Discourse cohesion in Xenophon's On Horsemanship through Sketch Engine

Victoria Beatrix Fendel <victoria_dot_fendel_at_classics_dot_ox_dot_ac_dot_uk>, University of Oxford https://orcid.org/0000-0001-6302-3726

Matthew T.Ireland <mti20_at_cam_dot_ac_dot_uk>, Sidney Sussex College, University of Cambridge https://orcid.org/0000-0003-0336-4419

Abstract

We build a Sketch Engine corpus for Xenophon's classical Greek scientific treatise *On Horsemanship*. Sketch Engine is a web-based corpus-analysis tool that allows the user to inspect the lexical makeup of a text (cf. keyword lists), explore the surroundings of select items (cf. concordances) and identify fixed expressions in a text (cf. n-grams). We make available our corpus-preparation tool and our corpus configuration file for Sketch Engine. We use the Sketch Engine corpus to detect discontinuous verbal multi-word expressions, specifically support-verb constructions (e.g. *to take a decision*). We examine how support-verb constructions – through their structural and lexical properties – aid discourse coherence and cohesion throughout Xenophon's treatise. We furthermore examine how the recurring support-verb constructions in the treatise reflect the scientific register of the text. The article shows how an understudied category of lexico-syntactic device (support-verb constructions) in classical Greek majorly aids discourse cohesion, structurally and contextually speaking. It also shows how an understudied text in the form of a technical treatise (*On Horsemanship*) majorly furthers insight into scientific literacy of the classical period. Finally, by making available our corpus-preparation tool and code, we hope to further collaboration and adaptation and thus improvement of existing tools and counteract the multiplication of tools.

The classical Greek historian Xenophon (5th / 4th c. BC) is best known for his literary works (*Anabasis*, *Hellenica*) describing war-time challenges. However, amongst his minor works is a treatise *On Horsemanship*. Xenophon's hands-on guide to choosing, caring for, and training a horse differs from the descriptions of the equines of Greek literature, e.g. Achilles' Xanthos (in Homer's *Iliad*) and Bellerophon's Pegasus (in Hesiod's *Theogony*). In the epic, Xanthos is immortal (Iliad 16.148ff), weeps after Patroclus' death (Iliad 17.426ff), and prophecies Achilles' destiny (Iliad 19.392ff). Pegasus is immortal, winged, and born out of Medusa's blood (Hesiod, Theogony 280ff). In the literary genres, the focus lies with the horses as protagonists rather than with the real-life challenges of horse keeping. Conversely, Xenophon's *On Horsemanship* is interested in the latter. It reflects attention to detail [Greer 2015] along with literacy in scientific discourse. [1]

The treatise *On Horsemanship* is comparatively short (7,139 words), divided into twelve chapters [Bowersock and Marchant 2014], and written in prose. It covers (i) the conformation and character of the horse so as to be fit for the intended purpose [sections 1–3], (ii) the care for the horse by the groom [sections 4–6], (iii) the ridden education of the horse [sections 7–8], (iv) special cases (e.g. the spirited horse, the warhorse, and the parade horse) [sections 9–11], and (v) the arms for horse and rider [section 12]. Each section finishes with a brief interims summary. [2]

Despite this clear sectioning, Xenophon's text 'flows', that is information is communicated effectively, comprehensibly, and systematically to the reader. Linguistically, we cast this flow into the notions of coherence and cohesion. Coherence refers to how building blocks of a sentence are tied together; cohesion refers to the tying together of clauses and sentences [Webster 2019, 41]. Both are needed for information to be communicated in such a way that the reader can establish links with the preceding discourse and the author can add new pieces of information incrementally. Classical Greek relies on a range of morpho-syntactic, lexical, and pragmatic strategies to achieve coherence and cohesion. [3]

Our aim in this article is threefold. We want to show how to use an application (Sketch Engine) widely used for modern languages for classical Greek and provide the reader with a corpus preparation tool that can be used (and adapted) for their own purposes, thus counteracting the multiplication of digital tools and refocusing on their functionalities. We want to show how an understudied category of lexico-syntactic device (support-verb constructions) in classical Greek majorly aids discourse cohesion, structurally and contextually speaking. We want to show how an understudied text in the form of a technical treatise (*On Horsemanship*) majorly furthers insight into scientific literacy of the classical period. These aims translate into the following three research questions which underpin the following sections: [4]

1. How can we choose and facilitate the application of existing corpus-analysis tools for Ancient Greek in a way that is cohesive, text-agnostic, scalable, flexible and freely-reusable?
2. How do support-verb constructions by means of their structural properties aid discourse cohesion and coherence?
3. How do support-verb constructions by their register-related properties aid discourse cohesion and coherence?

Research question 1 explains our choice of Xenophon's *On Horsemanship*. The treatise is not part of and differs in genre and register from the large literary classical Attic corpus which the Sketch Engine corpus was originally built for.[1] We thus showcase adaptability and extensibility of our tool. [5]

The article falls into five sections. Section 1 introduces the notions of coherence and cohesion and explains our choice of Sketch Engine. Section 2 outlines the process of building the corpus-preparation tool, adapting it for Xenophon's *On Horsemanship*, and implementing the corpus into Sketch Engine. Section 3 uses this Sketch Engine corpus to examine support-verb constructions, a discontinuous lexical device creating both coherence and cohesion. Section 4 moves beyond coherence and cohesion achieved through structural means, assessing the adherence to the scientific register. Section 5 summarises the results and offers conclusions. [6]

1 Introduction

Greek relies on a range of morpho-syntactic, lexical, and pragmatic strategies to achieve coherence and cohesion. Relevant single-word items can relatively easily be extracted from a lemmatised corpus of text. However, the same is not true for multi-word items, especially discontinuous ones (e.g. *to have an idea* in *I had a great idea yesterday*). The below describes means to achieve coherence and cohesion in classical Greek discourse before introducing the reader to the structures of choice and the challenges they pose for automated extraction. [7]

Morpho-syntactically, coherence and cohesion can be achieved by means of structuring particles (e.g. *μέν – δέ* "on the one hand – on the other hand," *οὖν* "thus," *γάρ* "for") [Bonifazi et al. 2016], conjunctions (e.g. *ἀλλά* "but" and *ὅτι* "because"), and pronouns, which must have a clear anaphoric or kataphoric link to the item they stand in for (e.g. *I met Will yesterday. He is very well.* where "he" can only be used once we have established who it refers to) [Halliday 1976] [Luraghi 2003] [Rysová 2017]. Moreover, word-order choices play a role since Greek word order is information-structure driven [Celano 2013a] [Celano 2013b] [Dik 1995] [Mastronarde 2013]. Slots in the word-order frame are not indicative of an argument function (e.g. subject, object), but of an information-structural value (e.g. topic / known information, focus / new information). E.g. in Herodotus, *Histories* 5.3.5 *νόμοισι δὲ οὗτοι παραπλησίοισι πάντες χρέωνται κατὰ πάντα* usages.dat – yet.particle – these.nom – alike.dat – all.nom – they.use – in everything "Yet all of them have similar habits in all things," initial *νόμοισι* is the topic component, an unmarked element follows [Beschi 2018, 181–182]. In Sophocles, *Oedipus Colonus* 1656 *Μόρῳ δ' ὁποίῳ κεῖνος ὤλετ᾽* death.dat – yet.particle – what.dat – that.nom – destroy.aor.pass.3sg "Yet through which type of death did he die?," *μόρῳ* is the topic component, *ὁποίῳ* is a focus component, and an unmarked element follows [Beschi 2018, 179]. Setting, topic, and focus components are optional in the sentence, yet can aid discourse coherence [Beschi 2018, 181]. [8]

Lexically, coherence and cohesion can be achieved by the use of synonyms (e.g. *love* and *adore*) or antonyms (e.g. *love* and *hate*), the repetition of keywords (e.g. *ἱππασία* "horse exercise" in Xenophon's treatise) (cf. [Hutchinson 2017]), and scene-setting adverbial phrases, such as "in this treatise" or "to sum up the previous discussion." [9]

Pragmatically, coherence and cohesion can be achieved by the adherence to a register and genre, any derivation from which would derail the reader. A genre is a culturally determined [10]

norm for a text type [Biber and Conrad 2009, 16], with genre markers appearing at the start and end of the text, e.g. *once upon a time* and *they lived happily ever after* in English fairy tales. A register is a situationally conditioned shape of a text [Biber and Conrad 2009, 6], with register markers appearing throughout a text, e.g. *tu* as opposed to *vous* in French informal vs formal discourses (see also [Adams 2013, 107–110] [Bentein 2013] [Bentein 2016] [Willi 2003] [Willi 2010]). The choice of and adherence to a genre and register determines the amount of drawing on shared / background knowledge (cf. inference), the amount of repetition needed, and the way of presenting several events (e.g. impressionistic vs sequential).

