aristophanesquellen gibt, aber dennoch eine klare Differenzierung zwischen Weizen und Gerste und ihrer Rolle für die attische Gesellschaft zu beobachten ist.⁸¹

Richtet man den Fokus nun auf die Redner, fällt auf, dass auch bei den Reden Weizen nicht wesentlich häufiger als Gerste genannt wird. Denn Weizen lässt sich lediglich bei Lykurgos von Athen, Demosthenes und Deinarchos in insgesamt nur vier Belegstellen ausmachen und

77 Vgl. Pax 449.

78 Vgl. Olson 1998, S. 168.

79 Auch hier findet sich zum Abschluss in Pax 966 f. ein Witz durch die sexuelle Anspielung mit den Gerstensamen, die jedoch den sakralen Akt an sich nicht beeinflusst, vgl. Olson 1998, S. 254.

80 Vgl. Pax 1322.

81 Damit könnte einhergehen, dass sich der Getreideanbau in Attika zumindest noch zu Aristophanes' Zeiten auf Gerste konzentriert. Das bedeutet jedoch nicht, dass insbesondere die Bürger Athens auch überwiegend Gerste konsumierten, vgl. Schmitz 2004, S. 468 sowie Aristoph. eccl. 817, zumal Gerste und dessen Abfallprodukte beim Dreschen auch als Futter für Nutztiere verwendet wurden.

damit auch nur bei den Rednern der späten Klassik. Dabei handelt es sich bei Lykurgos⁸² auch nur um eine Schwurformel, in der Weizen neben Gerste – Gerste folgt auf Weizen⁸³ – in einer Aufzählung verschiedener attischer Nahrungsmittel auftaucht.

Die beiden Belegstellen für Demosthenes befinden sich wiederum in der 19. Rede *Über die Truggesandtschaft*, die sich gegen Aischines und Philokrates richtete und ihnen Selbstbereicherung aus Landgütern ehemaliger Bündner, insbesondere mit dortigen Holz- und Weizenenerträgen, vorwirft.⁸⁴ Hier handelt es sich aber um Weizenanbau außerhalb Attikas, sodass daraus ersichtlich wird, dass in ausländischen Besitzungen mit besseren klimatischen Bedingungen Weizen angebaut wurde.

Die zweite Belegstelle stammt aus einem Privatprozess gegen Phormio, der angegeben hatte, aufgrund eines Schiffsunglücks nicht nur Ware, sondern auch sein gesamtes Geld verloren zu haben und daher dem Kläger die geliehene Summe nicht mehr zurückzahlen könne. Die Passage, in der Weizen explizit genannt wird, befasst sich mit den Wohltaten, die der Kläger dem attischen Volk habe zukommen lassen und aufgrund derer er die Hörerschaft dazu verleiten möchte, seiner Klage zu folgen. So habe er zu Verteidigungszwecken ein Talent gespendet, als Alexander in Theben einfiel⁸⁵ und:

ὄτε δ' ὁ σῖτος ἐπετιμήθη τὸ πρότερον καὶ ἐγένετο ἑκκαίδεκα δραχμῶν, εἰσαγαγόντες πλείους ἢ μυρίους μεδίμνους πυρῶν διεμετρήσαμεν ὑμῖν τῆς καθεστηκυίας τιμῆς, πέντε δραχμῶν τὸν μέδιμνον, καὶ ταῦτα πάντες ἴστε ἐν τῷ πομπείῳ διαμετρούμενοι.

„And when grain earlier advanced in price and reached sixteen drachmae, we imported more than ten thousand medimni of wheat, and measured it out to you at the normal price of five drachmae a medimnus, and you all know that you had this measured out to you in the Pompeium.”

