DIGITAL CLASSICS ONLINE

Mapping the Words
Experimental visualizations of translation structures between Ancient Greek and Classical Arabic

Torsten Roeder

Abstract: The article deals with presentation forms of linguistic transformation processes from ancient Greek sources that were translated into classical Arabic from the 9th to 11th century AD. Various examples demonstrate how visualizations support the interpretation of corpus structures, lexical differentiation, grammatical transformation and translation processes for single lexemes in the database project Glossarium Graeco-Arabicum. The database contains about 100,000 manually collected word pairs (still growing) from 76 texts and their translations. The article discusses how the project utilizes Sankey diagrams, tree maps, balloon charts, data grids and classical coordinate systems to point out specific aspects of the data. Visualizations not only help beginners to understand the corpus structure, they also help editors and specialized users to identify specific phenomena. A well-documented interface design is crucial both for usability and interpretative work.[1]

1. Introduction

1.1.   Translations as Research Field in the Digital Humanities

Analyses of languages and their structures belong to the fields that have applied computational methods to a humanities-related subject in a very early stage. An open question may stay here whether this was due to the abstract structures of information technology and linguistics, which were particularly easy to combine, or to the large amounts of data that suggested the use of computers, or even to further factors that may have been involved.[2] Eventually it has turned out as a major feature of computational linguistics that digital recording of language-related data would not need to confine itself to the creation of an electronic equivalent of a printed reference book. To the contrary, its great potential lies in the possibility of complex quantitative and structural analyses, which could unveil leads and clues to structures, processes and developments in language.[3]

While digital resources that deal with just one language are available in numerous and diverse forms, bilingual resources still exist only in niches. Yet these possess great importance for investigations of cultural transfer: Assumed that, in any particular cultural context, an interest existed to express or to understand a certain message in another language, translation belongs

---

1   This is a translated, reworked and slightly expanded version of an article which was originally published in the German digital humanities journal "Zeitschrift für digitale Geisteswissenschaften"; see Roeder (2016). Further, the topic has been subject of a lecture in the Digital Classicists Berlin Seminar; see Arzhanov/Roeder (2016).

2   For a thorough comparison of the disciplines of corpus linguistics, literary computing, and computational linguistics cf. Lüdeling/Zeldes (2009).

3   E.g. the Regionalsprache.de (REDE) project (http://www.regionalsprache.de), which uses advanced digital methods to investigate German dialects.

to the most significant facets of intercultural exchange. Not only does the translation process transport information, but it also partly transfers also lexical expressions and structural elements into the other language. In this way, translations provide impetus for cultural development, which is reflected in vocabulary and grammar. Systematically designed bilingual resources are therefore of major relevance for the understanding of cultural transfer. For this reason, the resources that come into question for investigating historical transfer studies are, amongst others, written translations.

Further, such transfer processes normally work only in one direction, i.e. they are not reciprocal in their effect: This would only be the case when an equivalent transfer process also occurred in the other direction, which is usually not applicable in the case of translations (or at least it does not happen in the same area of interchange). Thus, it is often sufficient to consider only one direction of transmission to examine a cultural trend.

Another more special case are receptions of earlier cultures, such as the reception of Ancient Greek literature in the Renaissance or in the medieval Middle East, which took place only unilaterally. In the following, the Glossarium Graeco-Arabicum, a project that deals with Arabic translations of ancient Greek writings, will work as an example of such receptive cultural transfer processes. In addition, a look at a follow-up project will be given, where the special role of early Syriac translations as a bridging language from Greek to Arabic will be the focus of investigation.

1.2.   The Glossarium Graeco-Arabicum

Figure 1: Aristotle teaching, from Kitāb naʿt al-hayawān, 7th century AH (13th century AD).[4]

---

4   British Library, MS Or. 2784, fol. 96r, Reproduced and cited from: Nasr (1976), 50, plate 22. Any copyrights are to be observed.

The Glossarium Graeco-Arabicum (short: GlossGA) undertakes the lexicographical analysis and description of a corpus that is in many aspects of particular interest for studies of cultural transfer. During a flowering period of Arabic translation movements, which lasted approximately from the 3rd to the 5th century AH (9th to 11th century AD), an intensive Arabic reception of classical Greek writings took place in the area of Baghdad.[5] The writings of Aristotle, Plato, Euclid, Galen and many others were translated, summarized and paraphrased into Arabic, mostly by scholars with Christian backgrounds, and then utilized for the study of philosophy, mathematics, medicine, astronomy, and other sciences (cf. figure 1). It is thus a historic transfer process of ancient Greek literature into classical Arabic culture.

Language	Greek	Arabic
Lexeme	λόγος	قول
Root / Stem		q w l – – / –
Part of Speech	noun	noun
Expression	οἱ Σωκρατικοὶ λόγοι	al-aqāwīlu l-mansūbatu ilā Suqrāṭa
Annotation	–	paraphr.
Quotation	οὐδὲν γὰρ ἂν ἔχοιμεν ὀνομάσαι κοινὸν τοὺς Σώφρονος καὶ Ξενάρχου μίμους καὶ τοὺς Σωκρατικοὺς λόγους	wa-ḏālika annahū laysa lanā an nusammiya bi-māḏā yušāriku ḥikāyātu wa-tašbīhātu l-šāʿiri Sūfruna wa-Kasānarḫusa wa-l-aqāwīlu l-mansūbatu ilā Suqrāṭa
Reference	Arist. Poet. 1, 1447b11	220.20
Lexica	Search for translation in Perseus	Search for translation in Perseus

Figure 2: Glossarium Graeco-Arabicum, Word record for λόγος / قول.[6]

The corpus of the Glossarium Graeco-Arabicum[7] is based on 76 original texts and their translations. It includes just over 100,000 manually selected word entries so far. One single entry consists of a word from the Greek original text, the corresponding word (or phrase) in the Arabic translation and additional grammatical properties (base lexeme, part of speech and root). Furthermore, a complete entry (cf. figure 2) contains accurate records of the original quotations. Additionally, the database provides contextual links and data connections to other digital resources in the field of the Graeco-Arabica, such as Perseus Digital Library[8], G2A Web Application for Literary Computing[9] and Digital Corpus for Graeco-Arabic Studies[10]. Since early 2016, the texts are also associated with the names of their authors and (if known) their translators, which are, in most cases, connected to library standard catalog entries (VIAF, GND) and to English Wikipedia pages.

---

5 Cf. D'Ancona (2013).

6 https://telotadev.bbaw.de/glossga/glossary.php?id=131860 (screenshot: January 17, 2016).

7 Glossarium Graeco-Arabicum, European Research Council; Ruhr-Universität Bochum; Berlin-Brandenburgische Akademie der Wissenschaften, http://telota.bbaw.de/glossga (March 5, 2017); the current development version can be found at https://telotadev.bbaw.de/glossga (March 5, 2017). Cf. also Endress et al. (2013); Arzhanov/Roeder (2013).

8 http://www.perseus.tufts.edu/hopper (March 5, 2017).

9 http://g2a.ilc.cnr.it (March 5, 2017).

10 http://www.graeco-arabic-studies.org (March 5, 2017).

Figure 3: File cards used to compile GALex and Glossarium Graeco-Arabicum. Photo: private.

