DIGITAL CLASSICS ONLINE

Pain in Classical Greek Texts

Vojtěch Linka, Vojtěch Kaše

Abstract: Texts from the Classical Greek period play a crucial role in the historical development of Western science and philosophy. The concept of pain is of key importance, especially for two areas of human knowledge, namely medicine, and ethics. While the concept of pain is important for both areas, in the majority of scholarship, the areas are studied separately. We approach them together while analyzing the whole extant body of the digitized ancient Greek literature from the 5th and 4th centuries BCE. This is enabled by our methodological framework, which combines traditional interpretative approaches with computational approaches from the area of distributional semantics, making it possible to study a vast amount of textual data in a controlled way. When we look at the context of the usage of individual words denoting pain across the texts covering various genres or topics, we identify relatively stable semantic clusters to which pain words relate, such as pathologies, emotions, or morality. Thus, we can capture the role of pain words, their meaning, and mutual relations in the corpus. Also, our approach enables us to discern the role of various textual subcorpora (philosophical, medical) in how pain was conceived in the period under scrutiny.

Introduction[1]

In Classical Greek philosophy and medicine, pain plays a peculiar role. Even though some medical authors include pain in their definitions of medicine, Aristotle emphasizes its relevance in ethics, Plato discusses its role in a good human life,[2] the reader finds very few explicit passages in which pain is defined or in which its origin, nature, and implications are explained.[3] Also, a reader interested in how the ancients conceived pain faces the question of whether pain was understood as a unique phenomenon by the authors, who use several words for it. Finally, the Classical Greek period (from the mid-5th to the end of the 4th century BCE) is associated with the emergence of three important textual corpora with a tremendous impact on the development of both medicine and philosophy in the subsequent centuries, namely the *Corpus Hippocraticum*, the *Corpus Platonicum,* and the *Corpus Aristotelicum*, the reader may wonder how the general conception of pain was (re)shaped within this period. In this study, based on the nature of the studied area just mentioned, we elaborate tentative answers for the three questions, i.e. the ontology of pain, the taxonomy of pain, and the role of philosophical and medical texts in shaping the discussions about pain in the Classical Greek period. We hope this analysis might improve our understanding of pain in Classical Greek thinking. Methodolo-

1 This work was financially supported by Charles University Grant Agency, project no. 78120, entitled “Aristotle and Hippocrates on Pain”, implemented at the Faculty of Arts of Charles University.

2 See for example Hippocrates, Flat. 1 (6,90 L = 102,1–103,4 Jouanna), Vict. I,15 (6,490 L = 136,27–28 Joly-Byl); Aristotle, Eth. Nic. 1104b3–16, 1152b1–8, 1172a19–25; Plato, particularly Philebus 31d–33b, Gorgias 493d–498a, Resp. 583a–588a, Leg. 732d–734e, Prot. 352d–354e.

3 Plato Phlb. 32a8–d4; Aristotle, Eth. Nic. 3.12, 1119a21–25; Hippocrates, Nat. Hom. 4 (6,40 L = 172,13–174,10 Jouanna).

gically, we approach these questions through a combination of computational text analysis methods of distributional semantics (distant reading) together with more traditional interpretative approaches. Such a hybrid approach is driven by the scarcity of passages explaining what pain is on the one hand and by the extensive textual corpus, which covers all texts written in Greek in the Classical Greek period, on the other. Even though close reading expertise is indispensable for this task, since the computational methods are not always well suited for analyzing individual passages, distant reading methods are the prominent lens through which we approach the studied problem. Our main objective is thus to broaden the general understanding of pain in the Classical Greek period using an innovative set of methods and, at the same time, to show how distant reading methods, when complemented by classical expertise, can contribute to the study of ancient thinking in general. Since it is not possible to analyze the whole body of Classical Greek text by traditional close reading methods, the computational methods we use in this paper enable us to answer questions that would otherwise be left unanswered.

1. The Role of Pain in Classical Greek Literature: An Overview

As already mentioned above, pain is in no way absent from the Classical Greek texts. What we are missing, however, is a theoretical explanation of it. In the medical literature collected in the so-called Hippocratic Corpus (CH), words denoting pain occur frequently.[4] Nonetheless, the medical authors do not express any need to explain what they mean by these terms and how their meaning differs. Pain words usually occur in passages describing pathological conditions that patients suffer from with specifications of where and in what intensity pain is felt.[5] In some more 'theory-laden' treatises of CH, we find pain words in passages explaining the constitution of the human body and emergence of disease, the definition of medicine, or the nature of the physician's profession.[6] While it seems that on the level of individual writings, it is possible to find answers to questions such as "What is the conception of health or disease?" by using traditional close reading methods, in the case of pain this approach appears to be unsuitable. This is also mirrored in the modern scholarly discussion about pain in CH. While some authors emphasize the absence of any 'theory of pain' in the corpus,[7] others try to distill such a theory from reading particular passages.[8] However, the latter approach is problematic in generalizing the findings based on one or a few treatises and applying them to the whole corpus, neglecting the fact that the particular treatises were written by various authors. In our previous study, we faced this challenge in a distributional semantic analysis of CH and demonstrated, for instance, a strong semantic association between pain and pathological states expressed by some pain words (ἀλγ*, ὀδυν*) but not by others (λυπ*, πον*).[9]

In the field of Classical Greek philosophy, pain is usually discussed in the context of ethics. This holds true not only for the two classical figures Plato and Aristotle but for some of the 'pre-Socratic' authors as well, especially Democritus.[10] However, Plato and Aristotle are the main figures shaping the discus-

4 According to Horden, the word ἄλγος appears over 400 times, λύπη 59 times, ὀδύνη 772 times, πόνος over 700. See Horden (1990), 295–315 at 298.

5 See for example Epid. I, 2,6 (II,636 L = 19.1 Jouanna), II,1,11,2 (IV,82 L = 28 Smith), VII,1,11 (VII,382 L = 58,21 Jouanna), III,3,17(12) (III,136 L = 107,7–9 Jouanna), VI,3,20 (V,302 = 230 Smith). V,1,67 (V,244 L = 30,12 Jouanna), Nat. Mul. 2 (VII, 312 L = 3,12 Bourbon), Mul. I,5 (VIII,110 L = 114 Potter), Superf. 38 (VIII, 506 L = 297,9 Bourbon), Acut. A 6 (II,264 L = 43,26–44,10 Joly).