(Demosth. or. 34,39; Übers. Murray 1936)

Die Übersetzung ist hier etwas ungenau, da es sich an dieser Stelle nicht um einen ‚normalen‘, sondern den ‚angeordneten‘ (καθεστηκυίας) Preis der Polis handelt, die eine bestimmte Getreidemenge zur Grundversorgung zu einem Festpreis ausgab.⁸⁶ Der Kläger hätte, da er sein Getreidekontingent unabhängig von der Grundversorgung verkaufte, einen höheren Preis verlangen können, entschied sich aber offenbar dafür, zum Wohle Athens seine Menge zum

82 Vgl. Leokr. 77.

83 Vgl. Kapitel 3.

84 Vgl. Demosth. or. 19,145: τῆ μὲν δὴ πόλει ταῦτ' ἐκ τῆς εἰρήνης γέγονεν, ὧν οὐδ' εὐρεῖν αἰσχίῳ ῥάδιον: τοῖς δὲ πρέσβεσιν τί τοῖς ταῦτα πράξασιν; τὰ μὲν ἄλλα σιωπῶ πάνθ', ὅσ' ἐοράκαθ' ὑμεῖς, οἰκίας, ξύλα, πυρούς, ἀλλ' ἐν τῇ τῶν ἀπολωλότων συμμάχων χώρα κτήματα καὶ γεωργίαι παμπληθεῖς, Φιλοκράτει μὲν τάλαντον ἔχουσαι πρόσοδον, τούτῳ δὲ Αἰσχίῳ τριάκοντα μνᾶς. – „That is what the peace has brought to the city: you could not easily invent anything more dishonorable. What has it brought to the ambassadors who contrived that dishonor? I say nothing of the wealth that lies before your eyes – houses, timber, grain; but in the country of our ruined allies there are estates and extensive farms bringing in a rental of a talent to Philocrates and half a talent to Aeschines.“ (Übers. Vince 1926). Auffallend ist hier jedoch, dass durch den Hinweis τὰ μὲν ἄλλα σιωπῶ πάνθ', ὅσ' ἐοράκαθ' ὑμεῖς, οἰκίας, ξύλα, πυρούς offensichtlich auch Weizen in Attika angebaut wurde.

85 Vgl. Demosth. or. 34,38.

86 Vgl. Engels 2000, S. 111f. sowie insbesondere ibid., S. 112 FN 58.

angeordneten Preis abzugeben und damit auf Gewinn zu verzichten. Er hat demnach mehr als zehntausend Medimnoi für den festgesetzten Preis von fünf Drachmen pro Medimnos veräußert.

Deutlich häufiger finden sich bei Demosthenes jedoch Verweise auf Gerste, so auch in der 42. Rede, auf die in den Kapiteln 2, 20 und 24 drei der sechs ‚Gerstentreffer‘ entfallen. Diese Rede richtet sich gegen Phainippos und zielte darauf ab, dessen Vermögen höher einzuschätzen und ihn damit in den Stand eines Leiturgiepflichtigen zu bringen.⁸⁷ Um dieser Pflicht zu entgehen, soll er τὰς κριθὰς [...] καὶ τᾶλλα⁸⁸ aus den versiegelten Gebäuden entfernt haben. Dieser Fall ist dahingehend interessant, da er ein Beleg für den intensiven Anbau von Gersten in Ostattika ist.⁸⁹ In 42,20 nennt der Kläger zudem einen Wucherpreis, mit dem besagter Phainippos seine Gerste verkauft haben soll, nämlich mit 18 Drachmen pro Medimnos, was gemäß der oben genannten Stelle or. 34,39 fast das Vierfache des dort öffentlich festgeschriebenen Weizenpreises gewesen wäre. Es ist zwar naheliegend, dass es sich bei dieser Preisangabe um eine Übertreibung aus rhetorischen Gründen handelt, um die Leiturgiepflicht für den Beklagten durchzusetzen, doch ändert das an der grundlegenden Konstellation, dass offensichtlich Gerste zu sehr hohen Preisen abgesetzt werden konnte, nichts. Diese Preisunterschiede lassen sich jedoch durch lokale Unterschiede hinsichtlich des Getreidebedarfs erklären. So hatte die Lauriotike durch die vielen im Silberbergbau tätigen Sklaven⁹⁰ einen sehr hohen Getreidebedarf, der sicher phasenweise gravierende Preisunterschiede auf dem freien Getreidemarkt beziehungsweise den lokalen ἀγοραὶ nach sich gezogen haben wird und für einen solchen Wucherpreis verantwortlich gewesen sein könnte.⁹¹