It was the original aim of the project to create the printed lexicon "A Greek and Arabic Lexicon" (short: GALex).[11] Starting in the 1980s, the team began to collect data on handwritten file cards (cf. figure 3), and later, in the course of the "digital turn", scanned them into TIFF images and began to transcribe them manually[12] into a relational database.[13] This database has since been titled Glossarium Graeco-Arabicum.[14] The overcoming of the manifold technical challenges that came with the parallel use of Greek and Arabic characters with diacritics and vowel signs respectively, involving issues from presentation to search indexing, was an essential step towards a stable basis for scientific work with modern usability.[15] Thanks to appropriate funding[16] and infrastructural support, it was possible to redesign gradually the database from a file card management system into a research instrument: It has evolved from a relatively simple and, initially, only internally used database application to a freely accessible digital encyclopedia which now exists parallel to the printed GALex. Indeed, print and digital versions serve different purposes: While the GALex volumes provide consistent entries, currently from letters Alif to Bā', the database covers the full lexical range, while the collecting and editing of data is an open process. Despite its work-in-progress nature, the database is already used regularly for research and frequented by about 30–40 recurrent international users per month.[17]

---

11   Cf. Endress/Gutas (1992).

12   OCR (Optical Character Recognition) and ICR (Intelligent Character Recognition) were not options, as all file cards were filled in manually, with many individual handwriting styles.

13   Cf. Arnzen et al. (2012).

14   The Glossarium Graeco-Arabicum is hosted by the Berlin-Brandenburg Academy of Sciences since 2008.

15   Cf. Roeder (2015).

16   Most recently (2010–2015) under the ERC project "Greek into Arabic. Philosophical and Linguistic Bridges" (Advanced Grant 249431, http://greekintoarabic.eu). The results were presented, amongst other venues, at the international workshop "Plotinus East and West. The Enneads in Arabic and Latin" (Pisa, November 3–6, 2014) by Yury Arzhanov, Gerhard Endreß and Torsten Roeder.

17   As analyzed by Piwik (http://piwik.org), usage data from August 2014 to December 2015.

1.3. Towards Comparative Translation Studies

The availability and the further development of the database's content and functionality is currently guaranteed for the medium term at least, mainly due to a new project titled "Transmission of Classical Scientific and Philosophical Literature from Greek into Syriac and Arabic" which was launched in mid-2016.[18] The planned database "HunaynNet" (named after Hunayn ibn Ishaq, one of the most famous Christian-Arabic translators) will also take earlier Syriac translations into account (cf. figure 4). That those Syriac translations influenced later Arabic versions is a general assumption that still needs to be confirmed. To investigate the question of the role of Syriac in the transmission process and to tackle the problem of continuity from the period of Syrian to Arabic translations, HunaynNet will integrate all records from the Glossarium Graeco-Arabicum to form a trilingual research platform for translation studies. One declared aim of HunaynNet is to reveal how the Syriac translations, which were used primarily by Syriac Christians, might have influenced or prepared the prosperity of the Islamic sciences. The project intends to design the underlying database flexibly enough to incorporate further relevant languages, e.g. Latin, for future projects.

Language	Greek	Syriac	Arabic
Lexeme	λόγος	ܡܠܬܐ	قول
Root / Stem		m l l – – / –	q w l – – / –
Part of Speech	noun	noun	noun
Expression	ὁ λόγος κατὰ τοῦ ὑποκειμένου	ܡܠܬܐ ܚܠܦ ܗܘ ܕܣܝܡ ܣܝܡ	القول يحمل على الموضوع
Translation	definition	definition	definition
Annotation	—	—	—
Quotation	ὥστε καὶ τοὔνομα καὶ ὁ λόγος κατὰ τοῦ ὑποκειμένου κατηγορηθήσεται	ܡܕܝܟ ܢܩܒܠ ܘܡܠܬܐ ܚܠܦ ܗܘ ܕܣܝܡ ܣܝܡ ܕܐܟܕܟ	فيكون الاسم والقول يحملان على الموضوع
Reference	Arist. Cat. 2a26	ed. King, 100.35	BN 160a16
Lexica	Search in Perseus		Search in Perseus

Figure 4: A preliminary example of a tri-lingual entry in the future database HunaynNet, displaying Syriac and Arabic translations for λόγος.

---

18 The project is directed by Dr. Grigory Kessel, hosted at the Austrian Academy of Sciences (ÖAW) and funded for five years as an ERC Starting Grant, cf. ÖAW, "Vier neue ERC-Starting Grants", November 24, 2015 http://www.oeaw.ac.at/oesterreichische-akademie-der-wissenschaften/die-oeaw/article/vier-nachwuchswissenschaftlerinnen-an-der-oeaw-erhalten-erc-starting-grants-copy-1/ (April 29, 2016).

1.4. Visualization and Interpretation of Big Data

As a collective of pioneer disciplines, the Digital Humanities seek technologically supported alternatives to traditional philological research methods, in order to gain evidence of previously unrecognized or unconsidered phenomena by using digital techniques.[19] This includes also exploratory approaches, e.g. such as "serendipity"[20], which are less appreciated by the humanities, since they are not guided by any predefined, specific interest. Indeed, quantitative analyses have to accept the criticism that their primary results consist of purely numerical statements. However, the oceanic quantities of digitally available information often does not suit hermeneutic approaches, so that alternative methods need to be evaluated, even if eventual insights are impossible to foresee.

Also in the Glossarium Graeco-Arabicum, the steadily increasing amount of data makes it increasingly difficult to gain an overall view of the underlying corpus, or even of excerpts. Although typical database mechanisms, providing search forms and result lists, cover basic requirements, this type of functionality suffices only in rare, special cases: When searching over a corpus of more than 100,000 records, a general search query would in most cases lead to a vast amount of results, unless the user is able to narrow down the results. Thus, the conventional functionality will no longer suffice in the long term, especially while the database is still growing: With extending size, the content of a database becomes more and more opaque for the user, and it becomes increasingly unclear how to evaluate individual information within the context of the whole corpus.

Therefore, in order to provide a solid foundation for research, it is necessary to present the material in alternative ways. Particularly, visually oriented interfaces offer a great opportunity in this respect, as they are able to depict a variety of semantic levels simultaneously, and they offer an attractive method of access for academics or scholars as well as the interested public. A good visualization can give immediate insights to the user into the relationship between data structures and content, and may even give decisive clues to new interpretive possibilities. On this basis, it is possible to generate new impulses for research in the humanities, even from large amounts of data.

---

19   Cf. Oakes/Ji (2012).

20   Cf. Thudt et al. (2012).

2. Method

2.1. »Mapping« the Words?[21]

Figure 5: Astronomers working, from Shāhanshāhī-nāmah, 10th century AH (16th century AD).[21]

The illustration (cf. figure 5) shows an astronomical laboratory in Istanbul in the 16th century. Scholars are handling all sorts of scientific tools, such as quadrant, astrolabe, hourglasses and celestial globe, and transmitting their results on maps and in journals. With their devices, they try to measure and to depict the happenings in the sky, in order to better understand and interpret the movements of the heavenly bodies. This variety of monitoring tools and the detailed recording of the results act here as crucial prerequisites for understanding complex processes. Knowledge production, as pictured here, is thus a process of measuring, mapping and interpretation. Mapping assumes a special role in this process, since it mediates between data collection and measurement on the one side and interpretation and understanding on the other side.

The Glossarium Graeco-Arabicum aims primarily at translation studies, as introduced before, and consists mainly of a large amount of Greek-Arabic word pairs, each associated with their

---

21 University Library Istanbul, MS No. FY 1404. Reproduced and cited from: Nasr (1976), 113, plate 65 (detail). Any copyrights are to be observed.

sources and translators. Geographically, the Arabic side of the corpus concentrates on the area of Baghdad, because in this region the richest production of translations took place. In this respect, the title "Mapping the Words" does not aim at geographical, or even – as the presented image could suggest – at astronomical dimensions, but rather on linguistic structures and their alterations through translation processes. This is the point where visualizations come into play: Graphical illustrations should allow the viewer to recognize (maybe even to "read") structures and information like on a geographical map.[22] Then, in the best case, such maps could form an independent reference for research.