The present article focuses on a type of lexico-syntactic device which contributes to discourse cohesion at the lexical, syntactic, and pragmatic levels, that is support-verb constructions, such as *ἱππασίαν X ποιέομαι* to "do X horse-exercise." Support-verb constructions are lexical devices in that their two elements (the verb and the noun) form a unit as regards meaning [Lipka 1992]. However, given their frequent discontinuity, the two elements must be linked together by means of syntactic dependency, with the noun filling the object slot of the verb. Support-verb constructions contrast with simplex verbs of similar meaning, e.g. for *ἱππασίαν ποιέομαι*, there would be *ἱππεύω* "to ride" and *ἱππάζομαι* "to drive / to ride" (Xenophon, *On Horsemanship* 10.13), which are regular denominal formations (in *-ευω* and *-άζω/άζομαι*) [van Emde Boas 2019, 274–275].

Support-verb constructions are characterised by their variability (e.g. *he breaks/broke her/the heart*), ambiguity (e.g. *he broke her chocolate heart* which has a literal reading), and discontinuity (e.g. *he broke her young heart*) [Pasquer et al. 2018, 2585]. All three characteristics are problematic for automated extraction, since "if MWEs [sc. multi-word expressions] are treated by general, compositional methods of linguistic analysis, there is … an overgeneration problem" [Sag et al. 2002, 3]. Overgeneration means that false positives are included in the output and impact on the F1 score, which indicates the accuracy of the result:

$$F_1 = \frac{2}{recall^{-1} + precision^{-1}} = 2\frac{precision \times recall}{precision + recall} = \frac{2tp}{2tp + fp + fn}$$

$$precision = \frac{number\ of\ true\ positive\ results}{total\ number\ of\ positive\ results}$$

$$recall = \frac{number\ of\ true\ positive\ results}{total\ number\ of\ samples\ that\ should\ have\ been\ identified\ as\ positive}$$

tp = a "true positive" (a support-verb construction the algorithm returns that is one)

fp = a "false positive" (a support-verb construction the algorithm thinks is one but is not)

fn = a "false negative" (a support-verb construction the algorithm does not think is one but is)

Example 1.

The F1 score oscillates between ideal 1, i.e. maximum accuracy, and 0.

Furthermore, automated tools rely on training data, such as the annotated PARSEME shared task (https://typo.uni-konstanz.de/parseme/index.php/results/shared-task) for 18 modern languages, and very large and expandable corpora (e.g. [Scheible et al. 2013]).[2] Conversely, the classical Greek corpus of data is comparatively small and cannot be expanded easily, such that training data is difficult to obtain (see also [Sheinfux et al. 2019]). Furthermore, when drawing on previous work on languages other than classical Greek[3], it must be borne in mind that support-verb constructions are language-specific, such that cross-linguistic studies can only point the direction (e.g. *eine Reise machen* vs *to take a trip*). Finally, contiguity / adjacency is not compulsory in support-verb constructions [Pasquer et al. 2018] [Sheinfux et al. 2019]. In Greek, no word-order constraint akin to the English subject-verb-object exists (e.g. Lysias, Speech 3.22 *συνθήκας πρὸς αὐτὸν ποιησάμενος* agreements.acc – with him – making "making agreements with him").[4]

Fully automated approaches rely on algorithms that classify the input data into output categories (here support-verb construction vs not support-verb construction).[5] These are trained on a suitably prepared training corpus. Techniques of analysis combatting the above issues have been suggested in the literature, such as "listing words with spaces, hierarchically organized lexicons, restricted combinatoric rules, lexical selection, idiomatic constructions, and simple statistical affinity" [Doucet and Ahonen-Myka 2004] [Sag et al. 2002, 15]. Furthermore, marking up support verbs in the training data improved support-verb-construction identification [Cap et al. 2015]. [6] [Cordeiro and Candito 2019, 103] find a noticeable difference in performance between seen structures (max. F1 = 0.83) and unseen structures (max. F1 = 0.31), across classifiers evaluated.

We are not aware of existing work that algorithmically extracts support-verb constructions from classical Greek texts, nor have we opted for a fully automated approach. We build a corpus for Sketch Engine [Kilgarriff et al. 2014]. Sketch Engine is a web-based corpus-analysis tool that allows the user to inspect the lexical makeup of a text (cf. keyword lists), explore the surroundings of select items (cf. concordances), and identify fixed expressions in a text (cf. n-grams) (see also [Maiko 2020] [McGillivray and Kilgarriff 2013]).[7] Sketch Engine combines the following functionalities: (i) operation on lemmata rather than word forms, (ii) definition of any corpus, (iii) concordancing, and (iv) creation of n-grams.

Tools that operate on the word form as attested in the text are less informative for classical Greek than for e.g. English as classical Greek is a morphologically rich language, such that e.g. verbal lemmata change stems through the tenses (e.g. "to see" – *ὁράω* present, *ὄψομαι / ὀφθήσομαι* future, *εἶδον / ὤφθην* aorist, *ἑώρακα / ἑώραμαι / ὄπωπα / ὦμμαι* perfect). Tools that predefine or/and allow for limited modification of the corpus of analysis only necessitate extensive manual correction when a different corpus is selected due to research objectives.

Concordances and n-grams make possible the extraction of support-verb constructions when one component is known (either the verb or the noun).[8] Our approach does not take the human researcher out of the equation but facilitates their analysis by providing ranked and easily interpretable collocation metrics. For example, Sketch Engine provides the feature of analysed concordance tables, which can be sorted by various lexical affinity measures. We opt below for the *logDice*. The *logDice* is a measure of lexical affinity between two items with a maximum value of 14 [Rychlý 2008]:

$$logDice = 14 + log_2\frac{2f_{xy}}{f_x + f_y}$$

f_x = number of occurrences of word X

f_y = number of occurrences of word Y

f_{xy} = number of occurrences of words X and Y

Example 2.

Co-occurrence needs to be defined either with regard to distance (i.e. number of intervening items) or with regard to structure (e.g. syntactic relationship between X and Y). Co-occurrence is here defined as appearing within 5 items of each other and in the same syntactic projection. Sketch Engine allows for individual definitions of co-occurrence to be applied.

2 The Fabric of the Text: Building a Sketch Engine Corpus for Classical Attic

We built a corpus for Sketch Engine based on a large sample of literary classical Attic historiography, oratory and prose[9] and adapted the code for *On Horsemanship*.

Approach

It is necessary to tag each word in the source text (the input) and prepare the tagged corpus for import to Sketch Engine (the output). We rely on existing tools for tagging and analysis as far as possible and present a Python program that applies an existing tagging tool (Perseus) to each word in the original text and prepares the corpus for import to Sketch Engine.

Choice of External Tools

For analysing words, correctness and completeness are critical. We chose the Perseus Digital Library Project as an established, widely reviewed, and freely-available source of analyses for individual words. We chose the commercial Sketch Engine tool due to its large feature set and ease-of-use. We are confident in the correctness of the tool due to widespread commercial use in other languages. Commercial applications enable development of a large feature set and easy-to-use user interface that could not be built within the scope of a research project.

Overview

A Python program was written that converts TEI[10]-conformant Greek texts into a tagged corpus for the Sketch Engine tool. The program is extensible and could easily be used with other TEI-conformant texts. To enable anyone to run our open-source program and convert their own texts to a format that can be used by Sketch Engine, we make our program accessible in two formats:

1. as a web-based tool (powered by Google Colab), so that anyone can use our tool on their own texts with the click of a single button in their web browser, without having to download or install software;
2. for advanced users, a Python script in a git repository, so that our program can be used within other scripts, and any changes that users wish to make can be easily shared with us.

We hope that releasing our code as a web-based tool (that does not require any software to be downloaded or installed, or the user's computer to be configured in any way) will enable the community to reproduce our results at the click of a single button. We also hope that by making all of our code open-source and easily accessible, the community may be able to apply Sketch Engine in their own work.

Accessing and Running the Program

Our system is easy to use and we facilitate access through two platforms. The first option is suitable for annotating small texts; the second option is more useful for large corpora.