Zudem gibt die Stelle beiläufig einen weiteren, wichtigen Hinweis hinsichtlich des Anbauverhaltens in Attika, da er den Gewinn mit Gerste erzielt, das er in den entlegensten Gebieten Attikas (ἐσχατιὰς) angebaut habe. Somit handelt es sich hier um einen literarischen Hinweis aus spätklassischer Zeit, wo also nach wie vor vor allem die Gebiete der attischen Peripherie genutzt wurden, um Gerste – und keinen Weizen – anzubauen.

In Kapitel 24 dieser Rede folgt schließlich noch die Feststellung des Klägers, der Beklagte habe zwar für die Festsetzung der εἰσφορὰ seinen Wagen, dafür aber bei den Gersten-, Wein- und anderen Erträge nur zu einem Zehntel der tatsächlichen Menge angegeben.

In spätklassischer Zeit wurde Gerste also weiterhin in größerem Umfang angebaut und hatte eine große Bedeutung. Dass die Versorgungslage nicht nur mit Weizen, sondern eben auch mit Gerste durchaus schwanken konnte, zeigt der Wucherpreis, den Phainippos für die Gerste eingestrichen haben soll.

87 Der Kläger ist nicht bekannt, ebenso ist die Gegenrede des Phainippos nicht überliefert, vgl. MacDowell 2009, S. 147.

88 Vgl. Demosth. or. 42,2.

89 Die Farm des Phainippos lag im Demos Kytherros nahe der Ostküste Attikas, vgl. MacDowell 2009, S. 148.

90 Die Schätzungen gehen hier von bis 100000 im 4. Jahrhundert aus, wobei eine mittlere fünfstellige Zahl als grobe Orientierung dienen kann, vgl. Kalcyk 1982, Untersuchungen zum attischen Silberbergbau, S. 159–165. Bei einer solchen mittleren fünfstelligen Zahl berechnet Rhill 2010, S. 115, einen Tagesbedarf von ca. 4000 Kilogramm an Nahrungsmitteln, um die Arbeiter in und an den Silberminen zur versorgen.

91 Zumal die Lage der Farm nahe der Ostküste Attikas dafür spricht, dass das Getreide auf den Märkten der Südostküste abgesetzt wurde, entweder über die Handelsstraße an der Ostküste, oder per Schiff oder Floß an einen der zahlreichen kleine Häfen der Südostküste, vgl. Goette 2000, S. 62–65.

Auch ein Blick auf die deklinierten Formen von σῖτος lässt bei Demosthenes Interpretationsspielraum zu. Den 2 Belegstellen für Weizen und 6 für Gerste stehen insgesamt 72 Belegstellen für σῖτος gegenüber. Somit scheint auch für Demosthenes die Getreideart eine untergeordnete Rolle zu spielen, die er nur benennt, wenn sie für den Kontext der Rede auch tatsächlich von Bedeutung war. Der Hang zum allgemeinen Begriff σῖτος spricht also dafür, dass Gerste und Weizen immer seltener unterschieden wurden.⁹²

Einen weiteren interessanten Aspekt bringt eine Belegstelle bei Isaios mit sich:

παιδὸς γὰρ οὐκ ἔξεστι διαθήκην γενέσθαι: ὁ γὰρ νόμος διαρρήδην κωλύει παιδὶ μὴ ἐξεῖναι συμβάλλειν μηδὲ γυναικὶ πέρα μεδίμνου κριθῶν.

”For a minor is not allowed to make a will; for the law expressly forbids any child – or woman – to contract for the disposal of more than a bushel of barley.”