However, some crucial questions remain. In which way should visualizations be designed to support research questions in an actually useful way? Which methods of multidimensional representation are available? In which state does a visualization supply more information, or display it more clearly, than a textual results list of a database query? Do visualizations have the potential to provide a reference for research at all, or are they merely a simplistic view?

Visualizations are not per se useful or useless, but much depends on the quality of its design. Edward Tufte, in his monograph "The Visual Display of Quantitative Information", formulated some guidelines for the design of visualizations under the heading "Principles of Graphical Excellence"[23] and pointed out the following key aspects:

- The informative value of visualization increases with the substantial dimensions displayed and correlated at the same time.
- A good visualization provides information compactly without loss of transparency and diversity.

According to this, formal and structural design is actually a central aspect of visualization. When thinking of the users of a visualization, it seems crucial to explain not only the material, but also the method of presentation. From this perspective, graphical style is less relevant when utilized for visual effects, as well as for suggestive design, which puts only prefabricated statements in the foreground. It is, instead, crucial to raise the viewers' interest by simultaneously displaying possibly correlated information and by doing this transparently, to leave interpretation ultimately to them. This reflects the idea of mapping as demonstrated in the above picture.

The following chapters will exemplify this idea in four use cases from the Glossarium Graeco-Arabicum. As a general strategy, the following aspects need to be clarified before implementing a visualization:

- which specific aspect to investigate into, and which research interest to follow,
- which dimensions are necessary for this, and which correlations are expected in this context,
- which visualization strategies are adequate for this use case, and how the display can be optimized for the viewer.

2.2.   Defining Topics and Interests

The development of visualizations for the Glossarium Graeco-Arabicum should support the following aspects of the database:

1. How can an overview of the whole corpus be created, in order to understand the chronological arrangement of the texts and their respective relevance? Here, the scope of the texts needs to be correlated with the history of sources and their translations. This is

---

22   This was demonstrated e.g. for a historic speech corpus of English; cf. Alexander (2010).

23   Cf. Tufte (2001), 51.

of particular importance for new users, who are not yet familiar with the database corpus (cf. section 3.1: Corpus Overview).

2. Which lexical differences exist between sources and their translations in general? Do translations use a broader vocabulary and tend to circumscribe phrases with more words, or do they tend to synthesize information? Here it would be necessary to compare the lexical variance of source and target language (cf. section 3.2: Terminological Consistency).

3. How did the texts transform on the grammatical level? How did the translators react to the structural differences between Greek and Arabic? Which differences exist among the translators? Which Greek parts of speech transformed into which Arabic ones, and which differences become visible by comparing authors or sources to one other (cf. section 3.3: Part of Speech Transformations)?

4. How can the origins and translation variants for a specific lemma, including the relevance of authors and translators in the process, be displayed at a glance? The aim would be to show how the various translators treated a specific word in correlation to its usage by different authors or in different sources (cf. section 3.4: Translation Variants).

2.3. Choosing Visualization Techniques

After having defined the topics and the specific research interests, it is the right time to decide for a data visualization. Which visualization techniques are actually adequate to map the above-mentioned aspects of the database? Obviously, typical two-dimensional diagrams, e.g. bar charts or pie charts, will not suffice to depict the mentioned correlations without compromising the subject's complexity. Therefore, the following approaches favor visualizations that depict more than two parameters simultaneously.

The implementations presented in the next chapter will apply three basic models of visualization:

- Coordinate systems, which represent information usually as two-dimensional points that are further qualifiable by color, size and shape, resulting in five simultaneous dimensions (or even more, if e.g. different patterns or line widths would also be applied).
- Tree maps, which represent information as areas of quantity-related size, arranged in a way that reflect hierarchic relationships.[24] Colors, forms and patterns are suitable to display further dimensions.
- Flowcharts, which illustrate multistage processes, and amongst these especially Sankey diagrams[25], which depict quantitative connections between nodes.

From the variety of visualization approaches shown here, it becomes evident again that the decision for a suitable visualization is already determined by the data itself and by the intention of which aspect to investigate. It also becomes clear that visualizations not only reflect quantities, but also qualities like typologies, relationships, and processes.

---

24   For a historical overview and an introduction to the various algorithms, cf. Shneiderman/Plaisant (2009) and Leydesdorff/Welbers (2011).

25   Named after the Irish engineer Matthew Henry Phineas Riall Sankey (1853–1925), who developed a graphical representation of simultaneous data flows that displays direction and quantity at the same time. A collection of application examples can be found on the website "Sankey Diagrams. A Sankey diagram says more than 1000 pie charts", http://www.sankey-diagrams.com (April 29, 2016).

3. Implementation

On the technical level, all following implementations rely on the API *Google Chart Tools*[26], which provides a highly reliable, sufficiently flexible and no-cost way to create visualizations on a web site.

3.1. Corpus Overview

The researcher looks at the recently acquired encyclopedia in the bookshelf. Intuitively, he recognizes the scope of this publication in comparison to other books on the shelf. He takes one of the volumes and browses through the pages, receiving a quick insight into the content structure. Most databases do not provide an equivalent possibility: In many cases, the information about scope and structure of a database corpus remains hidden, and overviews for the users are rarely implemented. However, especially in large corpora, this entrance procedure is of great importance, as it can guide the user's expectations into the right direction before research takes place. For the diversified corpus of the Glossarium Graeco-Arabicum a tree map was implemented to provide exactly such a preliminary overview, to show to the user what he can actually expect to find in the database.
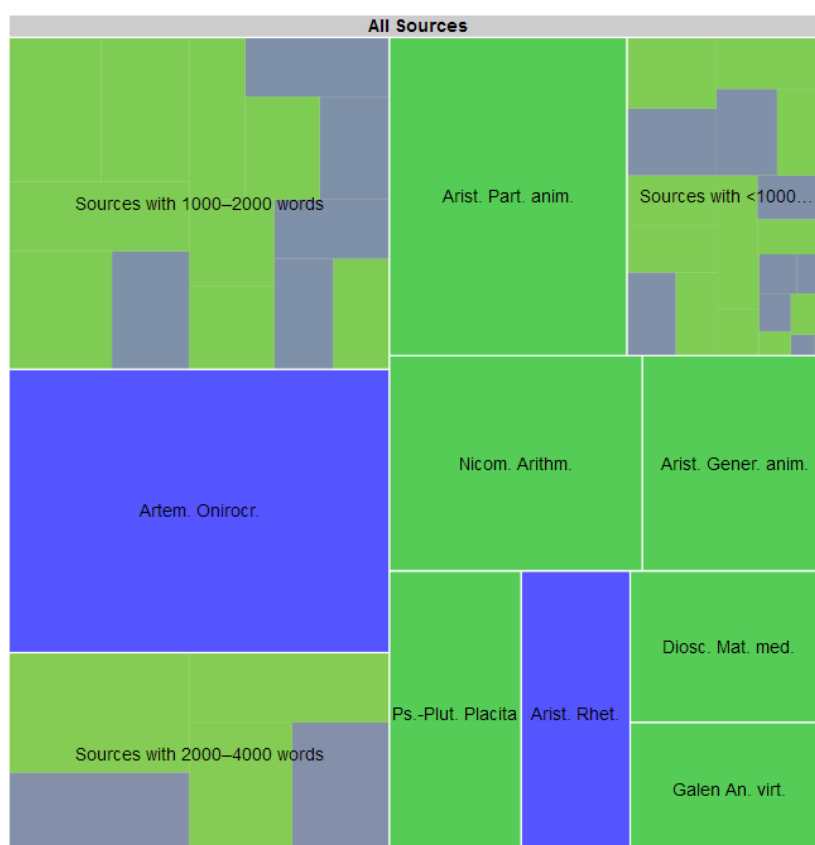

Figure 6: Glossarium Graeco-Arabicum, Corpus Treemap by Source.[27]

---

26 Google Chart Tools, version 44, February 23, 2016 https://developers.google.com/chart/interactive/docs/release_notes#-february-23-2016 (April 29, 2016).