Option 1: Run the Tool in Your Web Browser

Open the Google Colab notebook: https://colab.research.google.com/drive/1JEuEWVe1t0AyBROb3NwnVMfufPOtzX_8?usp=sharing

Click "Runtime -> Run All." When prompted, choose an input file to upload. When the tagging is finished, the output vertical file will be downloaded automatically, as shown in Figure 1.

Figure 1. Web-based conversion tool, running in Google Colab

Option 2: (Advanced) Clone the Git Repository

The repository can be found here: https://github.com/MatthewIreland/xml_lemmatiser_tagger.

The Python Program

In the main loop, an analysis is looked up in Perseus and then appended to the output file for Sketch Engine (and also recorded in the Analysis Cache in case the word form appears again later in the text). Care is taken to make sure that metadata corresponding to markers in the source text (such as section numbers, chapter numbers, or line numbers) is also recorded in the correct place in the output file so it can be used in Sketch Engine. If an error occurs (e.g. Perseus cannot analyse a word), we record this. Some words that cause errors can be manually annotated (the analysis cache can be initialised with these words on startup). The flowchart in Figure 2 provides an overview of the process:
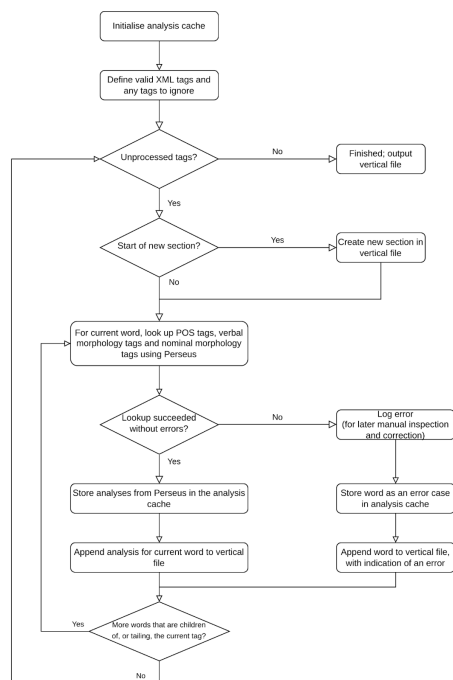
Figure 2. Flowchart of Python program execution

If the same word appears multiple times in the source text, it is only looked up once in Perseus and the same results are printed to the vertical file wherever the word appears. After the word has been looked up using Perseus for the first time, it is added to the Analysis Cache. If the word appears again, the analysis will be retrieved from the Analysis Cache rather than from Perseus. This reduces the overall execution time as fewer network requests are required.

30

Input and Output Formats

The input files are obtained from the Perseus Digital Library. They are in a TEI-compliant format. The output format is a "vertical file" that can be imported into Sketch Engine. For each word in the corpus, the vertical file includes a lemma, POS tag, and nominal/verbal morphology tags.[11]

31

Errors

We focus on the reliability of our results. All errors (i.e. where Perseus fails to lemmatise or tag a word) are logged and manually inspected. Where errors can be corrected (e.g. through manual lemmatising or tagging), the manual annotations are added to the analysis cache on startup so that Perseus is not required for lemmatising/tagging and the words will be correctly lemmatised/tagged on a second pass through the data.[12]

32

Design Choices

Many have previously built parsers for TEI-compliant texts. Why build another one? We needed a fine level of control over the parser in order to manage errors and ensure reliability of results. Hence, we used Python's XML parsing libraries rather than an existing parser specifically for TEI texts.

33

Another tool considered was Morpheus. However, we found that the results from Morpheus were very sensitive to the changes in the build configuration. Results may change when Morpheus has been built on different computers, such that it is difficult to trust the results. Hence, we opted to use the online version of Perseus, which seems to generate consistent and more reliable analyses.

34

The Perseus API (Application Programming Interface) permits the lemmata and tags to be returned in a structured format. However, a bug means that API parameters are not correctly parsed when data is returned in XML format. This bug was fixed by the Perseus authors when data is returned in HTML format, but not when data is returned in XML format. To see the bug (source version 20110527), look at the Perseus source code and compare *a* and *b* in Example 3:

35

a) ./sgml/reading/src/perseus/controllers/document/MorphController.java

b) ./sgml/reading/src/perseus/controllers/document/XmlMorphController.java

Example 3.

Note in MorphController.java the presence of the following code to decode a UTF-8 input parameter:

36

```
word = URLDecoder.decode(word, "utf-8");
word = new String(word.getBytes("8859_1"),"UTF-8");
if (language.equals(Language.GREEK)) {
    word = GreekEncodingAnalyzer.transcode(word, "PerseusBetaCode");
}
```

This is not present in the XmlMorphController but would be required in order to correctly parse inputs. Note in MorphController, the presence of the comment "//don't need this because user is only entering BetaCode," which is not correct because forward and back slashes may be used in BetaCode to represent acute and grave accents respectively.[13] Since they will be encoded as URL parameters, they need decoding correctly in the Java source.

37

Modifications for *On Horsemanship*

A small modification was required for Xenophon's *Minor Works*, which *On Horsemanship* is one of, as it included a "group" XML tag that had not been seen in any previous text. By

38

default, the program throws an error if it sees an unknown tag, so that the user can decide how this tag should be processed (e.g. should metadata from this tag be copied to the output). However, as there is only one "group" tag in the text, the software did not need to take any further action upon seeing this tag. Therefore, the only modification required was to add "group" to the list of known XML tags so that this tag did not cause an error. The Perseus analysis succeeded for every word in the *On Horsemanship* text and there were no errors. The processing of this text executed in 11 minutes and 4 seconds.

Implementation in Sketch Engine

In order for our .vert file to be correctly read by Sketch Engine, especially with regard to (i) the column names (lemma, POS tag, verbal morphology, nominal morphology), (ii) the section and sentence markers and (iii) the types of metadata we include, we produced a modified corpus configuration file: https://github.com/MatthewIreland/linguini/blob/main/files/configFile.txt. [14]

`39`

Accuracy Tests

The errors.txt file is created by the Python program while looking up each word form of a text on Perseus. The word forms for which the lookup fails are appended to errors.txt. Word forms can cause errors related to the POS tag ("cannot find POS tag"), the morphological tag ("cannot find verbal / nominal morphology"), and the lemma ("cannot find lemma"). The error.txt file is then processed to find unique errors, i.e. the minimum set of words that would have to be manually annotated and used for initialising the analysis cache to reduce the total error count to zero. Table 1 summarises the error rates for the classical literary corpus[15], a manually annotated sub-sample of the literary classical Attic corpus[16], and *On Horsemanship*.

`40`

Corpus	Number of Words	Number of Unique Errors in error.txt	Percentage of Error
Corpus of text	492,620	451	0.09%
Test sample	117,783	56	0.05%
On Horsemanship	7,144	0	0.00%

Table 1. error.txt File

The second measure is manual assessment of the Test sample using the *Thesaurus Linguae Gracae* (TLG henceforth), in Text search – proximity – lemma. The results for select items are presented in Table 2.

`41`

Lemma	*Thesaurus Linguae Graecae*	Sketch Engine Test Sample	LogDice				
	Total	Co-occurrence with Predicative Noun	*logDice*	Total	Co-occurrence with Predicative Noun	*logDice*	error vis-à-vis TLG data
δίκη (Predicative Noun)	114	n/a		113	n/a		
λαμβάνω	257	28	11.27	253	26	11.18	0%
δίδωμι	171	53	12.57	171	47	12.40	1.4%
ἀπολείπω	15	ø	n/a	15[17]	ø	n/a	
φεύγω	92	3	8.898	102	3	8.84	0%
συμμαχία (Predicative Noun)	55	n/a		55	n/a		
ποιέω	608	17	9.71	622	15	9.50	2.2%
ἀνίημι	8	2	10.02	80[18]	ø	n/a	
ἀφίημι	34	2	9.52	62[19]	ø	n/a	
ὅπλον (Predicative Noun)	75	n/a		75	n/a		
ἔχω	760	11	8.75	756	11	8.76	0%
παραδίδωμι	58	12	11.53	58	10	11.27	2.3%

Table 2. *Thesaurus Linguae Graecae* vs Sketch Engine

As both the number of attestations and the number of co-occurrences factor into the *logDice*, we calculate the percentage of error for the *logDice* as an overall measure of performance. Sketch Engine performs at the 2.5% threshold, i.e. the percentage of error for all the measurements taken falls below this threshold.