6 Nat. Hom. 4 (6,40 L = 172,13–174,10 Jouanna), Nat. Hom. 2 (6,34–36 L = 168,4–9 Jouanna), Vict. II,66 (6,582–584 L = 188,18–19 Joly-Byl), Med. Vet. 14 (1,602 L = 136,10–16 Jouanna), Med. Vet. 16. (1,606–608 L = 136,10–16 Jouanna).

7 Horden (1990), 295–315.

8 Rey (1995), 17–23, Villard (1998), 124.

9 Linka / Kaše (2021), 54–71.

10 See for example B 189 (Stobaeus III,1,14), B 235 (Stobaeus III,18,35), B 178 (Stobaeus II,31,56), B 179 (Stobaeus II,31,57).

sion about pain by implementing it into their theories of ethics, the good life, happiness, etc. Yet, even though the passages where pain words occur are numerous, a particularly formal and contextual feature occurs: pain words (usually λύπη) in the texts from these authors almost always occur together with the opposite of pain – pleasure (ἡδονή).[11] Both philosophers usually spent more of their time explaining and analyzing pleasure with the occasional indication that pain is its opposite. This approach is to some degree amplified by modern scholarly discussions, in which pleasure is discussed much more deeply than pain.[12] Although this approach was challenged by Cheng or Linka,[13] pain in Plato and Aristotle still fails to receive appropriate attention.

Pain is of course in no way exclusive to philosophy and medicine, as can be shown from its occurrences in other genres of Classical Greek literature, such as tragedy, history, or oratory. In Sophocles' *Philoctetes*, for example, the painful conditions of the main hero are the driving motif of the whole plot.[14] However, pain is not explicitly problematized, although for different reasons than in works by philosophical and medical authors: pain here is a punishment sent from God for Philoctetes' abuse of Apollo's priest Chryses.[15]

What the examples above have in common is the fact that even though the pain words are used abundantly, their authors usually do not feel the urge to provide theory-laden explanations about the nature, origin, or meaning of pain.[16] Instead of focusing on these few passages where they do it, thus, in what follows, we will look at how pain words are used through the lens of distant reading and distributional semantics which will offer more promising results because we no longer have to search passages where the authors explicitly say what pain is and does, but rather to catch the meaning of it in the broader contexts of the textural corpora.

2. Computational Text Analysis

The computational analyses introduced in this article are based on a subset of ancient Greek words from the LAGT dataset.[17] LAGT stands for Lemmatized Ancient Greek Texts, and, as its title implies, the main feature of the collection is that the texts are lemmatized, i.e. they are in dictionary-like form (nouns in nominative singular, verbs in 1st person singular indicative present, etc.). In addition, the lemmatized version of the text is filtered in a such way that it contains only nouns, proper names, adjectives, and verbs. This is because these word categories tend to be the most semantically loaded, and the LAGT dataset is primarily designed to facilitate a semantically focused type of computational analysis of the texts. The texts in LAGT originate from two open-source corpora of ancient Greek texts: the Canonical Greek Literature dataset from the Perseus Digital Library[18] and the First Thousand Years of Greek dataset of the Open Greek & Latin project.[19]

11 See for example Aristotle, Eth. Nic. 1104b3–16; II,5 1105b21–23; 1106b19–2; 1107b4–7, Hist. An. 535a12, 581a30–31, Part. An. 666a11–12; De An. 408b1–6; 409b16–17; 413b23, MA 701b35–36, Sens. 436a10; Plato, Ph. 83d4, Phil. 27d5, Tim. 42a6, Leg. 654d2.

12 Bostock (2000), 143–160; Brodie (1991), 313–365; Frede (2016), 255–275; Frede (1992), 425–463; Gosling / Taylor (1982); Harte (2014), 288–318; Taylor (2008); Owen (1977), 92–103; Wolfsdorf, (2013).

13 Cheng (2015), W. Cheng (2018) 1–25, Linka (*forthcoming* 2023).

14 Budelmann (2014) 443–467, 1325–35.

15 Sophocles, Philoc. 191–6, 1325–35.

16 This claim holds stronger in the case of the 'Hippocratic' writings, in the case of Plato and Aristotle, there are some passages explicitly talking about pleasure. In comparison to pleasure, however, there are scarce.

17 Kase, LAGT (v2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7221150 (last access 21.02.2023).

18 Cerrato et al. (2020).

19 Crane et al. (2020).

It is important to note that the lemmatization of the texts in the LAGT dataset has been produced automatically, using machine learning algorithms. For a vast majority of texts, the LAGT dataset inherits the lemmata from the GLAUx corpus, a recently published dataset of ancient Greek texts using state-of-the-art methods of artificial intelligence for rich morphological annotations of the texts. [20] For texts unavailable within the GLAUx corpus, the lemmatization has been conducted using a more traditional rule-based algorithm developed for an older version of LAGT.

As LAGT is a compilation of other collections of ancient Greek texts, it also inherits some useful metadata concerning the texts. First of all, LAGT employs a canonical reference system for ancient works based on CITE architecture.[21] Using this system, each work in this dataset is represented by a unique identifier, which is shared with other collections of ancient Greek texts, e.g. the Perseus Digital Library or Thesaurus Linguae Graecae. Thus, for instance, the identifier of Aristotle's *Nicomachean Ethics* is "tlg0086.tlg010", where "tlg0086" is the code for Aristotle and "tlg010" is the code for this notional work within a subset of works assigned to him. These identifiers can be used to search for additional metadata in other databases, e.g. the date of origin of a genre of particular texts, which might be useful, especially for some lesser-known works.

As was already mentioned, in this study we focus only on a subset of texts from this large dataset. It is formed by a set of 371 texts by authors assumed to be active either during the 5th or 4th century BCE. As these texts are conventionally treated as the core of the canon of Classical Greek Literature, we call this subset CGL (= Classical Greek Literature) and will henceforth refer to it as such. Before the preprocessing steps described above, the texts consisted of 3,579,690 words; after we applied the filtering, the number of words decreased to 1,828,293. The 371 works are associated with 35 individual author identifiers. For instance, there are 52 works associated with the code "tlg0627" corresponding to Hippocrates. However, as is well known, these texts were produced by several different authors who were active in different periods and engaged in highly diverse writing agendas.[22] Similarly, the case of works under the author code "tlg0086" corresponding to Aristotle includes some works written in the Aristotelian school rather than by Aristotle himself (e.g. *Problemata*). Thus, the exact number of authors within CGL is unknown. However, we can still use these identifiers to sort the texts into some subcorpora. For pragmatic reasons, we find it reasonable to treat all texts associated with "tlg0086" as the *Corpus Aristotelicum*, all texts associated with "tlg0627" as the *Corpus Hippocraticum*, and all texts associated with "tlg0059" as the *Corpus Platonicum*. Our computational analysis of the CGL corpus starts with counting the occurrences of ten words associated with pain in the 5th and 4th century BCE and derived from four different word roots: λυπέω, λυπηρός, λύπη, ἄλγος, ἄλγημα, ἀλγέω, ὀδύνη, ὀδυνάω, πονέω, πόνος.[23] We calculated their frequencies independently for each author, which gave us an overall picture of which terms are typical for each of them (see authors_overview.csv).