27 https://telotadev.bbaw.de/glossga/source_charts.php?chart_type=corpus-treemap (screenshot: January 16, 2016).

The tree map (cf. figure 6) is a display of all 76 source texts of the corpus at the same time. Each area in this visualization represents one source, the size being proportional to the number of recorded words. For a better overview, the sources were divided into three groups of size (2,000–4,000 words, 1,000–2,000 words and less than 1,000 words). It is recognizable that about one-half of the corpus is determined by eight sources with a very high number of records. It should be noted here that the number of recorded words does not necessarily correspond to the actual length of a source's full text, because some sources are only represented by a representative selection of words.

In addition to the size, a color code provides information on the translator: green indicates a source with a known translator, while blue represents an unknown translator. The tree map demonstrates that the vast majority of translators is known, especially of the more extensively recorded sources; overall, the translators are known in 46 of 76 cases (about 60%).
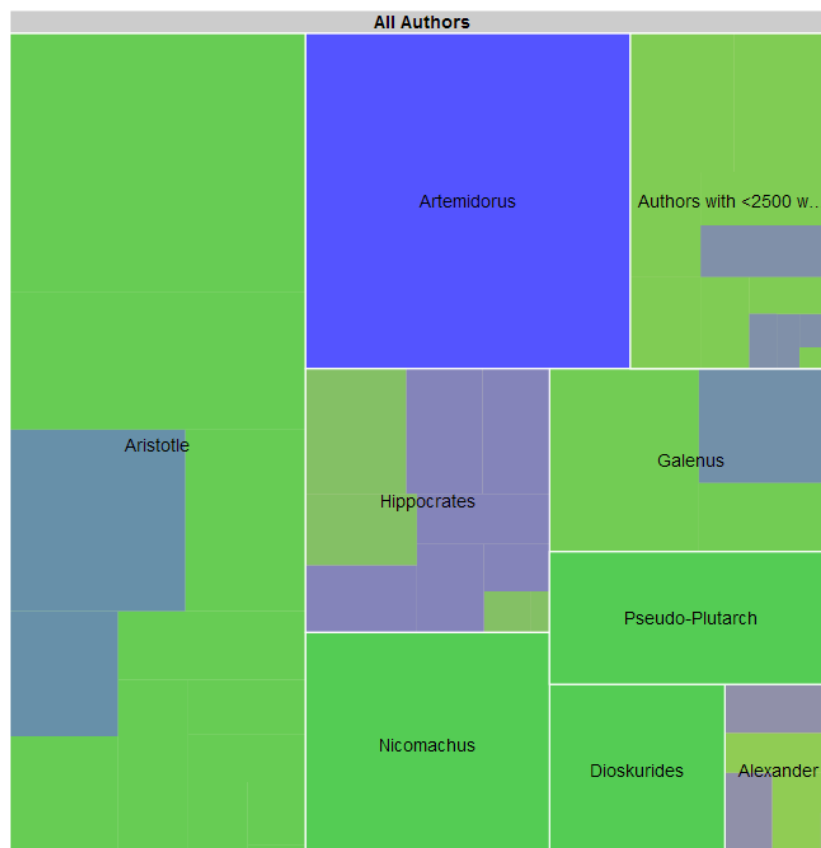

Figure 7: Glossarium Graeco-Arabicum,Corpus Treemap by Author.[28]

The alternative arrangement of the tree map by authors (cf. figure 7) basically gives a similar impression, but the information is focused on the relevance of the authors. Similar to the previous example, the image is determined by a few large areas, which is even more extreme in this case. However, it also becomes clear that e.g. Aristotle and Galen are represented by a large variety of texts, while Artemidorus, the author of the most comprehensive text, is represented only by that one single source. In addition, it becomes apparent that some translators, mainly those of Artemidorus and Hippocrates, are not yet identified. By clicking on the area, all texts related to that author are displayed as a tree map again (like a zoom-in on the corpus tree map).
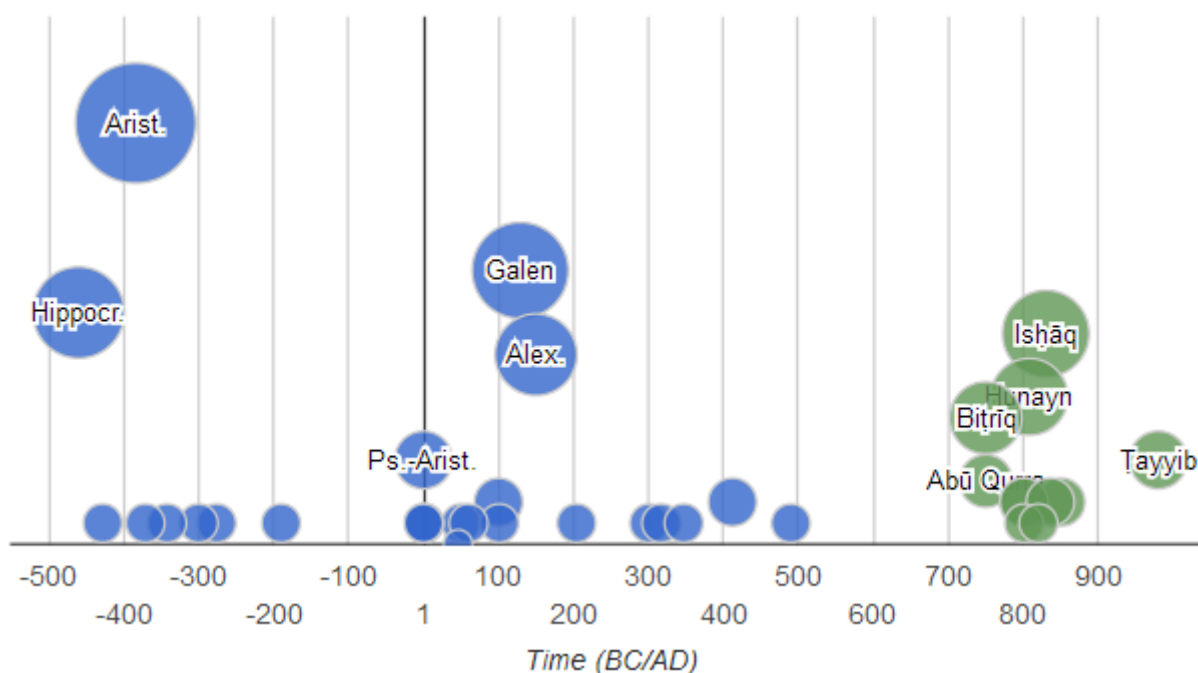

Figure 8: Glossarium Graeco-Arabicum, Author and Translator Time-map.[29]

For a chronological perspective that also takes the relevance of the respective author or translator for the corpus into account, a "Balloon Chart" was designed (cf. figure 8).[30] Each circle on the underlying timeline represents one person, while size and vertical position are proportional to the number of corresponding sources. This produces a clustering effect in the most important periods, while the more relevant authors and translators stand significantly out through the balloon effect, supporting legibility.

3.2. Terminological Consistency

When comparing texts and their translations, it is natural that some translators tend to differentiate the original terminology, while others make more use of generalizations. While a translation ideally would maintain terminological consistency as well as possible, a special interest of the translator, possibly related to the personal background, could determine a different

---

29   https://telotadev.bbaw.de/glossga/author_charts.php?chart_type=timeline (screenshot: January 16, 2016).