(Isaios 10,10; Übers. Forster 1983)

So wurde also noch zu Beginn des vierten Jahrhunderts der juristische Richtwert für die Geschäftsfähigkeit von Frauen oder Kindern – in diesem Fall des Demochares – an Gerste festgemacht, nämlich einem Medimnos, was zu der Zeit in etwa drei Drachmen entsprochen haben dürfte.⁹³ Dies zeigt einerseits die ursprüngliche Bedeutung der Gerste für Athen, da es sich hierbei um ein Gesetz aus vorklassischer Zeit gehandelt haben dürfte.⁹⁴ Andererseits veranschaulicht es die kontinuierliche Relevanz von Gerste, da diese weiterhin trotz schwankender Preise als Richtlinie galt, um die Geschäftsfähigkeit von Frauen und Minderjährigen zu reglementieren.⁹⁵ Gleichzeitig diente Gerste auch hier als gesellschaftliches Differenzierungsmerkmal, da vor allem Frauen durch die Kopplung ihrer Geschäftserlaubnis an einen Medimnos der Getreidesorte in ihren Handlungsmöglichkeiten eingeschränkt wurden.⁹⁶ So fehlten diesen nicht nur viele bürgerliche Rechte, sondern sie blieben auch wirtschaftlich in einer gewissen Weise eingeschränkt, auch wenn der festgelegte Richtwert von einem Medimnos Gerste durchaus Handelsaktivitäten auf dem Markt ermöglicht haben müsste.⁹⁷

92 Auch wenn sich dies freilich schon bei Herodot und Thukydides beobachten lässt, vgl. oben, S. 9.

93 Vgl. Ussher 1986, S. 217.

94 Vgl. Schmitz 2004, S. 443 FN 228 geht von einem solonischen Gesetz aus.

95 Vgl. Griffith-Williams 2009, S. 227f.

96 Vgl. hierzu auch Aristoph. Eccl. 1025, wo die Konsequenz des Dekrets durch umgekehrte Anwendung auf Männer deutlich wird, sowie hierzu auch Schol. Aristoph. Ekk. 1025 und die Ausführungen in Rutherford 1896, S. 554. Die zweite Belegstelle bei Isaios ist in der elften Rede, Kapitel 43. Dort wird die Gerste allerdings ebenfalls nur als Teil einer Vermögensaufstellung aufgezählt. Das Beispiel von Lykurgos wurde bereits oben behandelt und war, neben Weizen, Teil eines Schwurs, vgl. Lyk. Leokr. 77.

97 Vgl. Schmitz 2004, S. 443 FN 228.

Fazit

Der methodische Ansatz des *Blended Reading* konnte wichtige Beiträge zur Forschungsdiskussion über den Getreidekonsum in Attika liefern. So zeigte die Frequenzanalyse des Greek Word Study Tools, dass die Gerste im Laufe der klassischen Zeit keinesfalls seltener in den literarischen Quellen genannt wurde, sondern bei den Rednern der Spätclassik weiterhin Gegenstand juristischer Auseinandersetzungen war und weiterhin gesellschaftliche Relevanz besaß. Die Kookkurrenzanalyse zeigte nicht nur, dass die antithetische Herangehensweise wenig zielführend ist, da beide Begriffe in ca. einem Viertel aller Belegstellen gemeinsam genannt werden und auch insgesamt gleichmäßig in den literarischen Quellen vorkommen, sondern erweiterte auch die Feststellung der Frequenzanalyse dahingehend, dass Gerste und Weizen in klassischer Zeit beide einen großen Stellenwert hatten. Denn nach der Begrenzung der automatisch durch die Kookkurrenzanalyse generierten Perseus-Treffer auf die klassischen Autoren war eine Kookkurrenz sogar in fast jeder zweiten Belegstelle zu finden. Zudem verdeutlichte die Kookkurrenzanalyse, dass die antiken Autoren bei der gemeinsamen Nennung beider Getreidearten deutlich dazu tendierten, Weizen zuerst zu nennen. Dies scheint mit der Bedeutung von Weizen einherzugehen, der durch seine Backeigenschaften begehrt, aber selbst in den semiariden Gebieten Attikas schwieriger anzubauen war und deshalb größtenteils importiert werden musste.