3. Distributional Semantic Analysis

After these preliminary analyses, we turned to the methods from the area of distributional semantics. Distributional semantics draws on the so-called distributional hypothesis, which posits that there is a correlation between distributional similarity and the semantic similarity of words:[24] words appearing in similar language contexts that tend to have similar meanings. It implies that it should be possible to use the former (i.e. data on the distribution of words within a language corpus) to approximate the lat-

20 Keersmaekers (2021), 39–50.

21 Blackwell / Smith (2019), 73–94.

22 Eijk (2008), 385–412; Eijk (2015).

23 These four word roots were chosen because they represent the four most frequently used pain words in classical antiquity. See Cheng (2018), 1–25.

24 Harris (1954), 146–162; Sahlgren (2008), 33–53.

ter. Formally, this approach relies to a substantial extent on the toolbox of linear algebra, representing the distribution of words using multidimensional vectors called word embeddings.

In particular, here we employ the CBOW (continuous bag-of-words) model, a variant of the famous word2vec model.[25] It is a neural network-based approach for obtaining multidimensional vector representations of words. The vectors are generated by training a model aiming to predict a target word from one or more context words. During the training process, the model receives a large number of short sequences of words extracted from a corpus as its inputs. With each sequence, one word is hidden from it (the target word), and the model attempts to predict it using the remaining words (context words). In the model, each word is represented by a set of weights or parameters. These weights are gradually updated as the model learns from its successes and mistakes. After the training ends, these weights are extracted from the model and treated as vectors representing individual words. Thus, in the case of word2vec, the vector representations of words are a byproduct of a model trained to perform a different task: to predict a hidden word. It has been demonstrated that when trained on a sufficient amount of data, neural-based embeddings usually outperform other more straightforward models.[26]

We train the word2vec models by using the CGL dataset described above. During the learning process, we feed the model using a randomized mixture of word sequences of three different forms: The first type of word sequences are the lemmatized and filtered sentences described above; the second type of sequences are continuous word bigrams extracted from the sentences; and, finally, the third sequences are continuous word trigrams also extracted from the sentences. Thus, for instance, the opening sentence of Aristotle's *Nicomachean Ethics* (πᾶσα τέχνη καὶ πᾶσα μέθοδος, ὁμοίως δὲ πρᾶξίς τε καὶ προαίρεσις, ἀγαθοῦ τινὸς ἐφίεσθαι δοκεῖ) is represented by the lemmatized and filtered sentence (πᾶς τέχνη πᾶς μέθοδος πρᾶξις προαίρεσις ἀγαθός τις ἐφίημι δοκέω), overlapping continuous bigrams such as πᾶς τέχνη, τέχνη πᾶς, πᾶς μέθοδος, etc., and overlapping continuous trigrams such as πᾶς τέχνη πᾶς, τέχνη πᾶς μέθοδος, πᾶς μέθοδος πρᾶξις, etc. This implies that one occurrence of a two-word phrase is represented several times within the data feeding the model. It has two valuable properties: first, it makes the input data denser and bigger, which is always valuable when training a machine learning model; second, it is sensitive to the local proximity of words, paying the highest attention to words appearing in the immediate neighborhood.

Within this article, we introduce five models, each of them based on a slightly different subset of texts from the CGL dataset. First, we train a full model based on all 371 texts within the CGL corpus. Furthermore, we generate a model trained on all texts from CGL, except for texts from the *Corpus Hippocraticum*, i.e. based on a subset of 319 works and covering 1,632,716 lemmata. Analogically, we generate a model based on CGL with the exception of the *Corpus Aristotelicum* (336 works, 1,446,704 lemmata) and a model based on CGL with the exception of the *Corpus Platonicum* (334 works, 1,543,372 lemmata). Finally, we generate a model based on CGL with the exception of both the *Corpus Aristotelicum* and the *Corpus Platonicum* (299 works, 1,161,783 lemmata).

These five models allow us to study the impact of the exclusion of an individual subcorpus upon the semantic relationship within the corpus as a whole. When we quantitatively compare the data from these models, we can analyze the extent to which the semantic similarities between any two words in the overall corpus are driven by their usage within an individual subcorpus. Alternatively, we could build independent models for each subcorpus. However, this would imply the necessity to train models on significantly smaller textual data, and we know that the size of the data used for developing the models is crucial for obtaining the appropriate results. On the contrary, with our approach, we see that each of the five models is based on texts covering more than 1 million words.

25 Mikolov et al. (2013).

26 Sahlgren / Lenci (2016).

All five models employ the same vocabulary, which consists of the 3,462 words appearing in the CGL corpus at least 50 times (as above, only lemmatized nouns, verbs, and adjectives are included here). Within each model, there are 3,462 vectors, each one corresponding to one word within the vocabulary. Due to our parametrization of the training algorithm, our vectors have 150 dimensions, a number quite often employed in the literature and sufficient to capture the semantic features we are interested in. The assumption is that vectors occupying similar positions within this 150-dimensional space are semantically related. While our perceptual system is not suitable to capture spatial proximity within a multidimensional space, the mathematical apparatus of linear algebra is free of these limitations. A common algebraic technique to measure the proximity of vectors within a multidimensional space is cosine similarity, defined as the cosine of the angle between two vectors. The resulting score is on a scale between -1 and 1, where 1 means the complete identity of the angle. In other words, cosine similarity allows us to measure the extent of correlation between two rows of numbers such as our word vectors.