30   This visualization type is related to the so-called poppy chart, implemented by Valentina D'Efilippo and Nicolas Pigelet on http://www.poppyfield.org (April 29, 2016).

terminological rendition in the translation. The question at hand is whether such a tendency existed when the Greek texts were translated into Arabic, and to what extent this depended on translators or even on the authors. A first approach to answering this question is comparing the numbers of distinct words (counting each word only once, regardless of how often it is used) in each source and its respective translation. A comparison of these numbers could indicate to which extent the lexical diversity of the original text was modulated by the translation.
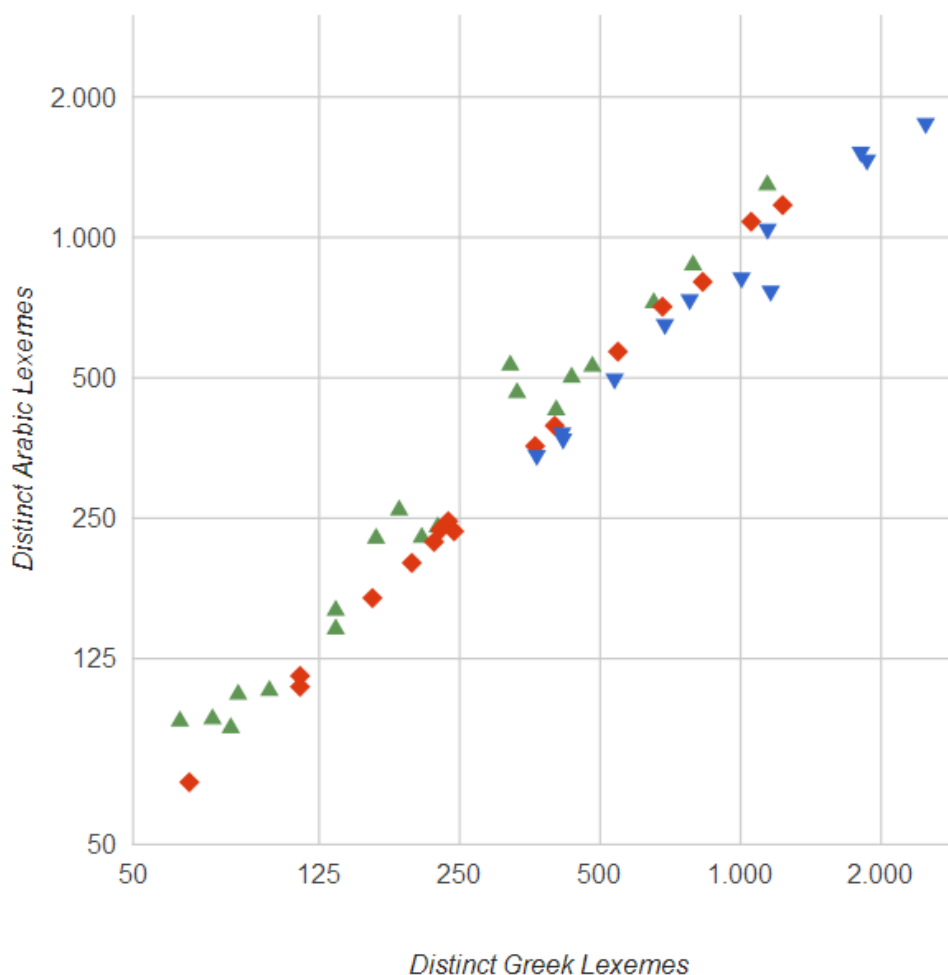

Figure 9: Glossarium Graeco-Arabicum, Distinct Lexemes by Source.[31]

This diagram (cf. figure 9) will help to compare this phenomenon throughout the whole corpus. Each shape corresponds to one text. The x-axis represents the number of distinct Greek lexemes, while the y-axis corresponds to the same in Arabic. Due to the numerical distribution, a logarithmic scale is applied. An exactly equal distribution of distinct lexemes (inflections ignored) in the Greek source and the Arabic translation would appear on the center diagonal; a trend to generalization in the Arabic text would result in a position toward the lower right, while the opposite case, a trend to diversity, would result in a position toward the upper left. In addition, a shape/color coding is used: A variation of less than 5% is highlighted as a red diamond (◆), while higher deviations are marked as a green triangle pointing up (▲) for more diversity in Arabic, or a blue triangle pointing down (▼) for the opposite case.

First, it becomes clear that the majority of translations tend towards a more diversified terminology. Yet, this trend is visible mostly in the lower half of the diagram (less than 500 words),

31   https://telotadev.bbaw.de/glossga/source_charts.php?chart_type=-distinct-lexemes (screenshot: January 16, 2016).

while it is balanced in the middle range (500–1,000 words) and reverted in the upper region (more than 1,000 words). However, this does not imply that texts with more recorded words are terminologically less diverse: In larger texts, a disproportional number of overlaps levels the clear tendency towards differentiation in smaller texts, and thus creates a statistical counter effect.
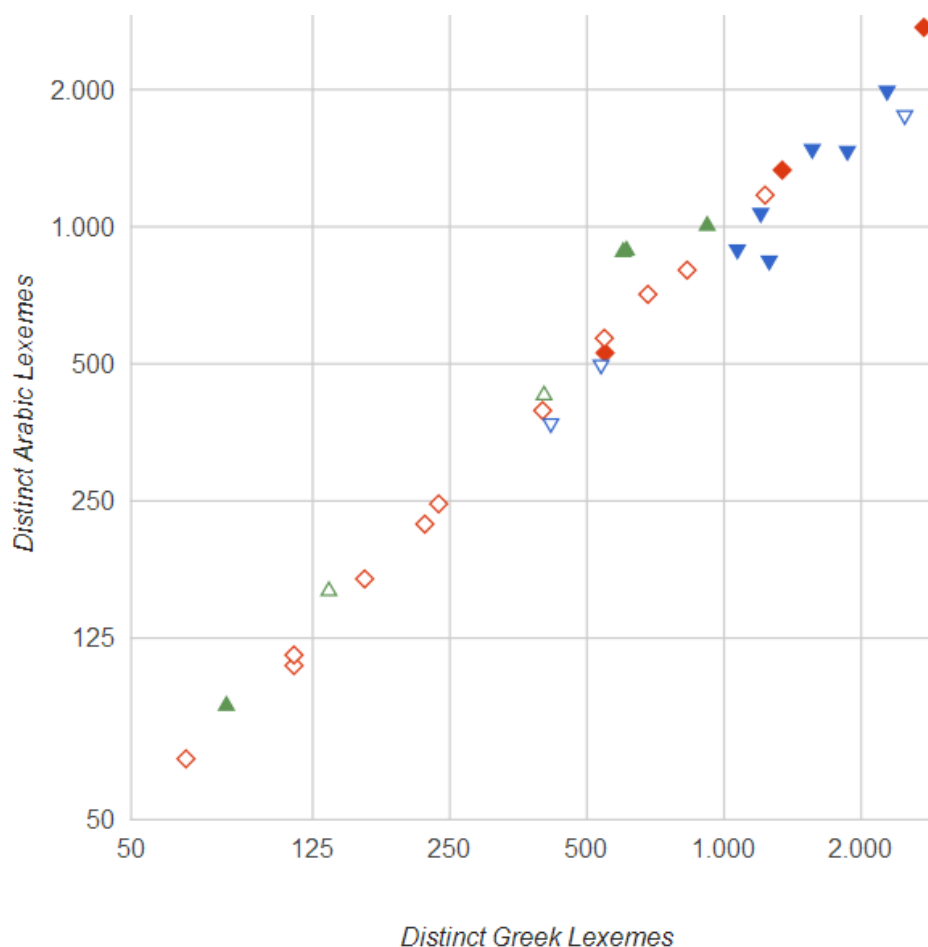

Figure 10: Glossarium Graeco-Arabicum, Distinct Lexemes by Translator.[32]

A diagram variant which displays the same setting, now focused on translators (cf. figure 10), demonstrates that this is not only a random spike in the statistics, but also confirms the latter assumption. As all texts by one author are merged into a larger group, resulting in some very large sets of words, a clear tendency towards generalization is visible; and this occurs despite a contrary tendency in the single texts.