Bestätigt wurden die Erkenntnisse aus den Ergebnissen der Textminingtools auch durch die qualitative Analyse. Zunächst im Hinblick auf die These, Weizen habe Gerste zunehmend verdrängt. Dies kann nicht nur zahlenmäßig, sondern auch nach Durchsicht der Quellen kaum bestätigt werden, da es zahlreiche Beispiele – insbesondere bei Demosthenes – gibt, in denen auch in der Literatur zum Ende der klassischen Zeit der Gerste eine zentrale Ernährungsfunktion zukam. So baute Demosthenes seine Klage gegen Phainippos u.a. auf die widerrechtliche Aneignung von Gerste von seinen Besitzungen auf, die im Zuge eines Antidosis-Prozesses beschlagnahmt worden waren. Zudem führt er in dieser Rede den Wucher des Phainippos an, der mit Gerste einen erstaunlichen Preis von 18 Drachmen pro Medimnos erreicht habe, was besonders im Hinblick auf die Rede *Gegen Phormio* erstaunt, in der ein von der Polis festgesetzter Weizenpreis auf gerade einmal 5 Drachmen pro Medimnos beziffert wird, wobei hierfür, wie oben beschrieben, die lokal unterschiedlichen Getreidebedarfe eine plausible Erklärung sind.

Somit ist ein Bedeutungsverlust der Gerste zu Gunsten von Weizen bis in die spätclassische Zeit anhand der literarischen Quellen weder qualitativ noch quantitativ auszumachen. Auch die Charakterisierung des Unverschämten mit Hilfe der Gerste bei Theophrast sowie die charakterisierenden Stellen in Aristophanes' *Ειρήνη* lassen den Schluss zu, dass Gerste auch als literarisches Stilmittel sowohl in der Mitte der Klassik als auch gegen Ende gleichermaßen zur Beschreibung der ärmeren Einwohner Athens diente – zumal schon die Versorgung der Nutztiere und vielen Sklaven, die nicht nur in der Landwirtschaft, sondern auch zu Zehntausenden in der Lauriotike arbeiteten, einen enormen Bedarf an Gerste nach sich zog.

Im Hinblick auf die Forschungsdiskussion zwischen Garnsey/Sallares/Gallant und Jardé/Whitby/Morley legen die Ergebnisse nahe, dass der Import von Getreide nicht nur sehr hoch war und möglicherweise auch den Bedarf an Gerste abdecken musste. Hierfür konnten im Zuge der qualitativen Auswertung weitere Indizien zusammengetragen werden, wie die oben

genannte Klage gegen Phainippos oder auch die Nikiasrede bei Thukydides.

Die textmininggestützte Analyse der Begriffe *πυρός*, *κριθή* und *σίτος* konnte demnach zeigen, dass über den gesamten Zeitraum des klassischen Athens für beide Getreidesorten ein hoher Bedarf bestand und in den Quellen lediglich eine gewisse Wertung zu Gunsten des Weizens zu erkennen ist – was aber auch mit den natürlichen Eigenschaften von Weizen korreliert.

Abschließend lässt sich festhalten, dass die meisten Berechnungen über den tatsächlichen Getreideanbau in Attika doch recht spekulativ erscheinen, insbesondere wenn man einen etwas breiteren methodischen Ansatz wählt. Einerseits zeigen die Textmining-Tools, dass sich die Frage nach einer Tendenz Weizen oder Gerste nicht beantworten lässt und damit auch kein Ertrag aus der attischen Landwirtschaft wirklich präzise berechnet werden kann. Denn der Ertrag für Gerste wäre deutlich zuverlässiger und konstanter als der für Weizen. Hinzu kommen die naturwissenschaftlichen Ansätze: Einerseits ist der klimatische Unterschied zwischen arider bis subarider Zone im Süden Attikas und semiarider Zone im Norden zu nennen, der ebenfalls eine besondere Beachtung in den Ertragsberechnungen finden müsste. Andererseits ist die Pollenanalysen unter der Leitung von Katerina Kouli in der Marathon- und der Brauronebene zu berücksichtigen, die in den beiden fruchtbaren Gegenden Attikas insbesondere in der klassischen Zeit einen nicht sehr ausgeprägten Getreideanbau zu Gunsten von Oliven und Pistazien nahelegen. Somit sind die Berechnungen über die Anbauflächen in Attika und ein daraus resultierender Ertrag, die von Jardé über Garnsey bis Moreno aufgestellt wurden, zwar interessante Annäherungen an ein landwirtschaftliches Potential Attikas. Vom realen Anbau dürften sich diese Hochrechnungen aufgrund der Möglichkeit des Imports sowohl von Weizen als auch von Gerste jedoch deutlich unterscheiden.