`42`

Items that are likely to produce errors when using our code on a new corpus include (i) *hapax legomena*, (ii) personal names (that are not catalogued in standard dictionaries), (iii) words that are misspelled, and (iv) any not yet considered punctuation marks or textcritical marks that are fused with word forms and thus prevent the lookup from being successful. Moreover, e.g. epic, dialectal, or post-classical word forms (and associated lemmata) cause errors due to the limitations of the standard dictionaries which Perseus accesses.

`43`

3 Cohesion / Coherence through Discontinuity: Support-Verb Constructions

Support-verb constructions are combinations of a verb and a noun that fill the predicate slot of a sentence (e.g. *I took the decision to do it*) [Butt 1995]. The noun is the semantic head of the construction [Nagy et al. 2013, 329] and since filling the predicate slot needs to be predicative or reconceptualised as predicative (e.g. *I took a picture of him* [Radimský 2011]).[20] The agent of the event referred to by the noun must be co-referential with the grammatical subject of the verb, at least in non-causative and non-passive support-verb constructions (e.g. *I paid attention to the talk* vs *I caught his attention* vs *I directed their attention to him*).[21] The noun and the verb contribute to the argument structure (e.g. *I gave him the impression that I wanted to leave*, where the verb is the head of the indirect object (sc. *him*) but the noun is the head of the direct object (sc. *that I wanted to leave*)), unlike in auxiliary-verb constructions (e.g. *I have read him the book*) [Bowern 2008] [Butt 2010] [Butt et al. 2013] [Butt and Lahiri 2013]. The noun and the verb contribute to the semantic structure (e.g. *I gave him the impression that* … vs *I got the impression that* …), such that removal of either component would break up the support-verb construction.[22] Support-verb constructions in classical Greek have not received much scholarly attention, with the exception of the support verb *ποιέομαι* "to do" [Baños 2015] [Cock 1981] [Fendel 2021] [Jiménez López 2011] [Jiménez López 2012] [Jiménez López 2016] [Jiménez López 2021] [Marini 2010] [Zilliacus 1956] [Zilliacus 1967].

`44`

For their functioning as devices to further discourse cohesion, three aspects of support-verb constructions are of interest (see also [Storrer 2009]):

`45`

1. They are multi-morphemic, thus (often) allowing modification of either component (e.g. *He confidently gave a long speech*, where the verb is modified by the adverb of manner *confidently*, whereas the noun is modified by the adjective of degree *long*).[23] This allows for the fine-tuning of the predicate expression [Didakowski and Radtke 2020]. Moreover, their being multi-morphemic and in many cases internally analytic allows for the condensation of several support-verb constructions by deletion of a recurring support verb (e.g. *He made a suggestion and an assumption at the same time* [Gross 1998]); it also allows for the expansion of a support-verb construction across a stretch of discourse, e.g. by means of relativisation (e.g. *The idea which I had yesterday was really useful*) and pronominalisation (e.g. *I had a great idea*

yesterday, I suddenly had it on the train to London).[24] Finally, it allows for the noun to be used recurringly in the discourse without being part of the same support-verb construction at all times (see *ἱππασία* "horsemanship / horse exercise" in *On Horsemanship*) [Jackson 2016, 16–21].

2. They are discontinuous, thus (often) allowing for items to intervene between the noun and the verb, while the support-verb construction is held together by the syntactic dependency relation between the verb and the noun (e.g. *I had* a great *idea*).[25] This allows for the bracketing of pieces of information, thus assigning them unequivocally to the support-verb construction, e.g. Lysias, Speech 3.22 συνθήκας πρὸς αὐτὸν ποιησάμενος agreements.acc – with him – making "making agreements with him." More generally, this ties in with Lakoff and Johnson's observation that a semantic link is reflected in the formal expression (cf. principle of iconicity) [Lakoff and Johnson 1980, 130]. They conclude that *I taught Greek to Harry* and *I taught Harry Greek* differ in that only the latter refers to the acquisition of Greek, which is reflected in the formal expression by the positioning of *Harry* (see also [Frenda 2017]).

3. They are semi-compositional, thus (often) developing a meaning different from and more specific than the related simplex verb [Sanroman Vilas 2009] [Storrer 2009] *unterrichten* "to teach" vs *Unterricht erteilen* "to give a lesson"). Support-verb constructions reflect a range of degrees of compositionality. For example, in *to have an idea*, abstracting from concrete *to possess / to have* to *to belong to / to have* explains the meaning of the support-verb construction [Hermann 2020, 58–61]; in *to take a picture*, we need to reconceptualise the meaning of the noun to refer to the process resulting in the object rather than the concrete object [Radimský 2011]; in *to take heart*, we need to metaphorically extend the meaning of the noun to refer to feelings / emotions and specifically courage [Nunberg et al. 1994] [Sheinfux et al. 2019]. They are also semi-productive in the lexicon, such that they cannot be generated at random or according to a fixed set of rules [Kamber 2008, 143], e.g. *to make a trip* is unnatural in English.[26] Rather, lexical affinity between the verb and the noun governs the creation of support-verb constructions.

In Xenophon's short treatise, three support-verb constructions recur, *ἱππασίαν ποιέομαι* "to do horse exercise" including the topical keyword *ἱππασία*, along with *τεκμήρια παρέχομαι* "to give an indication" and *πεῖραν λαμβάνω* "to put to the test." The latter two refer to the scientific method of finding evidence and drawing conclusions based on it. In the below we focus on those support-verb constructions. A full list of support-verb constructions in *On Horsemanship* can be found in the appendix.

46

ἱππασία is a keyword in the treatise. It appears 17 times in differing syntactic environments, as shown in Figure 3.
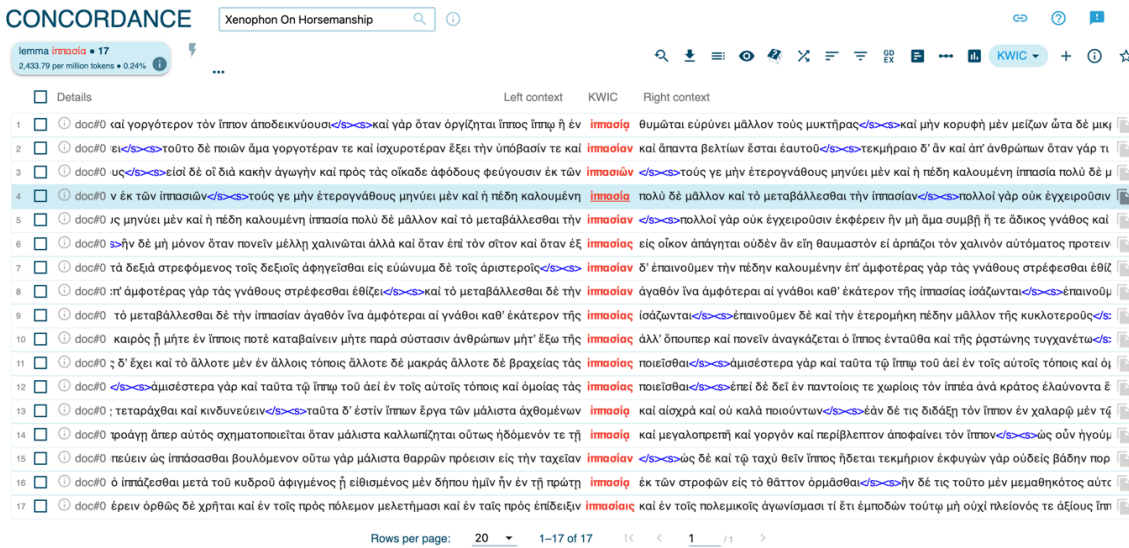
47

Figure 3. Sketch Engine Concordance of *ἱππασία* in *On Horsemanship*

In 8 of 17 instances, *ἱππασία* appears in an adverbial phrase (with a preposition)(rows 1, 3, 6, 9, 10, 15, 16, 17). In 2 of 17 instances, *ἱππασία* takes the form of an adverbial dative with a verb of emotion (rows 13 and 14). In 4 of 17 instances, *ἱππασία* is the subject (row 4) or object (rows 5, 7, 8) of the sentence. The remaining three passages are instances of support-verb constructions.