In Figure 1, we see a matrix of cosine similarities between the vectors corresponding to the ten pain words in the full model.[27] This figure enables us to see the level of (dis)similarity between particular pain words, which is a valuable first step in the classification of pain in general. Unsurprisingly, pain words derived from the same radix have relatively high scores, on average around 0.5. On the other hand, the fact that some words have higher scores with words from other word families than from their own is surprising. This is most clearly seen in the case of ἄλγημα – ὀδύνη (0.65) and ἄλγημα – ὀδυνάω (0.67). Other words from the ἀλγ* family have a slightly lower but noticeable relationship to the ὀδυν* family. This indicates that these two pain word families are used in similar contexts to denote similar phenomena.

This figure also captures the fact that some pain words have a very low score. This is particularly the case of the λυπ* family: λυπέω – ἄλγημα (-0.02) λυπηρός – ὀδύνη (0.02) λύπη – ὀδυνάω (0.01). The score between λυπ* and πον* families is higher, but it is still clear that the λυπ* family is distant from other pain word families. The link between the πον* family and other families is quite constant between 0.2 to 0.3. This figure thus indicates that while ὀδυν* and ἀλγ* families are very close to each other, πον* has a significantly weaker link to them, whereas λυπ* has almost none. We can thus presuppose that ὀδυν* and ἀλγ* will show close semantic similarities, while λυπ* and πον* will be somewhat distant from them. We elaborate upon this observation in the following paragraphs.

27 We used the four pain word families that are most common in classical Greek texts.

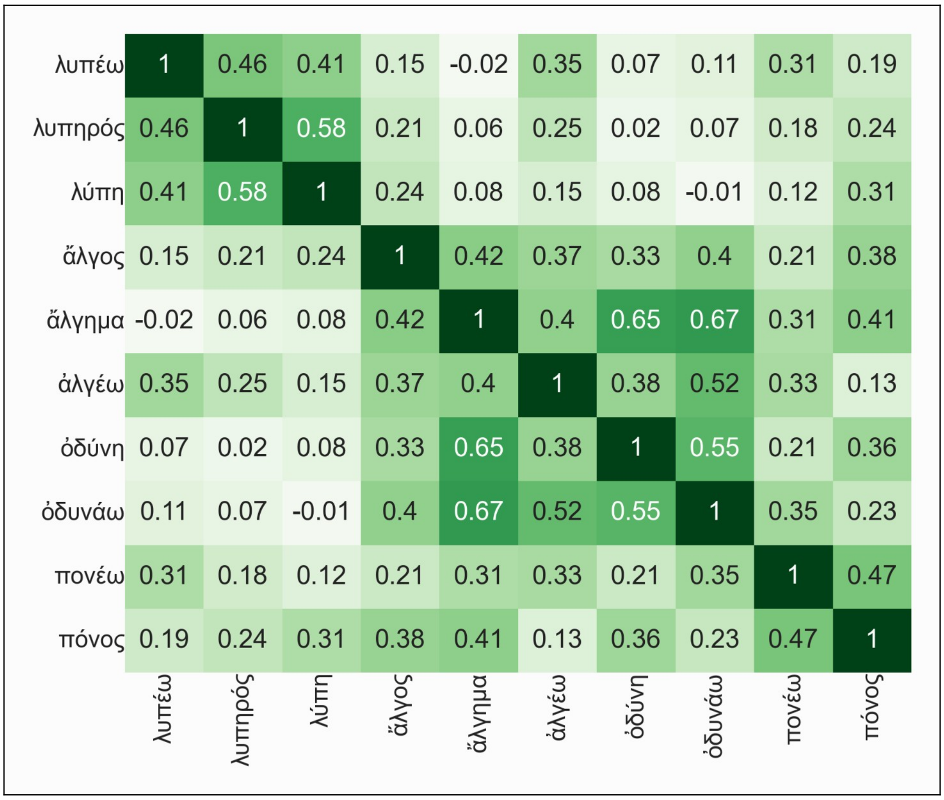

Fig. 1: Cosine similarity matrix of vectors of ten pain words within the full word2vec model.

Furthermore, for each of the ten words, we can inspect which are their nearest neighbors within the vector space. In particular, for each target word (i.e. pain words), we identify a subset of words with vectors having the highest cosine similarity with the vector of the target word. In Table 1, we see the 10 nearest neighbors for each of the pain words. For instance, we observe that among the nearest neighbors of the ἀλγ* words, there are some ὀδυν* words and vice versa. We can also notice that these two groups of words share several neighbors, e.g. ἰξύα, κενεών, νείαιρα or βουβών. Generally speaking, ἀλγ* and ὀδυν* share the same neighbors or share neighbors belonging to the same categories, such as body parts and pathologies. In contrast, λυπ* and πον* words have very specific neighbors absent in ἀλγ* and ὀδυν*. Neighbors of both λυπ* and πον* include only one pain word family, namely ἀλγ. As for other neighbors, λυπ* is close to words denoting pleasure – thus the opposite of pain – such as ἡδύς, ἡδονή, ἥδομαι, or emotions and terms connected with morality. In the case of πον*, we see that the majority of its neighbors are connected to hard work, labour, exercise, etc. Thus, we can see that every word family relates to the closest similar words. Thus, we can see that in every pain word family, there is quite a different tendency of most close words.

λυπέω	λυπηρός	λύπη	ἄλγος	ἄλγημα
χαίρω (0.51)	λύπη (0.58)	λυπηρός (0.58)	τάλας (0.66)	ἰξύα (0.75)
εὐφραίνω (0.5)	ἡδύς (0.58)	ἡδονή (0.57)	μέλεος (0.6)	ὀδυνάω (0.67)
ἀκόλαστος (0.49)	ἀλγεινός (0.52)	ἐπιθυμία (0.57)	δύστηνος (0.6)	ὑποχόνδριος (0.67)
ἥδομαι (0.49)	βλαβερός (0.51)	σωματικός (0.53)	πότμος (0.57)	φρικώδης (0.67)
ἄχθομαι (0.46)	ἀηδής (0.48)	ἀκολασία (0.51)	οἰκτρός (0.56)	θέρμη (0.67)
δυσχερής (0.46)	συζάω (0.47)	ἀληδών (0.48)	πῆμα (0.56)	ὀσφῦς (0.67)
λυπηρός (0.46)	λυπέω (0.46)	θυμός (0.48)	τλήμων (0.54)	τράχηλος (0.66)
ἀπολαύω (0.45)	ἀπολαύω (0.45)	ἐγκράτεια (0.47)	τλάω (0.54)	κενεών (0.65)