A real statistical glitch is instead caused by the texts *De virtutibus et vitiis and Divisiones quae vulgo dicuntur Aristoteleae*, which both display a very high lexical diversity in Arabic. This can be explained by the fact that this text was translated two times, namely by Theodore Abū Qurra (ca. 750–ca. 823) and later by Ibn al-Tayyib (980–1043).

Overall, these observations give some important insights about how to interpret a statistical statement on terminological consistency: It is relative to the total amount of compared words, while the absolute ratio does not serve as an indicator. However, future experiments should take this into consideration to normalize the statistical behavior. Additionally, an abnormally high divergence from the average points very probably towards special conditions that are unrelated to a single author or translator.

---

32 https://telotadev.bbaw.de/glossga/author_charts.php?chart_type=distinct-lexemes (screenshot: January 16, 2016).

3.3. Part of Speech Transformations

This example investigates linguistic structures and their transformations through the translation process. The grammars of Greek and Arabic are significantly different from each other, so that in many cases the part of speech of a word had to be changed. For example, a direct equivalent of the Greek gerund does not exist in the Arabic grammar, forcing a translator to choose another part of speech. Additionally, the idiomatics of a language and the interpretative function of the translator exert additional strong influences on part of speech transformation. This creativity in the translation process can also be visualized for (or communicated to) users who do not possess expertise in either or both of the languages.
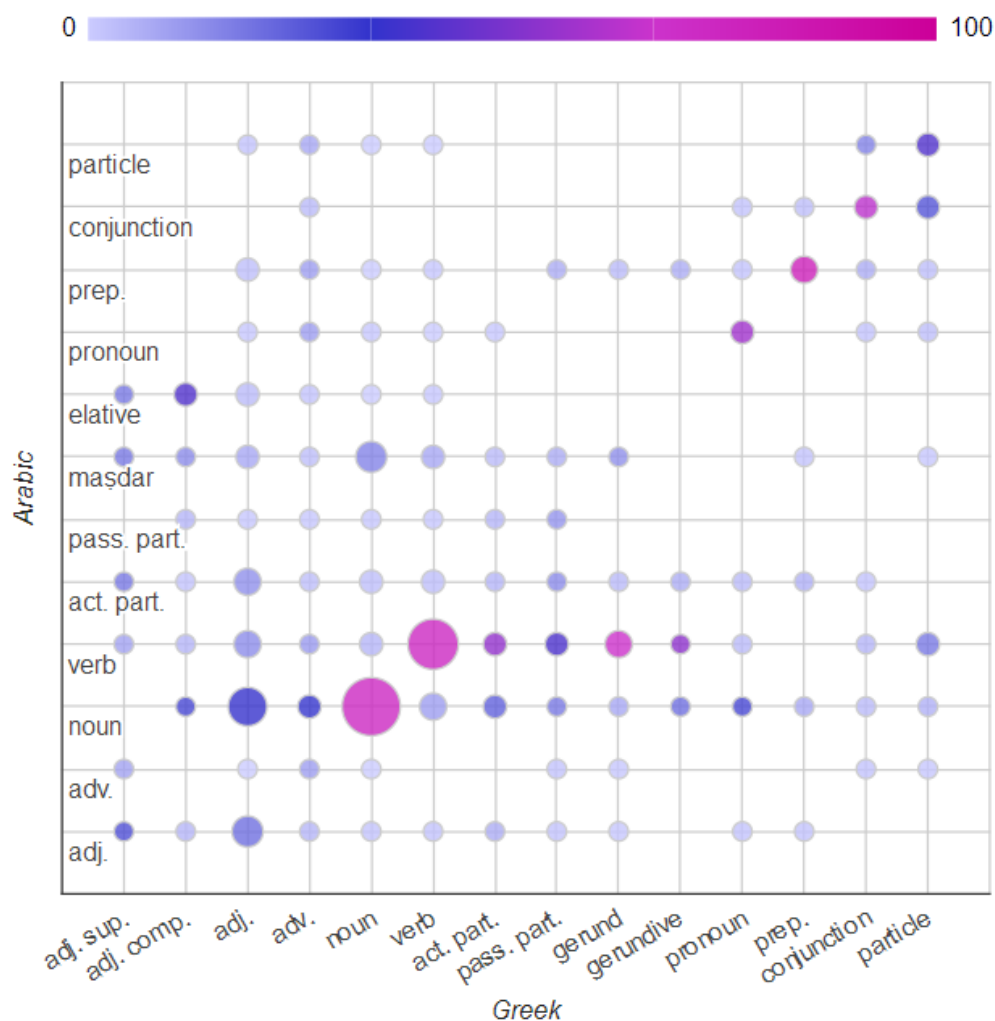

Figure 11: Glossarium Graeco-Arabicum, Compared Parts of Speech for Ḥunayn b. Isḥāq.[33]

The diagram (cf. figure 11) correlates Greek parts of speech (on the x-axis) to Arabic parts of speech (y-axis) based on the frequency of transformation, in this case focused on the translator Ḥunayn b. Isḥāq (808–873 AD). The size of each intersection represents the absolute frequency, while the color reflects the relative frequency for the respective part of speech. It becomes visible, for example, that Hunayn transformed Greek gerundives and gerunds most frequently into finite verbal forms; and among the parts of speech that exist in both languages, he transformed adjectives often into nouns, and nouns sometimes into maṣdar (a nominal form that

---

33 https://telotadev.bbaw.de/glossga/author_charts.php?person_id=100039&chart_type=compared-pos-grid (screenshot: January 18, 2016).