48

In Example 1, *ἱππασία* is combined with the support verb *ποιέομαι* to mean "to do horse exercise / to exercise the horse." The qualifying adjectives (*μακράς* "long," *βραχείας* "short," and *ὁμοίας* "similar") are spaced out across the sentence.

49

> Ὀρθῶς δ' ἔχει καὶ τὸ ἄλλοτε μὲν ἐν ἄλλοις τόποις, ἄλλοτε δὲ μακράς, ἄλλοτε δὲ βραχείας τὰς ἱππασίας ποιεῖσθαι. ἀμίσεστερα γὰρ καὶ ταῦτα τῷ ἵππῳ τοῦ ἀεὶ ἐν τοῖς αὐτοῖς τόποις καὶ ὁμοίας τὰς ἱππασίας ποιεῖσθαι.
>
> It is right to do exercise sometimes in different places, sometimes for a long time, sometimes for a short time. For these things (i.e. exercising in different places and with diversity of exercises) are less troublesome to the horse than to do exercise in the same places and the same exercise all the time.

Example 1. Xenophon, *On Horsemanship* 8.9

This allows for a contrast to be established between lengths of exercise (*μακράς, βραχείας*) and for these lengths of exercise to be linked to different spatial settings without repetition of the predicate but while maintaining that this refers to one and the same event ("to exercise the horse"). The same support-verb construction appears in the second sentence in Example 1, yet this time the adjective accompanying the noun is an adjective of manner forming the final part of a climax(*ἀεί* "always" – *ἐν τοῖς αὐτοῖς τόποις* "in the same places" – *ὁμοίας τὰς ἱππασίας* "the same exercises") highlighting what is to be avoided when training a horse.[27]

50

The use of a simplex verb would change the type of parallelism underlining the contrast in length in the first sentence and the type of climax underlining the degree of preference/dispreference in the second sentence. A simplex verb could only be modified externally, by means of an adverb, but not internally, by means of an adjective [Didakowski and Radtke 2020]. E.g. *to give a good speech* evaluates the content of the speech vs *to give a speech well* comments on the presentation of the speech. Equally, *to do an extended / short exercise* evaluates the exercise itself, whereas *to do exercise for a long / short time* comments on the execution of the exercise.[28] In the former, the focus is on the exercise itself and its type; in the latter, the focus is on the circumstances of doing the exercise. Xenophon's passage evaluates the type of exercise rather than the circumstances of the exercise.

51

In Example 2, *ἱππασία* is combined with the support verb *ἔχω* to mean "to have ridden work / to behave under saddle." Unlike in Example 1, the noun is not only qualified by adjectives of manner (*γοργοτέραν* "fiercer" and *ἰσχυροτέραν* "more powerful") but is also conjoined with a second noun *ὑπόβασιν* "movement."

52

τοῦτο δὲ ποιῶν ἅμα γοργοτέραν τε καὶ ἰσχυροτέραν ἕξει τὴν ὑπόβασίν τε καὶ ἱππασίαν καὶ ἅπαντα βελτίων ἔσται ἑαυτοῦ.

Doing this, (the horse) will have a fiercer and more powerful movement / gait and ridden work and he will be better than his former self in every way.

Example 2. Xenophon, *On Horsemanship* 1.14

While Example 1 illustrated how a support-verb construction can tie together many pieces of information about the event in question seamlessly, Example 2 shows how a support-verb construction can be used to condense information. Since the two co-ordinated nouns in Example 2 can both function as nouns in a support-verb construction and both accept the support verb *ἔχω*, one instance of the verb is deleted [Gross 1998].

[53]

Similar condensation appears with the one-off structures *ἀσχολίαν / ἀθυμίαν παρέχω* "to provide a lack of rest / a lack of confidence" in section 3.12. Their combination underlines the unpleasantness of the situation for the rider, accentuated by the parallel alpha privative at the start of both nouns. The focus on contrasting items by means of a support-verb construction may underlie *κόσμον παρέχω* "to adorn" in section 12.2, there describing the rider's headwear. The verb *παρέχω* "to give" is contrasted in the parallel structure with *δέχομαι* "to receive" and the noun *κόσμον* "ornament" with *τὸ πρόσωπον* "the face." While the parallel structure does not qualify as a support-verb construction, the parallelism underlines the interplay between the armour and its wearer.

[54]

The other two recurring support-verb constructions concern not the keyword but the key aim of Xenophon's treatise, that is a scientific approach to horse care and training. In Example 3, the support-verb construction *πεῖραν λαμβάνω* recurs three times in quick succession.

[55]

ἐπεὶ δὲ πολεμιστήριον ἵππον ὑπεθέμεθα ὠνεῖσθαι, ληπτέον πεῖραν ἁπάντων ὅσωνπερ καὶ ὁ πόλεμος πεῖραν λαμβάνει. ἔστι δὲ ταῦτα, τάφρους διαπηδᾶν, τειχία ὑπερβαίνειν, ἐπ' ὄχθους ἀνορούειν, ἀπ' ὄχθων καθάλλεσθαι· καὶ πρὸς ἄναντες δὲ καὶ κατὰ πρανοῦς καὶ πλάγια ἐλαύνοντα πεῖραν λαμβάνειν· πάντα γὰρ ταῦτα καὶ τὴν ψυχὴν εἰ καρτερὰ καὶ τὸ σῶμα εἰ ὑγιὲς δοκιμάζει.

Since we hypothesized that the horse fit for war is to be bought, it must be put to the test in everything which war usually puts it to the test in too. These aspects are (i) leaping across ditches, (ii) overcoming walls, (iii) jumping up a bank, (iv) leaping down from banks. In addition, (it is necessary that) by riding uphill and downhill and sideways he put (sc. the horse) to the test. For all these tests indicate whether (the horse) is strong in spirit and healthy in the body.

Example 3. Xenophon, *On Horsemanship* 3.7

Unlike *ἱππασίαν ποιέομαι*, *πεῖραν λαμβάνω* seems to assume a highly specific meaning, possibly acquired through lexicalisation.[29] *πεῖραν λαμβάνω* seems to mean specifically "to put to the test." The subject slot seems to be filled by who or what is doing the testing (*ὁ πόλεμος* "the war" and *ἐλαύνοντα* referring to the implied subject "the rider"). The entity to be tested (here the horse) is contextually salient but is not explicitly mentioned. This is not surprising given the general observation that support-verb constructions help to de-transitivise, that is to decrease the number of event participants (e.g. to suggest (something) vs to make a suggestion) [Marini 2010]. The focus is thus shifted to the event of testing and away from the patient of the event [Foley 2007]. The area of testing is indicated by means of an objective genitive (see *ἁπάντων* "everything"), thus it seems to be specifically the patient that the structure is intended to eliminate. Xenophon's choice of the support-verb construction ("to put to the test") over the simplex verb (*πειράω, πειράζω* "to examine / to test") focuses the reader's attention on the event of testing in more fact-based a way than the integration of a second participant (the horse) could.[30]

[56]

Similarly, *τεκμήρια παρέχομαι* "to give an indication of" in Example 4 and Example 5 reflects Xenophon's interest in the scientific method. However, unlike *πεῖραν λαμβάνω* "to put to the test" and more like *ἱππασίαν ποιέομαι* "to do horsemanship / horse exercise," he availed himself of the option of qualifying just one component of the support-verb construction. Thus, in both Example 4 and Example 5, we find adjectives of manner qualifying the noun (*πάνυ σαφῆ* "very clear" and *ἱκανά* "strong").

[57]

τοῦ μὲν τοίνυν ἔτι ἀδαμάστου πώλου δῆλον ὅτι τὸ σῶμα δεῖ δοκιμάζειν· τῆς γὰρ ψυχῆς οὐ πάνυ σαφῆ τεκμήρια παρέχεται ὁ μήπω ἀναβαινόμενος.

It is obvious that it is necessary to assess the body (sc. only) with regard to a still unbroken colt. For the one who has not yet been mounted does not give any clear indication of his temperament.

Example 4. Xenophon, *On Horsemanship* 1.1

ὅσοι δ' ἂν πεπονηκότες ἐθέλωσι πάλιν ὑποδύεσθαι πόνους ἱκανὰ τεκμήρια παρέχονται ταῦτα ψυχῆς καρτερᾶς.

All those (horses) who are willing to delve into work again after having worked out give (this as) a strong indication of a steadfast character.