κέρδος (0.45)	ἐπιθυμία (0.45)	Ἀφροδίσιος (0.47)	στένω (0.53)	ὀδύνη (0.65)
ἀγανακτέω (0.45)	παρουσία (0.44)	φθόνος (0.47)	πόθος (0.53)	βουβών (0.64)
ἀλγέω	**ὀδύνη**	**ὀδυνάω**	**πονέω**	**πόνος**
ὀδυνάω (0.52)	θέρμη (0.66)	μετάφρενον (0.7)	ταλαιπωρέω (0.6)	πλησμονή (0.54)
ἀλγεινός (0.49)	ἄλγημα (0.65)	ἰξύα (0.69)	πόνος (0.47)	ταλαιπωρία (0.5)
δακρύω (0.48)	δίψα (0.63)	ὑποχόνδριος (0.69)	γυμνάζω (0.44)	πονέω (0.47)
ψαύω (0.46)	βήξ (0.62)	ἄλγημα (0.67)	ἰσχναίνω (0.42)	ἀλγηδών (0.46)
βαρύνω (0.46)	στραγγουρία (0.61)	κενεών (0.65)	ταλαιπωρία (0.42)	εὐεξία (0.46)
νείαιρα (0.44)	νείαιρα (0.61)	ὕφαιμος (0.64)	εὐεξία (0.42)	ἄλγημα (0.41)
παραφρονέω (0.43)	ἰξύα (0.58)	βουβών (0.64)	βαρύνω (0.41)	καῦμα (0.41)
φρίκη (0.43)	κενεών (0.57)	πλευρόν (0.64)	διαπονέω (0.41)	ἐπίπονος (0.4)
ἰξύα (0.42)	σπασμός (0.56)	ὀδυνώδης (0.63)	κουφίζω (0.4)	γυμνάσιον (0.4)
ὑποχόνδριος (0.42)	φλεγμονή (0.56)	νείαιρα (0.63)	παραμένω (0.36)	φῦσα (0.39)

Tab. 1: Nearest neighbors of the three pain word families.

Looking at the data from Table 1, it appears that the nearest neighbors come from several discernible semantic categories such as bodily parts, dietetics, emotions, etc. Drawing on this observation, we subjected the terms within the table of nearest neighbors to manual coding. We repeated this procedure with all five tables, each corresponding to one model. Each term within these tables was assigned to one of the following categories:

- pain (pain words)
- suffering (more general than pain, not necessarily connected to the body)
- pathology (diseases, injuries, lack of health)
- bodily parts
- dietetics (words connected to work, exercise, eating, and drinking)
- emotions
- morality (virtues, vices, character traits)

The resulting list of coded words with categories assigned to them can be found in the file terms_translation_categories.csv available via Github.[28] Obviously, our decisions concerning how to code individual words can easily be problematized. However, despite these possible problems, we believe that our division of words into these categories is heuristically valuable since it covers both the aspects of pain discussed in the medical texts (pathology, body parts, dietetics) and philosophical texts (emotions, morality).

Subplot A in Figure 2 depicts the strength of semantic association between the pain words and the seven semantic categories within the full model. Each category is represented by the ten most frequent terms within the vocabulary used during the generation of the vector model. Strength is measured as average cosine similarity between the pain word on the one hand and the ten words most representat-

28 See https://github.com/kasev/PIPA/blob/main/data/terms_translation_categories.csv (last access 21.02.2023).

ive of the category on the other. The visualization of Figure 2 confirms what we have already mentioned above about the typical neighbors of particular pain words and also provides some new insights. A connection to the category 'pain' is unsurprisingly present in all pain words. In addition to that, however, we can see some relevant differences between the particular pain words. There seems to be a strong association between λυπ* words and 'morality' and 'emotions'. Both these categories are connected only to λυπ* words. At the same time, λυπ* words have a connection with the highest number of categories (five in the case of λυπέω, four in the case of λύπη and λυπηρός). There is no connection to 'bodily parts' or 'pathologies', but a slight connection to dietetics. We could thus tentatively claim that λυπ* designates pain that is more connected to psychic faculties, emotions, and moral aspects of life than other pain words are.[29] In contrast, there is a strong association between 'bodily parts' and ἀλγ* and ὀδυν* words. We have already observed this in the table above. What is more clear now is the fact that there is a semantic difference inside the ἀλγ* word family, since ἄλγημα is strong in its connection to pathology and bodily parts, while ἄλγος and ἄλγεω have no connections to more specific categories, only to pain and suffering. Thus, ἄλγημα is more similar to ὀδύνη and ὀδυνάω, and all three pain words are connected to 'bodily parts' and 'pathologies'. Thus, we could say that they denote bodily pain caused by injuries or diseases. Finally, πον* is specific in its connection to 'dietetics', which is absent in other pain words (except λυπέω). Thus, it seems that this word family is connected to hard work, exercise, etc.[30] We can thus say that there are clear differences between pain word families and, although there are exceptions, there are some general consistent semantic layers for particular pain words or pain word families.