Literatur

- Asheri u.a. 2007: D. Asheri et al. 2007, *A Commentary on Herodotus, Books I-IV*, Oxford.
- Bissa 2009: E.M.A. Bissa, *Governmental Intervention in Foreign Trade*, Leiden.
- Dalby 1992: A. Dalby, *Greeks Abroad: Social Organisation and Food among the Ten Thousand*, *JHS* 112, S. 16–30.
- Diggle 2009: J. Diggle, *Theophrastus. Characters*, Cambridge.
- Feng / Fu 2013: S. Feng / Q. Fu, *Expansion of global drylands under a warming climate*, in: *Atmos. Chem. Phys.* 13, S. 10081–10094.
- Gallant 1991: T. W. Gallant, *Risk and survival in Ancient Greece. Reconstructing the Rural Domestic Economy*, Cambridge.
- Garnsey 1988: P. D. A. Garnsey, *Famine and Food Supply in the Graeco Roman World. Responses to Risks and Crisis*, Cambridge.
- Goette 2000: H. R. Goette, *Ἡ ἀξιόλογος δῆμος Σούνιον - Landeskundliche Studien in Südost-Attika*, Rahden/Westf.
- Griffith-Williams 2009: B. Griffith-Williams, *A Commentary on selected Speeches of Isaios*, Leiden.
- Hansen 2006: M. H. Hansen, *The Shotgun Method: The Demography of the Ancient Greek City-State Culture*, Columbia.
- Hansen 1988: M.H. Hansen, *Three Studies in Athenian Demography*, in: *Historisk-Filosofiske Meddelelser* 56, S. 1–28.
- Harris 2003: E.M. Harris, *Law and Economy in Classical Athens: [Demosthenes], Against Dionysodorus*, in: C.W. Blackwell (Hg.), *Demos: Classical Athenian Democracy*, in: A. Mahoney and R. Scaife (eds), *The Stoa: A Consortium for Electronic Publication in the Humanities* [www.stoa.org]. (Zuletzt aufgerufen am 29.07.2018).
- Hornblower 2010: S. Hornblower, *Commentary on Thucydides III*, Oxford.
- How / Wells 1989: A. How / J. Wells, *A Commentary on Herodotus 1, Books I-IV*, Oxford.
- How / Wells 2002: A. How / J. Wells, *A Commentary on Herodotus 2, Books V-IX*, Oxford.
- Jardé 1925: A. Jardé, *Les Céréales dans l'antiquité grecque*, Paris.
- Kalcyk 1982: H. Kalcyk, *Untersuchungen zum attischen Silberbergbau. Gebietsstruktur, Geschichte und Technik*, Frankfurt am Main / Bern.
- Kouli 2012: K. Kouli et al., *Palynological investigation of Holocene palaeoenvironmental changes in the coastal plain of Marathon (Attica, Greece)*, *Geobios* 42, S. 43–51.
- Kouli 2009: K. Kouli, *Vegetation development and human activities in Attiki (SE Greece) during the last 5,000 years*, *VegetHistArchaeobot* 21, S. 267–278.

Lemke / Stulpe 2016: M. Lemke / A. Stulpe, Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung., in: M. Lemke, Matthias / G. Wiedemann (Hg.), Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse, Wiesbaden, S. 43–54.

Lohmann 1993: H. Lohmann, *Atene-Ἀθήνη*. Forschungen zu Siedlungs- und Wirtschaftsstruktur des klassischen Attika (2 Bde.), Köln / Weimar / Wien.