[31]

Example 5. Xenophon, *On Horsemanship* 3.11

The aspect indicated is specified in the form of an objective genitive in Example 4 and Example 5 (*τῆς ψυχῆς* "of the disposition" and *ψυχῆς καρτερᾶς* "of the steadfast character"). The simplex verb *τεκμαίρομαι* "to judge from signs" differs in meaning; the active *τεκμαίρω* "to make proof" comes closer to the meaning of the support-verb construction but is rare throughout the history of Greek as searches in the *Thesaurus Linguae Graecae* show (http://stephanus.tlg.uci.edu). Of the 51 instances only 3 are classical. These appear in Pindar's odes (6th / 5th c. BC) and Aeschylus' tragedies (6th / 5th c. BC).[32] Thus, it seems that the support-verb construction here may not only provide Xenophon with advantages as regards discourse cohesion but may also have filled a gap in the paradigm of the verb (see also [Aerts 1965] [Bentein 2016, 76] [Ledgeway and Vincent 2022]). Similarly, in the case of *ἐξουσίαν παρέχω* "to give the power to" in section 6.9, a related simplex verb only develops in later Greek.

[58]

4 Scientific Language as a Specialised Literacy: Cohesion through Register Continuity

Xenophon defines his intended audience at the beginning of his treatise as the next generation of horsemen (section 1.1 *τοῖς νεωτέροις τῶν φίλων* "the younger ones amongst the friends / companions"). Thus, he seems to expect his audience to belong to the same community of practice as himself, that is people interested in the purchase, care, and training of horses especially for cavalry purposes. A community of practice is a group of people that share knowledge and/or skills, often related to a specific area of expertise, through interactions [Unwin et al. 2007].[33]

[59]

Xenophon's treatise is only accessible to someone who has domain-specific literacy, that is a "competence or knowledge in a particular area" (OED s.v. literacy 2). However, in addition to basic background knowledge in the care and training of horses ([Maienschein 1998] science literacy), Xenophon's reader must also be familiar with his way of arriving at and presenting pieces of information (Maienschein's (1998) scientific literacy). This familiarity, the reader gains partly through engagement with the process of knowledge production in a scientific context (see also [Durant 1994] [Howell and Brossard 2021]), partly through engagement with the norms governing the communication of knowledge (the scientific register).

[60]

The support-verb constructions discussed in Section 3 reflect three aspects of scientific writing that are evident from Xenophon's text also elsewhere, that is (i) precision, through the use of specialised terms, (ii) methodical working, through the establishment of a clear structure, and (iii) incremental results, through the linking backwards and forwards in the discourse. We illustrate these three traits one by one before mapping them onto support-verb constructions.

[61]

Precision

In section 4.2, Xenophon considers the case of a horse that is unwell as indicated by the horse not taking in food normally (ὅταν [μὴ] ἐκκομίζῃ τὸν σῖτον ὁ ἵππος "when the horse does not take in his feed"). Xenophon considers the options of (i) too much blood in the body, (ii) exhaustion, and (iii) ailments such as laminitis (κριθίασις). κριθίασις is throughout Greek literature a word that is rare and which appears exclusively in treatises on horsemanship, as searches in the *Thesauraus Lingae Graecae* show.[34] κριθίασις seems etymologically related to κριθή "barley" / κριθάω "to be fed with barley." [Anderson 1961, 162 and 207] argues for it to refer to laminitis rather than colic, both specific illnesses affecting horses. Xenophon, by using the word, shows not only great attention to detail but displays familiarity with technical terms of the community of practice of horsemen. Moreover, κριθίασις fits in with the scheme of the creation of a "systematic scientific" vocabulary in the 5th / 4th c. BC by means of deverbal derivations in -σις, such as διαλέγω / διάλεξις "argumentation," περιλέγω / περίλεξις "circumlocution," καταλαμβάνω / κατάληψις "comprehension," κρούω / κροῦσις "attempts at deception" in Aristophanes, *Clouds* 316–318, all related to rhetoric [Galdi 2018, 242] [Langslow 2000] [Willi 2003, 67].

Methodical Working

The introduction to *On Horsemanship* shows Xenophon justifying himself as a competent author, stating unequivocally the objective of his work and aligning himself with past work in the same area of expertise.

[qualification of author] Ἐπειδὴ διὰ τὸ συμβῆναι ἡμῖν πολὺν χρόνον ἱππεύειν οἰόμεθα ἔμπειροι ἱππικῆς γεγενῆσθαι,

[objective of the work] βουλόμεθα καὶ τοῖς νεωτέροις τῶν φίλων δηλῶσαι ᾗ ἂν νομίζομεν αὐτοὺς ὀρθότατα ἵπποις προσφέρεσθαι.

[past research] συνέγραψε μὲν οὖν καὶ Σίμων περὶ ἱππικῆς, ὃς καὶ τὸν κατὰ τὸ Ἐλευσίνιον Ἀθήνησιν ἵππον χαλκοῦν ἀνέθηκε καὶ ἐν τῷ βάθρῳ τὰ ἑαυτοῦ ἔργα ἐξετύπωσεν·

[procedure] ἡμεῖς γε μέντοι ὅσοις συνετύχομεν ταὐτὰ γνόντες ἐκείνῳ, οὐκ ἐξαλείφομεν ἐκ τῶν ἡμετέρων, ἀλλὰ πολὺ ἥδιον παραδώσομεν αὐτὰ τοῖς φίλοις, νομίζοντες ἀξιοπιστότερα εἶναι ὅτι κἀκεῖνος κατὰ ταὐτὰ ἡμῖν ἔγνω ἱππικὸς ὤν· καὶ ὅσα δὴ παρέλιπεν ἡμεῖς πειρασόμεθα δηλῶσαι.

[qualification of author] Because we have been involved in horsemanship for a long time and believe to be experienced in horsemanship,

[objective of the work] we want to show to our younger friends how we think that they treat horses correctly.

[past research] In fact, also Simon has written about horsemanship, a man who offered a bronze horse at the Athenian Eleusinium and inscribed his own achievements on the pedestal.

[procedure] However, we do not leave out all those points in which we concur with him, but we will present these to the friends with great pleasure, considering them to be even more reliable because he too, such a great horseman, concurred. All those points which he has left out, we will try to clarify.

> **Example 6.** Xenophon, *On Horsemanship* 1.1

While the quoting of past works either verbatim or with modifications was a regular process in Greek literature [Adams and Ehorn 2015] [Finnegan 2011, chap. 8], Xenophon's scientific interest seems to be to fill gaps in the community's knowledge. He positions himself with regard to his contemporary's work in particular, Simon's *De forma et delectu equorum*, as Simon covered the same area of expertise. Xenophon's introduction in Example 6 is more than a condensed overview of the plot, it characterises him as a valid member of the community of practice to whose knowledge base he wants to contribute.[35]

Incremental Results

Throughout *On Horsemanship*, Xenophon facilitates it for the reader to tie pieces of the discourse together by means of interims summaries (e.g. sections 1.17, 2.3, 3.12, 9.1, and 10.17) and topic sentences foreshadowing what is to come next (e.g. 7.1, 9.1, and 12.1). In section 8.2, he explains why what may seem like repetition (διλογέω "to say twice") is necessary rather than redundant: ὅτε μὲν γὰρ ἐωνεῖτο, πειρᾶσθαι ἐκελεύομεν εἰ δύναιτο ὁ ἵππος ταῦτα ποιεῖν· νῦν δὲ διδάσκειν φαμὲν χρῆναι τὸν ἑαυτοῦ καὶ γράψομεν ὡς δεῖ διδάσκειν. "For when he bought (the horse), we said to try out whether the horse can do these things. But now we say that it is necessary to teach one's own horse and we will write down how this must be done." Xenophon's aim with these metatextual notes is to guide the reader through his work, so as to ensure that the reader appreciates the logical progression through the aspects discussed.

Support-verb constructions play a role in all three aspects – precision, methodical working, and incremental results. Firstly, they can develop, often through lexicalisation, a meaning that is different from that of the related simplex verb and often more specific. In Section 3, we saw the domain-specific support-verb construction τεκμήρια παρέχομαι "to give an indication of." Many of these lexicalised support-verb constructions are structurally rather fixed, although this is not a compulsory relation between structure and meaning ([Ledgeway and Vincent 2022]; see [Nunberg et al. 1994] vs. [Sheinfux et al. 2019].