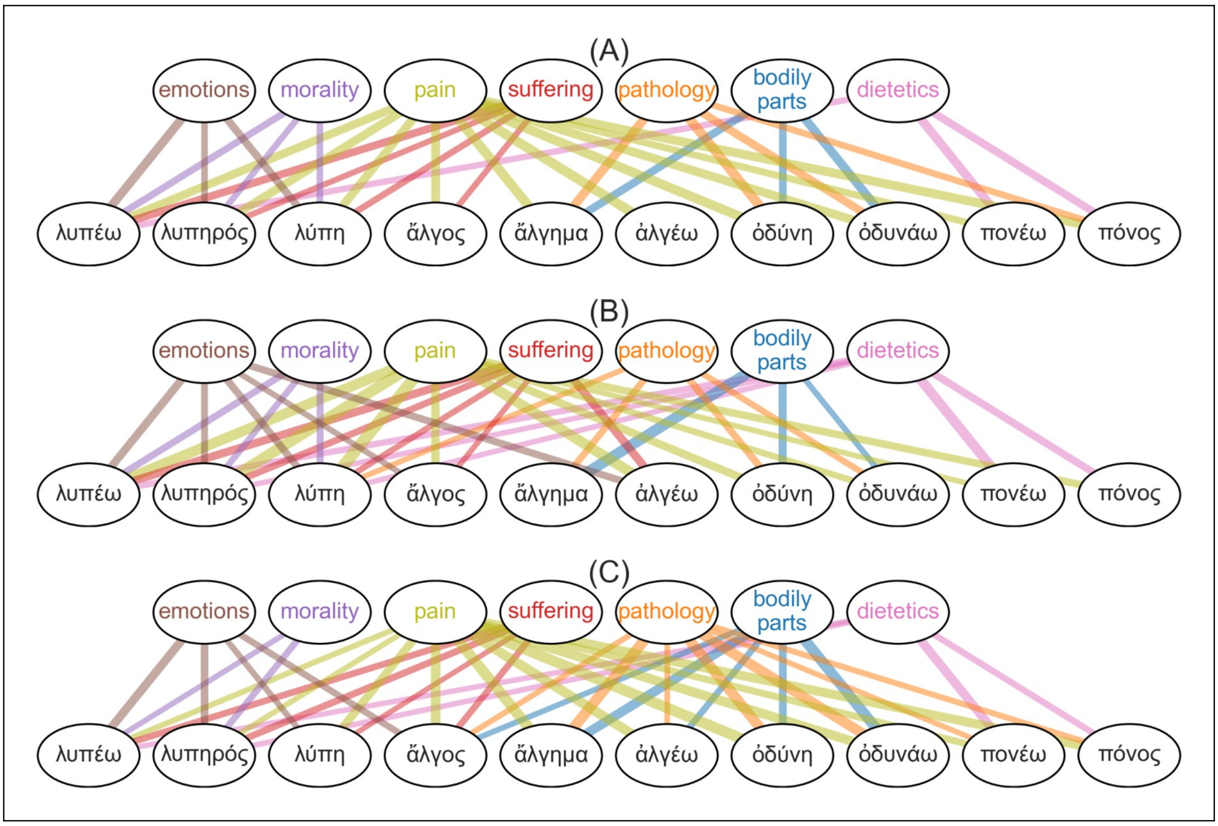

Fig 2: Association between pain words and selected semantic categories across three models: (A) – full model; (B) – model excluding *Corpus Hippocraticum*, (C) – model excluding both *Corpus Platonicum* and *Corpus Aristotelicum*. Line strength expresses the extent of cosine similarity, with a stronger line meaning a higher extent of similarity.

29 This fact is corroborated by Cheng (2018), 6.

30 This is not surprising, as in some important texts, for example the 'Hippocratic' *On Regimen*, πόνος means exercise or activity. See e.g., Vict. I,2 (6,470 L = 124,2–20 Joly-Byl), II,66 (6,588 L = 190,25–7 Joly-Byl), II,66 (6,582 L = 188,12–14 Joly-Byl).

After these explorations of the full model (i.e. a model based on the complete list of works within the CGL corpus), we can move on and focus on models trained on subsets of works excluding the individual subcorpora of our interest. In adopting this approach, we can ask: "To what extent are the associations between the words and semantic domains we have just observed driven by the usage of certain terms within the subcorpora and nowhere else?" Some useful insights can be obtained by looking at Figures 2-B and C.

Figures 2-B and 2-C adopt the same method as Figure 2-A, but are built on other vector models. Figure 2-B is based on the vector model trained on the subset of texts from the CGL corpus, except for the texts from the *Corpus Hippocraticum*. Thus, in Figure 2-B, in cases in which we observe a weaker line between a pain word and a semantic category than in Figure 2-A, it means that the association between the term and the domain was to a substantial extent driven by the *Corpus Hippocraticum* and that, after its exclusion, the association diminishes. Thus, we see that once we exclude the *Corpus Hippocraticum*, the association between ὀδυνάω and 'bodily parts' becomes weaker. In other words, this association is driven by this subcorpus. This is also true in the case of the category 'dietetics', which in the full model was dominantly connected only to πον*; now, λυπ* associates with it as well. The *Corpus Hippocraticum* also influences the association between 'emotions' and pain words: in the full model, it is associated only with λυπ*; now it is associated with ἄλγος and ἀλγέω as well. Another interesting finding is the link between λύπη and 'pathology', which was also absent in the full model.

We can also interpret Figure 2-C in the same vein. Perhaps most interesting here is the drop in association strength between λύπη and 'morality', and also between λύπη and other pain words and 'emotions'. This indicates that these associations are to a substantial extent driven by the two philosophical corpora. The exclusion of the two philosophical corpora has an effect on the category dietetics that is similar to the exclusion of the *Corpus Hippocraticum*: λύπη and λυπήρος associate with it. It is also worth mentioning that ἀλγέω and ἄλγος are now associated with 'pathologies' and 'bodily parts', which indicates that philosophical corpora substantially influence their semantics.

Finally, we can return to our former analysis of the mutual relationships between pain words and in what manner these associations are driven by individual subcorpora. For this purpose, we generated Figure 3, which captures the impact of the exclusion of individual subcorpora upon the strength of the semantic connection between individual pain words. The cell values in these matrices express the difference between values in the matrix of cosine similarities between pain words within the full model (depicted in Figure 1) and the matrix of cosine similarities between the pain words within a model excluding a subcorpus. For instance, in Figure 1, we could see that the cosine similarity between πονέω and ἀλγέω is 0.33. In Figure 3-A, we see the value -0.18. This is because cosine similarity between vectors corresponding to πονέω and ἀλγέω within the model excluding the *Corpus Hippocraticum* is 0.15[31] and 0.15 minus 0.33 is -0.18. That means that in the subset of texts from CGL, excluding the *Corpus Hippocraticum*, the semantic association between πονέω and ἀλγέω is substantially weaker than within the full model including this subcorpus, which implies that the association between these two words is driven by the given subcorpus. In other words, by reading the full CGL corpus regardless of individual authors, genres, or subcorpora, we would easily gain the impression that these two terms are used in a relatively similar manner (0.33). However, upon checking the sources more closely, we would realize that this is actually caused mainly by the usage of these terms within the *Corpus Hippocraticum* and is much weaker elsewhere. On all subplots of Figure 3, such cases are represented by negative values. On the other end of the spectrum, there are cases with positive values. These indicate an opposite trend. For instance, in subplot Figure 3-D, we see how similarities are affected once we exclude the two philosophical subcorpora, namely the *Corpus Platonicum* and the *Corpus Aristotelicum*. In Figure 1, we can see that the association between λύπη and ἄλγος tends to be rather mod-

31 See supplementary Figure sim_matrices_all.png available here: https://github.com/kasev/PIPA/blob/main/figures/sim_matrices_all.png (last access 21.02.2023).

erate, i.e. 0.24. But after we exclude the two philosophical corpora, we see that their connection becomes much stronger: 0.5. In such a case, the difference is a positive number, 0.26. It appears that with the *Corpus Platonicum* and the *Corpus Aristotelicum*, the similarity between these two terms is much weaker, as they are used differently than in the overall corpus, an observation we have already made with respect to the relationships depicted in Figure 2.