has similar qualities to gerund). Among the more syntactic parts of speech (pronouns, prepositions, conjunctions, particles), there is a lot of transformation within this group, as the "island" in the upper right corner displays. Generally, there is also a strong tendency of transformation into nouns or verbs, as can be seen in the strong accumulations on the respective horizontal lines. Furthermore, the visualization offers a link to the underlying data: A click on an intersection leads to a list of all respective word pairs in order to investigate them on a more detailed level.
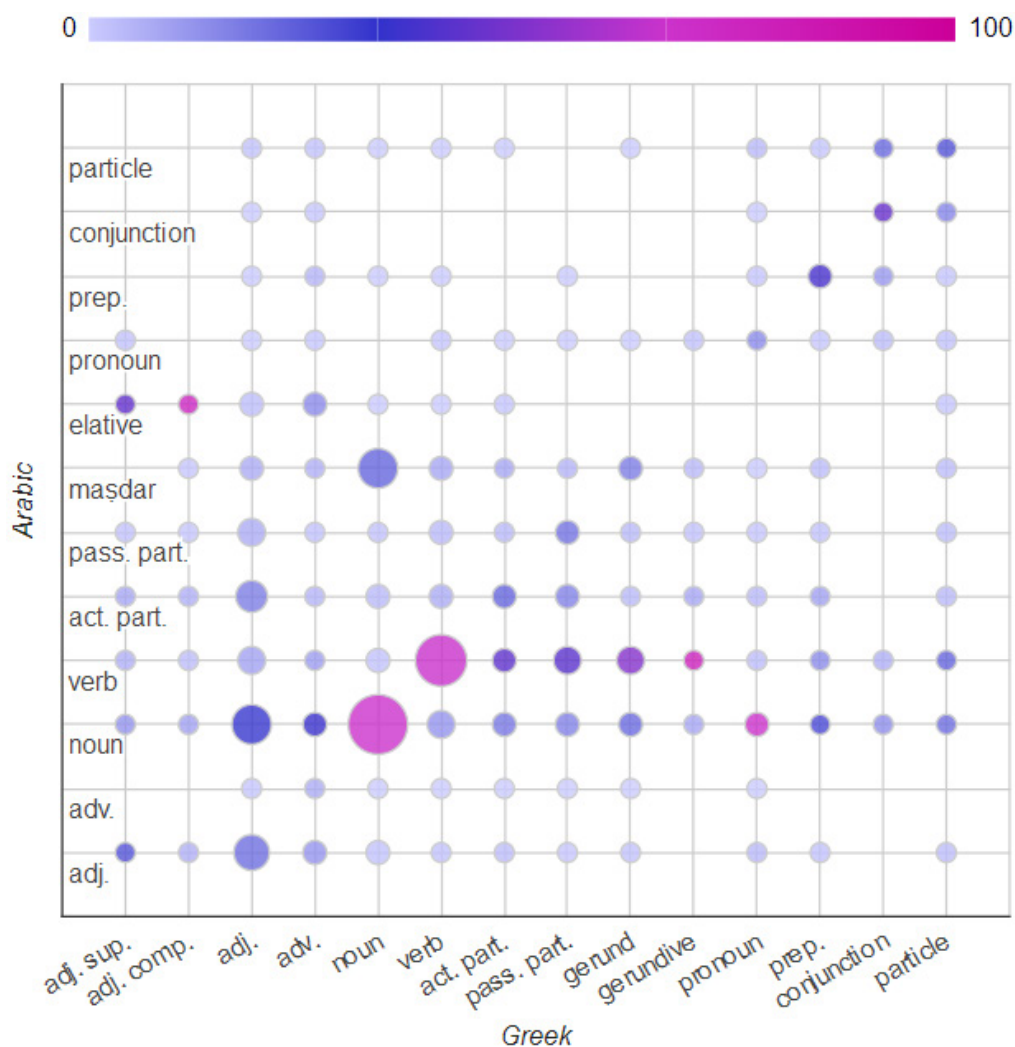

Figure 12: Glossarium Graeco-Arabicum, Compared Parts of Speech for Isḥāq b. Ḥunayn.[34]

The diagram for Isḥāq b. Ḥunayn (son of Ḥunayn b. Isḥāq, 830–910 AD, cf. figure 12) is significantly different: The "syntax island" in the upper right is less recognizable, while the tendency towards noun and maṣdar is much stronger, as well as the frequency of elative. This reveals that part of speech transformation is potentially a distinct characteristic of a translator.

---

34 https://telotadev.bbaw.de/glossga/author_charts.php?person_id=100026&chart_type=compared-pos-grid (screenshot: January 18, 2016).

3.4. Translation Variants

The last example shifts from the quantitative and grammatical to the semantic level, focusing on the translation processes of individual words. It belongs to the most important tasks of a bilingual dictionary to depict the variety of translation possibilities. The reason for variety is partly due to the different contexts in which a word is used, and secondly, to the interpretation of the translators. To start a deeper investigation of this issue, the example experiments with a Sankey diagram.

Figure 13: Glossarium Graeco-Arabicum, Translation process scheme for ἄνθρωπος.[35]

Read from left to right, the diagram (cf. figure 13) shows the translation process for the word ἄνθρωπος (anthropos, "man"). It appears in various texts by Aristotle, Artemidorus, Galen and Hippocrates, most frequently (as seen in section 3.1) in the extensively represented sources *Oneirocritica* and *Analytica posteriora*. In most cases it is translated with أنس (ʾns, "human"), but also often with رجل (rǧl, "man").[36] Apparently, the translators Biṭrīq and Ḥasan covered exactly these two cases, while Hunayn preferred the variant أهل (ʾhl, "people, group"). The variants من (mn, "someone") and قوم (qwm, "crowd") were particularly used by the unknown translator of *Oneirocritica*.

Within this diagram, the user can click on a node (i.e. a vertical bar) to highlight the corresponding connections to this node. When looking at the nodes on the right side, a greater diversification of variants أنس (ʾns, "man") and رجل (rǧl, "man") can be noticed, while the other variants, in almost every case, derive from only one translator each. This suggests that the latter versions are individual cases, due to a special meaning in the original text or a special interpretation by the translator.

---

35   https://telotadev.bbaw.de/glossga/glossary.php?gr_lexeme=ἄνθρωπος (screenshot: January 17, 2016).

36   The transliteration is performed in accordance with the rules of the DMG; cf. Glossarium Graeco-Arabicum, "Transliteration" https://telotadev.bbaw.de/glossga/transliteration.php (March 5, 2017).
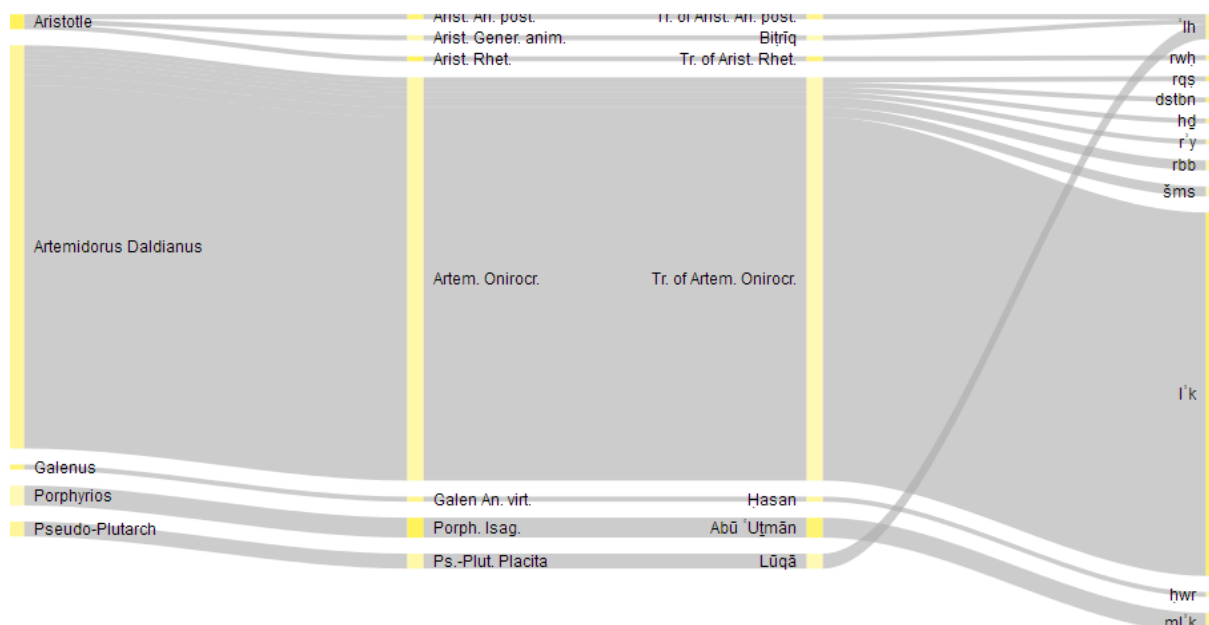
Figure 14: Translation process scheme for θεός.[37]

For a comparison, the Sankey diagram for the word θεός (theos, "god") is considered (cf. figure 14). Here, the scheme displays a completely different scenario: The translation of *Oneirocritica*, in which θεός translates most frequently into ملك (l'k, "angels"), dominates the diagram. This is a monotheistic interpretation of the Greek word "god", in the original context used in a plural form, degrading the Greek "gods" to Arabic "angels" (all under one God). For "god" in a singular meaning, which is less frequent in the Greek texts, also the today familiar Arabic form الله ('lh, "Allah") can be found, though it occurs mainly in the translations of Pseudo-Plutarchian and Aristotelian writings.[38]

However, Sankey diagrams have their limitations, as they produce less clarity and much more confusion in cases that are more complex: A display of a huge amount of variants would result in an unreadable, giant, spaghetti-like diagram. Data aggregation, clustering similar variants into groups, would help in such cases. Another desideratum in this case is the inclusion of chronological data; however, the chosen API is not yet sufficiently configurable (whether it will provide these options in one of the future releases remains an open question).