Secondly, support-verb constructions seem to be involved in the development of a systematic vocabulary for the purposes of a community of practice. As mentioned, the 5th / 4th c. BC saw an increase in deverbal derivations in -σις in the context of creating a systematic terminology for areas such as rhetoric. These nouns were integrated into the predicate slot of a sentence by means of the support verb ποιέομαι. This is visible when drawing up the collocational field of common support verbs in a large corpus of literary classical Attic by means of Sketch Engine: ἄγω "to act," δέχομαι "to receive," δίδωμι "to give," ἔχω "to have," λαμβάνω "to take / receive," παρέχω "to provide," ποιέομαι "to do," τίθημι "to put," φέρω "to bring," τυγχάνω "to get", and χράομαι "to use".[36] All the candidates except for ποιέομαι form strong collocations with a small number of nouns, which can receive a support-verb construction reading [Pasquer et al. 2018, 2583]. By contrast, ποιέομαι enters into a large number of lose combinations with nouns in -σις (deverbal formations) and -ία (deadjectival formations). This resembles Galdi's findings for later Latin *facere* "to do" support-verb constructions, which specifically in technical contexts flourish with nouns in -*io* (deverbal formations), e.g. *Mulomedicina Chironis*, 26 *simili modo sanguinis detractionem in eis facies, sicut in prioribus demonstravi* "in the same way you will do the extraction of blood in them, as I have described before" (vs *detraho* "to extract") [Galdi 2018, 239–240, 242].[37] ποιέομαι support-verb constructions almost seem like a default systemic means to integrate the newly-created technical terms (cf. [Langer 2004]). Section 3 examined topical ἱππασίαν ποιέομαι "to do horsemanship / horse exercise".

Thirdly, support-verb constructions serve to tie pieces of information together within sentences (coherence) and across sentences (cohesion) by means of their discontinuity and analyticity, as explained in Section 3. This happens through bracketing pieces of information with the predicate (cf. Lysias, Speech 3.22 συνθήκας πρὸς αὐτὸν ποιησάμενος "to reach agreements with him"), establishing long-distance relationships through pronominalisation, and/or anaphora of one component (e.g. Plato, *Gorgias 479d* καὶ ἀεὶ τὸν ἀδικοῦντα τοῦ ἀδικουμένου ἀθλιώτερον εἶναι καὶ τὸν μὴ διδόντα δίκην τοῦ διδόντος "and that the wrongdoer is always more wretched than the wronged and the unpunished than the punished"), and the repetition of keywords (cf. ἱππασία in *On Horsemanship*). Furthermore, support-verb constructions can serve to focus attention on the event by eliminating participants which could act as distractors, as shown for πεῖραν λαμβάνω "to put to the test" in *On Horsemanship*. This allows Xenophon to focus the audience's attention on the logical progression of events in the treatise.

5 Summary and Conclusion

We have built a Sketch Engine corpus for the classical scientific treatise *On Horsemanship*. We used this corpus to detect discontinuous verbal multi-word expressions, specifically support-verb constructions. We examined how support-verb constructions – through their structural and lexical properties – aid discourse coherence and cohesion throughout Xenophon's treatise.

Section 1 introduced the means of achieving coherence, which refers to how building blocks of a sentence are tied together, and cohesion, which refers to the tying together of clauses and sentences, in classical Greek. Section 2 presented the corpus preparation tool, which we built to tag and lemmatise a classical Greek corpus for Sketch Engine, and the modified corpus configuration file to implement the corpus into Sketch Engine. Section 3 discussed the support-verb constructions in *On Horsemanship* from the angle of coherence and cohesion, especially *ἱππασίαν ποιέομαι* "to do horsemanship / horse exercise," *πεῖραν λαμβάνω* "to put to the test," and *τεκμήρια παρέχομαι* "to give an indication (of)." Section 4 outlined the features of scientific writing in *On Horsemanship* and mapped these onto the support-verb constructions identified.

70

We find that our Sketch Engine corpus performs with an error rate below 2.5%. Support-verb constructions are detected through the generation of collocations and n-grams, with specific items. This implies a focus on the identification of seen structures rather than discovery of new structures [Pasquer et al. 2020]. This limitation concerns research on support-verb constructions in general, e.g. [Cordeiro and Candito 2019] find that at best 31% of unseen support-verb constructions are detected by their tool. The issues surrounding automatic discovery of support-verb construction is linked to (i) their variability, both internally (e.g. inflection of the verb and the noun involved) and externally (e.g. their discontinuous structure) [Constant et al. 2017, 844] [Pasquer et al. 2018], (ii) their mostly discontinuous structure which necessitates reliance on syntactic analysis [Constant et al. 2017, 842] [Pasquer 2017, 169] [Savary et al. 2018, 112 and 123], and (iii) their semi-productivity such that unseen structures with some support verbs abound because they appear rarely [Nagy et al. 2013, 331] [Savary et al. 2018, 128]. We can determine the most frequent support verbs in classical Greek based on a large corpus of literary classical Attic and draw up collocational fields using Sketch Engine for each support verb in question. Sketch Engine has been successfully used for the study of support-verb constructions in English [Sheinfux et al. 2019], Italian, and Russian [Maiko 2020], for instance.

71

In the future, the performance of Sketch Engine on classical Greek could be improved by writing and implementing a dependency grammar for classical Greek in order to disambiguate the analysis of word forms that return a long list of options (e.g. *ἀνήμι* with *ἄνευ* and *ἄνω*). Furthermore, the integration of classical Greek in projects such as PARSEME, which provides manually annotated corpora for verbal multi-word expressions in 18 languages including modern Greek could improve discovery of new items (yet [Savary et al. 2018] on issues).[38]

72

We find that support-verb constructions in *On Horsemanship* aid coherence and cohesion throughout the treatise, both structurally and lexically speaking, as exemplified for the recurring support-verb constructions: *ἱππασίαν ποιέομαι* "to do horsemanship / horse exercise" contains the keyword *ἱππασία* which appears throughout the treatise and thus ties pieces of information together through lexical repetition, thus aiding cohesion. Aiding coherence structurally, *ἱππασίαν* in *ἱππασίαν ποιέομαι* is used to spread out pieces of information across a sentence while tying them together through their linking to one event (see adjectival modifiers with *ἱππασίαν*); *ἱππασίαν* in *ἱππασίαν ἔχω* "to have riding-related skill" is co-ordinated with another predicative noun thus condensing information by fusing two support-verb constructions under the same support verb.[39] *πεῖραν λαμβάνω* "to put to the test" focuses attention on the event by eliminating event participants from the surface structure (here the horse). This helps the reader not to lose the golden thread. *τεκμήρια παρέχομαι* "to give an indication of" fills a gap in the lexicon since *τεκμαίρω* (in the active voice) remains rare throughout the history of Greek. Using a support-verb construction allows Xenophon to express precisely what he wants to say.

73

Extra-linguistically, all three recurring support-verb constructions reflect an aspect of scientific writing: *ἱππασίαν ποιέομαι* is a representative of the creation of a systematic scientific vocabulary in the 5th / 4th c. BC and its integration in the predicate slot by means of highly productive *ποιέομαι* "to do" [Savary et al. 2018, 128], thus broadcasting the author's methodical working; *πεῖραν λαμβάνω* de-transitivises the event expressed and thus focusses the reader's attention on the logical progression of events by eliminating distractors (such as event participants); *τεκμήρια παρέχομαι* fills a gap in the lexicon, thus reflecting the author availing himself of each and any productive pattern in order to achieve maximum precision of the expression. Thus, rather than just through their structural and lexical properties, support-verb constructions in Xenophon's *On Horsemanship* seem to align with the register of the text and thus align Xenophon with the community of practice of his intended audience.

74

We found that support-verb constructions are a mixed bag, as has been observed by others for modern languages (e.g. [Kamber 2008] [Nunberg et al. 1994] [Pasquer 2017] [Sheinfux et al. 2019] [Tutin 2016]). In Xenophon's *On Horsemanship*, they seem to reflect specificity of expression [Storrer 2009] as well as markedness for register [Fendel 2020], rather than be related to the colloquial sphere or to stylistic choices [Zilliacus 1956] [Zilliacus 1967]. Their variability makes their automatic identification and especially discovery difficult even in morphologically less rich languages with unlimited corpora of data. We therefore suggest Sketch Engine as a tool for the further study of support-verb constructions while focusing on seen structures, in order to improve tools for the discovery of support-verb constructions in the future [Pasquer et al. 2020].

75

Appendix: Support-verb constructions in Xenophon's *On Horsemanship* [18 in total]

*In square brackets, the number of attestations of the verbal lemma in *On Horsemanship* is provided. Only instances in which the lemma acts as a support verb are listed.