We can see that the exclusion of each subcorpus has a different impact on the relations between pain words. While there are no substantive changes when excluding the *Corpus Aristotelicum* (Figure 3-C), in the case of *Corpus Hippocraticum*, these changes are large (Figure 3-A). Of particular interest is the relation between the λυπ* pain word family and ἀλγέω, ὀδύνη, and ὀδυνάω, which increases after excluding the *Corpus Hippocraticum*. The relation between λυπ* and these three pain words is thus much weaker in this subcorpus than in the full corpus. Similarly, the pain word ἄλγημα evinces that it is used in a specific way in this subcorpus: in the *Corpus Hippocraticum*, ἄλγημα is much closer to other pain words than in the rest of the corpus. In general, in the *Corpus Hippocraticum*, associations between pain words are most specific in comparison to the full corpus and the exclusion of other subcorpora. Thus, the influence of the way pain words are used in the *Corpus Hippocraticum* on the whole CGL corpus is most substantial. Figure 3, therefore, corroborates our observation that not only associations between pain words and semantic domains are influenced by the specificity of various subcorpora, but the relations between pain words themselves are as well. While in the case of Figure 2, it seems that the exclusion of each subcorpus has a similarly relevant impact on the semantics of pain words, as is the case in Figure 3, where the influence of the *Corpus Hippocraticum* is the strongest.

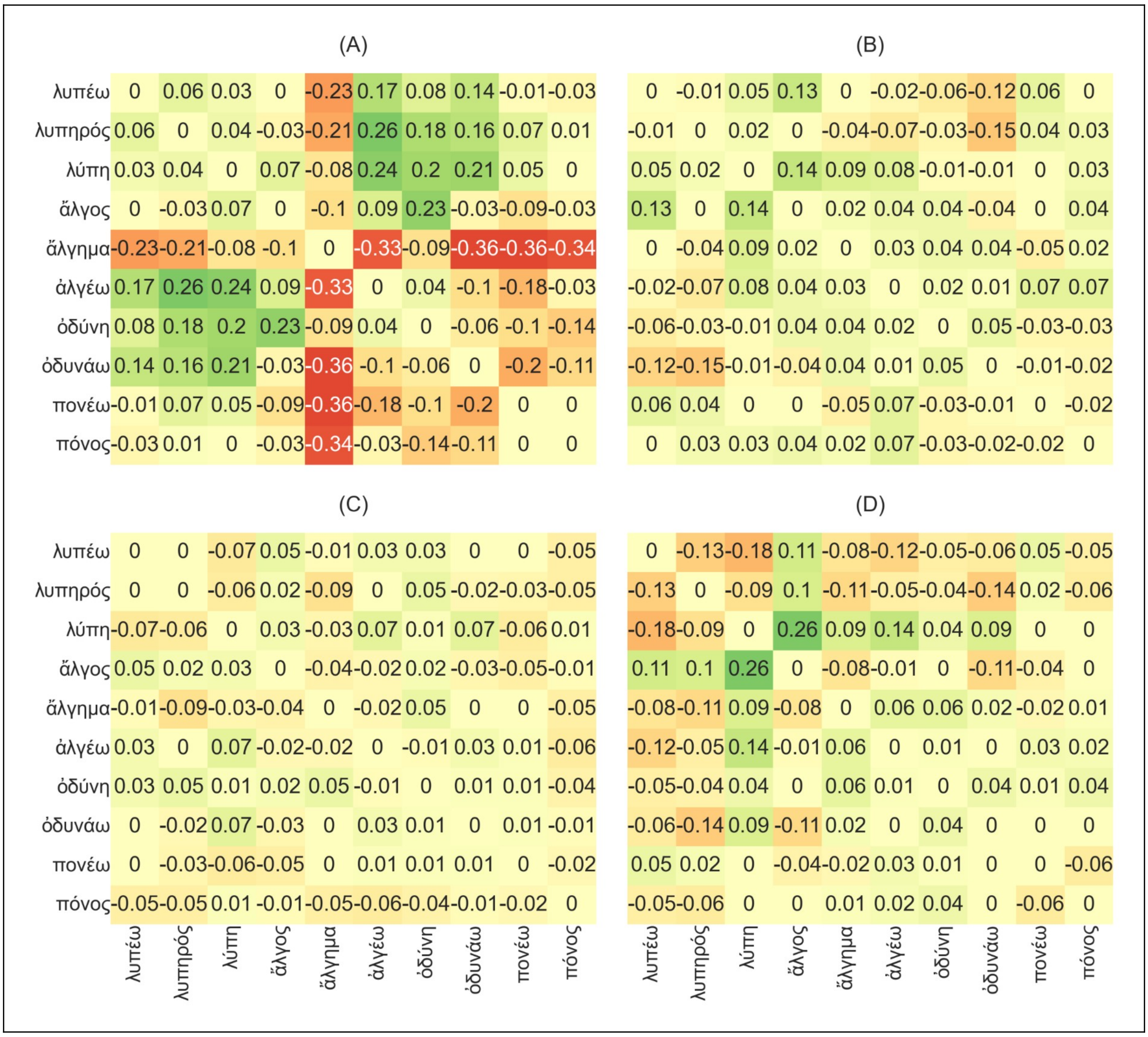

Fig. 3: Matrix of differences in cosine similarity between individual pain words after the exclusion of individual subcorpora: (A) – exclusion of the *Corpus Hippocraticum*; (B) – exclusion of the *Corpus Platonicum*; (C) – exclusion of the *Corpus Aristotelicum*; (D) – exclusion of both the *Corpus Platonicum* and the *Corpus Aristotelicum*.

Conclusions

In this paper, we computationally analyzed four pain word families, their role, meaning, and mutual relationship in Classical Greek texts. Although all four pain word families were used for denoting pain, we have shown that their meanings differ significantly according to the features of pain that they associate with. First, we analyzed relations between particular pain words and emphasized the closeness between ἀλγ* and ὀδυν*. Then we saw how particular pain word families differ in their associations to semantic categories: λυπ* is close to emotions and morality, ἀλγ* and ὀδυν* to bodily parts and pathologies, πον* to dietetics. This was specified by focusing on a particular pain word in pain word families. Nonetheless, the greatest contribution lies in revealing how particular subcorpora influence both the association of pain words with categories and the relations between pain words themselves. Especially in the case of the *Corpus Hippocraticum*, we have seen that its influence on the usage of pain words in the whole corpus was substantive. This study also showed that computational methods – together with close reading expertise – are appropriate tools for approaching traditional philosophical or historical problems, as we can analyze a vast amount of data and capture the meaning of relevant phenomena that are otherwise difficult to grasp by close reading alone.