4.   Results

The four presented examples demonstrate from very different angles how to visualize translation structures on various application levels, and how these visualizations can help to investigate various areas of research interest. Concerning corpus structures, lexical and grammatical differentiation, grammatical transformations and translation processes as well, visualizations help to identify phenomena that would not have been recognizable and comprehensible though a textually oriented interface. These revealed structures raise new questions regarding the material. Especially for a complex and large data collection like the Glossarium Graeco-Arabicum, which already tends to grow beyond human understanding, visual "distant reading" seems a fruitful approach to get a fresh perspective on the material.

---

37   https://telotadev.bbaw.de/glossga/glossary.php?gr_lexeme=θεός (screenshot: January 18, 2016).

38   GALex, however, gives still more records for "Allah" and covers more sources.

In addition, those structures are recognizable both for professionals as well as for amateurs, and thus fulfill both a professional and didactic purpose, from the corpus overview down to the single lexemes: Nonprofessionals can make a quick start through visual aids, while an expert can detect special phenomena, which could be worth a closer investigation. In all cases, the visualizations provide a good degree of transparency, as the underlying data is accessible through a simple link on the interactive display. Finally, yet importantly, even the database editors can profit from the visualizations, as they give an opportunity to check the integrity of the data and the consistency of the whole corpus.

Visualizations create impetus and inspiration through their direct, immediate effect. It is, however, crucial that such presentation forms provide also sufficient information about the database structure and the data representation. If those structures are not properly explained, there is a high risk that visualizations suggest correlations that later reveal as fallacies. Transparency is as a key here: Misunderstandings are avoidable by accurately documenting and explaining structures and collection methods of a corpus. Still, visualizations – understood in the meaning of mappings – will neither explain nor interpret anything about themselves: Instead, they recite data in their very special way, which still needs to be interpreted by the user, and explanations can only be given based on in-depth studies. Visualizations can support this in their function of a mediator, as long as it is clearly understandable what is actually being displayed and how it is displayed. The more consistently they do this, the more they will become a compact graphical narration and actually get closer to the ideal of becoming an independent scholarly media form.

5. Literature

Alexander (2010): Marc Alexander, "The Various Forms of Civilization Arranged in Chronological Strata. Manipulating the Historical Thesaurus of the OED", in: Michael Adams (ed.), Cunning passages, contrived corridors. Unexpected Essays in the History of Lexicography, Monza, 2010.

Arzhanov/Roeder (2013): Yury Arzhanov and Torsten Roeder, "The Glossarium Graeco-Arabicum. Linguistic Research and Database Design in Polyalphabetic Environments". Lecture at the Digital Classicists Berlin Seminar on November 19, 2013. http://hdl.handle.net/11858/00-1780-0000-0022-D548-B (permalink).

Arnzen et al. (2012): Rüdiger Arnzen, Yury Arzhanov and Gerhard Endress, "Griechische Wissenschaft in arabischer Sprache", in: RUBIN Wissenschaftsmagazin, spring 2012, 14–21. http://rubin.rub.de/de/griechische-wissenschaft-arabischer-sprache (March 5, 2017).

D'Ancona (2013): Cristina D'Ancona, "Greek Sources in Arabic and Islamic Philosophy", in: Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, 2013. http://plato.stanford.edu/entries/arabic-islamic-greek/ (March 5, 2017).

Endress/Gutas (1992): Gerhard Endress and Dimitri Gutas, A Greek and Arabic Lexicon (GALex), Leiden, 1992ff.

Endress et al. (2013): Gerhard Endress, Rüdiger Arnzen and Yury Arzhanov, "Griechische Wissenschaft in arabischer Sprache. Ein griechisch-arabisches Fachwörterbuch der internationalen Wissensgesellschaft im klassischen Islam", in: Studia graeco-arabica 3 (2013), 143–156. http://www.greekintoarabic.eu/uploads/media/BOCHUM_SGA_3-2013.pdf (March 5, 2017).

Leydesdorff/Welbers (2011): Loet Leydesdorff and Kasper Welbers, "The semantic mapping of words and co-words in contexts", in: Journal of Informetrics 5, No. 3 (July 2011), 469–475.

Lüdeling/Zeldes (2009): Anke Lüdeling and Amir Zeldes, "Three Views on Corpora: Corpus Linguistics, Literary Computing, and Computational Linguistics", in: Jahrbuch für Computerphilologie 9 (2009), 151–180. http://computerphilologie.tu-darmstadt.de/jg07/luedzeldes.html (March 5, 2017).

Nasr (1976): Seyyed Hossein Nasr, Islamic Science. An Illustrated Study, World of Islam Festival Publishing, 1976.

Oakes/Ji (2012): Michael P. Oakes and Meng Ji (ed.): Quantitative Methods in Corpus-Based Translation Studies. A practical guide to descriptive translation research (= Studies in Corpus Linguistics 51), Amsterdam/Philadelphia, 2012.

Roeder (2015): Torsten Roeder, "Alpha into Alif. Schnittstellen zwischen Schriftkunde und Informatik am Beispiel von Unicode im Glossarium Graeco-Arabicum", in: Studia graeco-arabica 5 (2015), 345–363. http://www.greekintoarabic.eu/uploads/media/14-SGA_V_2015_ROEDER.pdf (March 5, 2017).

Roeder (2016): Torsten Roeder, "Mapping the Words. Experimentelle Visualisierungen von Übersetzungsstrukturen zwischen Altgriechisch und Hocharabisch", in: Zeitschrift für digitale Geisteswissenschaften. http://dx.doi.org/10.17175/2016_006 (March 5, 2017).

Roeder/Arzhanov (2016): Torsten Roeder and Yury Arzhanov, "Experimental visualizations of translation structures between Ancient Greek and Classical Arabic", Lecture at the Digital Classicists Berlin Seminar on January 19, 2016. http://hdl.handle.net/11858/00-1780-0000-0029-C04E-B (permalink).

Shneiderman/Plaisant (2009): Ben Shneiderman and Catherine Plaisant, "Treemaps for space-constrained visualization of hierarchies", 2009–2014. http://www.cs.umd.edu/hcil/treemap-history (March 5, 2017).

Thudt et al. (2012): Alice Thudt, Uta Hinrichs and Sheelagh Carpendale, "The bohemian bookshelf: supporting serendipitous book discoveries through information visualization", in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), New York, 2012, 1461–1470. http://www.alicethudt.de/BohemianBookshelf/material/ThudtCHI2012.pdf (March 5, 2017).

Tufte (2001): Edward R. Tufte, The Visual Display of Quantitative Information, Creshire, 2001.

Author[39]

Torsten Roeder
Julius-Maximilians-Universität Würzburg
Institut für Musikforschung
Domerschulstraße 13
D-97070 Würzburg

Email: torsten.roeder@uni-wuerzburg.de