MacDowell 2009: D. M. MacDowell, *Demosthenes the Orator*, Oxford.

Moreno 2007: A. Moreno, *Feeding the Democracy. The Athenian grain supply in the fifth and fourth centuries B.C.*, Oxford.

Morley 2007: N. Morley, *Trade in Classical Antiquity*, Cambridge.

Ober 2016: J. Ober, *Das Antike Griechenland. Eine neue Geschichte*, Stuttgart.

Ober 2008: J. Ober, *Democracy and Knowledge. Innovation and learning in classical Athens*, Princeton / Oxford.

Olson 1998: S. D. Olson, *Aristophanes. Peace*, Oxford.

Osborne 2004: R. Osborne, *Greek History*, London / New York.

Raaflaub 2016: K. A. Raaflaub, *Die große Herausforderung. Herodot, Thukydides und die Erfindung einer neuen Form von Geschichtsschreibung*, HZ 302, S. 597–622.

Rau 2016: P. Rau, *Aristophanes. Komödien II*, Darmstadt.

Rihll 2010: T. E. Rihll, *Making Money in Classical Athens*, in: David J. Mattingly, John Salmon (Hrsg.), *Economies Beyond Agriculture in the Classical World*, London / New York, S. 115–142.

Ruffing 2012: K. Ruffing, *Antike Wirtschaft*, Darmstadt.

Rutherford 1896: W. G. Rutherford, *Scholia Aristophanica* (Bd. 2), London.

Sallares 2007: R. Sallares, *Ecology*, in: W. Schneidel, / I. Morris / R. Saller, (Hgs.), *The Cambridge Economic History of the Greco-Roman World*, Cambridge, S. 15–37.

Sallares 1991: R. Sallares, *The ecology of the Ancient Greek World*, Ithaca.

Schmitz 2007: W. Schmitz, *Haus und Familie im antiken Griechenland*, München.

Schmitz 2004: W. Schmitz, *Nachbarschaft und Dorfgemeinschaft im archaischen und klassischen Griechenland*, Berlin.

Triantaphyllou u.a. 2010: M. V. Triantaphyllou et al., *Paleoenvironmental changes since 3000 B.C. in the coastal marsh of Vravron (Attiki, SE Greece)*, *Quat. Int.* 216, S. 14–22.

Ussher 1986: R. G. Ussher, *Aristophanes Ecclesiazusae*, Bristol / New York.

Ussher 1993: R. G. Ussher, *The characters of Theophrastus*, London / Bristol.

Van Liefferinge 2014: K. Van Liefferinge et al., Reconsidering the role of Thorikos within the Laurion silver mining area (Attica, Greece) through hydrological analyses, *Journal of Archaeological Science* 41, S. 272–284.

Whitby 1998: M. Whitby, The grain trade of Athens in the fourth Century B.C., in: H. Parkins (Hg.), *Trade, Traders and the Ancient City*, London, S. 102–128.

Übersetzungen

Feix 1995: J. Feix, *Herodot. Historien*, Darmstadt.

Forster 1983: E. S. Forster, *Isaeus. Orationes*, Cambridge.

Klose 1988: *Theophrast. Charaktere*, Stuttgart.

Murray 1936: *Demosthenes. Orations IV (Private orations : XXVII – XL)*, Cambridge.

Vretska 2009: H. Vretska / W. Rinner, *Thukydides. Der Peloponnesische Krieg*, Stuttgart.

Vince 1926: C. A. Vince, *Demosthenes. Orations II (De corona and De falsa legatione)*, Cambridge.

Autorenkontakt⁹⁸

Sven-Philipp Brandt

Universität Göttingen
Althistorisches Seminar
Humboldtallee 21
37073 Göttingen

Email: sven-philipp.brandt@uni-goettingen.de

98 Die Rechte für Inhalt, Texte, Graphiken und Abbildungen liegen, wenn nicht anders vermerkt, bei den Autoren. Alle Inhalte dieses Beitrages unterstehen, soweit nicht anders gekennzeichnet, der Lizenz CC BY-SA 4.0 International.