Sources

Text editions

Littré (1839–1861): É. Littré (ed./transl.), Oeuvres complètes d'Hippocrate, Paris 1839–1861, repr. Amsterdam 1961–1962 & 1973–1991.

References

Budelmann (2007): F. Budelman, The Reception of Sophocles' Representation of Physical Pain, AJPh 128 (2007), 443–467.

Blackwell / Smith (2019): C. Blackwell / N. Smith, The CITE Architecture, A Conceptual and Practical Overview, in: M. Berti (ed.), Digital Classical Philology, Ancient Greek and Latin in the Digital Revolution, Berlin 2019, 73–93.

Bostock (2000): D. Bostock, Aristotle's Ethics, Oxford 2000.

Brodie (1991): S. Brodie, Ethics with Aristotle, New York / Oxford 1991.

Cerrato et al. (2020): L. Cerrato et al., PerseusDL/canonical-greekLit 0.0.2711 (Version 0.0.2711), Zenodo, https://doi.org/10.5281/zenodo.4067170 (Last access 21.03.2023).

Cheng (2015): W. Cheng, Pleasure and Pain in Context: Aristotle's Dialogue with his Predecessors and Contemporaries, PhD diss. Humboldt Universität, Berlin 2015.

Cheng (2019): W. Cheng, Aristotle's Vocabulary of Pain, Philologus 163/1 (2019), 47–71.

Craik (2015): E. M. Craik, The 'Hippocatic' Corpus, London / New York 2015.

Crane et al. (2020): G. R. Crane / L. Mueller / B. Robertson / A. Babeu / L. Cerrato / T. Koentges / R. Lesage / L. Stylianopoulos / J. Tauber, First1kGreek (Version 1.1.5070), Zenodo, https://doi.org/10.5281/zenodo.4091475 (Last access 21.03.2023).

Eijk (2008): P. van der Eijk, Medicine in the Formation of Early Greek Thought, in: P. Curd et al. (eds.), Oxford Handbook of Presocratic Philosophy, Oxford 2008, 385–412.

Eijk (2015): P. van der Eijk, On 'Hippocratic' and Non-'Hippocratic'' Medical Writings, in: L. Dean-Jones et al. (eds.), Ancient Conceptions of the 'Hippocratic', Papers Presented at the XIIIth International Hippocrates Colloquium Austin, Texas, August 2008, Leiden / Boston 2015, 17–47.

Frede (1992): D. Frede, Disintegration and Restoration, Pleasure and Pain in Plato's Philebus, in: R. Kraut (ed.), The Cambridge Companion to Plato, Cambridge 1992, 425–463.

Frede (2016): D. Frede, Pleasure and Pain in Aristotle's Ethics, in: R. Kraut (ed.), The Blackwell Guide to Aristotle's Nicomachean Ethics, Oxford 2016, 255–76.

Gosling / Taylor (1982): J.C.B. Gosling / C.C.W. Taylor, The Greeks on Pleasure, Oxford 1982.

Harris (1954): Z. S. Harris, Distributional Structure, Word & World 10 (1954), 146–62.

Harte (2014): V. Harte, The Nicomachean Ethics on Pleasure, in: R. Polansky (ed.), The Cambridge Companion to Aristotle's Nicomachean Ethics, Cambridge 2014, 288–138.

Horden (1999): P. Horden, Pain in Hippocratic Medinine, in: J. R. Hinnells et al. (eds.), Religion, Health and Suffering, London 1999, 295–315.

Kaše (2021): V. Kaše, LAGT: Lemmatized Ancient Greek Texts (v1.0.0) [Data set], Zenodo, https://doi.org/10.5281/zenodo.4552601 (last access 27.03.2023).

King (1998): H. King, Hippocrates' Woman, Reading the Female Body in Ancient Greece, London 1998.

Keersmaekers (2021): A. Keersmaekers, The GLAUx corpus: Methodological issues in designing a long-term, diverse, multi-layered corpus of Ancient Greek, Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021, 39–50. https://doi.org/10.18653/v1/2021.lchange-1.6 (last access 21.02.2023).

Linka / Kaše (2021): V. Linka / V. Kaše, Pain and Body in Corpus 'Hippocraticum', A Distributional Semantic Analysis, Digital Classics Online 7 (2021), 54–71. https://doi.org/10.11588/dco.2021.7.81212 (last access 24.04.2023).

Mikolov et al. (2013): T. Mikolov / I. Sutskever / K. Chen / G. S. Corrado / J. Dean, Distributed Representation of Words and Phrases and their Compositionality, in: C. J. C. Burges et al. (eds.), Advances in Neural Information Processing Systems 26 (2013), 3111–3119.

Owen, G. E. L. (1977): Aristotelian Pleasures, in: Barnes J. et al. (eds.), Articles on Aristotle, 2. Ethics and Politcs, London 1977, 92–103.

Sahlgren (2008): M. Sahlgren, The Distributional Hypothesis, Rivista di Linguistica, 20/1 (2008), 33–53.

Sahlgren / Lenci (2016): M. Sahlgren / A. Lenci, The Effects of Data Size and Frequency Range on Distributional Semantic Models. https://arxiv.org/abs/1609.08293 (last access 27.03.2023).

Taylor (2008): C. C. W. Taylor, Pleasure, Mind and Soul, Selected Papers in Ancient Philosophy, Oxford 2008.

Villard (2006): L. Villard, Vocabulaire et représentation de la douleur dans la Collection hippocratique, in: F. Proust et. al. (eds.), Penser et représenter le corps dans l'Antiquité, Rennes 2006.

Wolfsdorf (2013): D. Wolfsdorf, Pleasure in ancient Greek philosophy, Cambridge 2013.

Author contact information[32]

Vojtěch Linka
Faculty of Arts
Charles University
nám. Jana Palacha 2
11638, Praha 1
Czech Republic
E-Mail: vojtech.p.linka@gmail.com

Vojtěch Kaše
Faculty of Arts
University of West Bohemia
Sedláčkova 19
30614, Plzeň
Czech Republic
E-Mail: kase@kfi.zcu.cz