76

1 ἄγω [9]	none						
2 δίδωμι [6]	none						
3 φέρω [6]	none						
4 τίθημι [1]	none						
5 πάσχω [6]	none						
6 κομίζω [0]	none						
7 κτάομαι [4]	none						
8 λαμβάνω [12]	πεῖραν λαμβάνω "to put to the test"	3.7 ἐπεὶ δὲ πολεμιστήριον ἵππον ὑπεθέμεθα ὠνεῖσθαι, ληπτέον πεῖραν ἁπάντων ὅσωνπερ καὶ ὁ πόλεμος πεῖραν λαμβάνει. ἔστι δὲ ταῦτα, τάφρους διαπηδᾶν, τειχία ὑπερβαίνειν, ἐπ' ὄχθους ἀνορούειν, ἀπ' ὄχθων καθάλλεσθαι· καὶ πρὸς ἄναντες δὲ καὶ κατὰ πρανοῦς καὶ πλάγια ἐλαύνοντα πεῖραν λαμβάνειν·					
9 τυγχάνω	ῥαστώνης	7.19 ὅταν γε	11.5 ἡμεῖς γε				

[4]	τυγχάνω "to get relief"	μὴν καταβαίνειν ἤδη καιρὸς ᾖ, μήτε ἐν ἵπποις ποτὲ καταβαίνειν μήτε παρὰ σύστασιν ἀνθρώπων μήτ' ἔξω τῆς ἱππασίας, ἀλλ' ὅπουπερ καὶ πονεῖν ἀναγκάζεται ὁ ἵππος, ἐνταῦθα καὶ τῆς ῥαστώνης τυγχανέτω.	μέντοι τὸ κράτιστον τῶν διδασκαλίων νομίζομεν, ὥσπερ ἀεὶ λέγομεν, ἢν ἐν παντὶ παρέπηται τὸ ἐν ᾧ ἂν ποιήσῃ τῷ ἀναβάτῃ κατὰ γνώμην τυγχάνειν ῥαστώνης παρ' αὐτοῦ.							
10 δέχομαι [10]	κόσμον δέχομαι "to get decorated / to get adorned"	12.2 τοῦτο γὰρ ἅμα κόσμον τε παρέξει καί, ἢν οἷον δεῖ εἰργασμένον ᾖ, δέξεται ὅταν βούληται τῷ ἀναβάτῃ τὸ πρόσωπον μέχρι τῆς ῥινός.								
11 χράομαι [13]	ὅπλοις χράομαι "to use weapons / to fight"	8.10 ἐπεὶ δὲ δεῖ ἐν παντοίοις τε χωρίοις τὸν ἱππέα ἀνὰ κράτος ἐλαύνοντα ἔποχον εἶναι καὶ ἀπὸ τοῦ ἵππου τοῖς ὅπλοις καλῶς δύνασθαι χρῆσθαι, ὅπου μέν ἐστι χωρία ἐπιτήδεια καὶ θηρία, ἄμεμπτος ἡ ἐν θήραις μελέτη τῆς ἱππικῆς·								
12 ποιέομαι [42]	ἱππασίας ποιέομαι "to do horse exercise"	8.9 Ὀρθῶς δ' ἔχει καὶ τὸ ἄλλοτε μὲν ἐν ἄλλοις τόποις, ἄλλοτε δὲ μακράς, ἄλλοτε δὲ βραχείας τὰς ἱππασίας ποιεῖσθαι. ἀμισέστερα γὰρ καὶ ταῦτα τῷ ἵππῳ τοῦ ἀεὶ ἐν τοῖς αὐτοῖς τόποις καὶ ὁμοίας τὰς ἱππασίας ποιεῖσθαι.	ἕλκη ποιέω / ποιέομαι "to get sores / to make sores for themselves"	5.1 πολλάκις γὰρ κνῶν ὁ ἵππος ἐπὶ τῇ φάτνῃ τὴν κεφαλήν, εἰ μὴ ἀσινὴς ἡ φορβειὰ περὶ τὰ ὦτα ἔσται, πολλάκις ἂν ἕλκη ποιοίη. ἑλκουμένων γε μὴν τούτων ἀνάγκη τὸν ἵππον καὶ περὶ τὸ χαλινοῦσθαι καὶ περὶ τὸ ψήχεσθαι δυσκολώτερον εἶναι.						
13 παρέχω[13]	τεκμήρια παρέχω "to give an indication"	1.1 τῆς γὰρ ψυχῆς οὔ πάνυ σαφῆ τεκμήρια παρέχεται ὁ μήπω ἀναβαινόμενος.	3.11 ὅσοι δ' ἂν πεπονηκότες ἐθέλωσι πάλιν ὑποδύεσθαι πόνους ἱκανὰ τεκμήρια παρέχονται ταῦτα ψυχῆς καρτερᾶς.	ἀσχολίαν παρέχω "to provide a lack of rest / to make restless"; ἀθυμίαν παρέχω "to provide a lack of confidence / to make insecure"	3.12 οἱ δὲ ἢ διὰ βλακείαν ἐλάσεως πολλῆς δεόμενοι ἢ διὰ τὸ ὑπέρθυμοι εἶναι πολλῆς θωπείας τε καὶ πραγματείας ἀσχολίαν μὲν ταῖς χερσὶ τοῦ ἀναβάτου παρέχουσιν, ἀθυμίαν δ' ἐν τοῖς κινδύνοις.	ἐξουσίαν παρέχω "to empower (to)"	6.9 ὁ μὲν γὰρ ἄγαν πρὸς αὑταῖς τυλοῖ τὸ στόμα, ὥστε μὴ εὐαίσθητον εἶναι, ὁ δὲ ἄγαν εἰς ἄκρον τὸ στόμα καθιέμενος ἐξουσίαν παρέχει συνδάκνοντι τὸ στόμιον μὴ πείθεσθαι.	φόβον παρέχω "to scare"	6.15 οἱ δὲ πληγαῖς ἀναγκάζοντες ἔτι πλείω φόβον παρέχουσιν·	κόσμον παρέχω "to decorate / to adorn,"	12.2 τοῦτο γὰρ ἅμα κόσμον τε παρέξει καί, ἢν οἷον δεῖ εἰργασμένον ᾖ, δέξεται ὅταν βούληται τῷ ἀναβάτῃ τὸ πρόσωπον μέχρι τῆς ῥινός.
14 ἔχω [37]	σχῆμα ἔχω "to have a shape / form"	1.8 καὶ βιάζεσθαι δὲ ἥκιστ' ἂν δύναιτο ὁ τοιοῦτον σχῆμα ἔχων καὶ εἰ πάνυ θυμοειδὴς εἴη·	ὑπόβασιν ἔχω "to have power to go forward"; ἱππασίαν ἔχω "to have riding-related skill"	1.14 τοῦτο δὲ ποιῶν ἅμα γοργοτέραν τε καὶ ἰσχυροτέραν ἕξει τὴν ὑπόβασίν τε καὶ ἱππασίαν καὶ ἅπαντα βελτίων ἔσται	χαλεπότητα ἔχω "to have difficulty"	3.10 δεῖ δὲ καὶ εἴ τινα χαλεπότητα ἔχοι ὁ ἵππος καταμανθάνειν, εἴτε πρὸς ἵππους εἴτε πρὸς ἀνθρώπους, καὶ εἰ				

| | | | | ἑαυτοῦ. | | δυσγάργαλίς γε εἴη· | | | | |

Table 3.

Abbreviations

NOM	nominative case (subject case in classical Greek)
ACC	accusative case (object case in classical Greek)
DAT	dative case (indirect object case in classical Greek)
OED	Oxford English Dictionary (www.oed.com)

Table 4.

Data

.vert files for Xenophon's *On Horsemanship*: https://gist.github.com/MatthewIreland/81e75b4653a3812fca2c02741ba21e34

Notes

[1] Thucydides, Histories I–V; Xenophon, Anabasis I–IV, Memorabilia I–IV, Hellenica I–IV; Antiphon, Speeches 1–6; Isocrates, Speeches 1–6 and 13; Isaeus, Speeches 1–8; Lysias, Speeches 1, 3, 7, 12, 14, 19, 22, 30–32; Demosthenes, Speeches 1–4, 6, 9, 18; Plato, Gorgias, Phaedrus, Republic I–III; Aristotle, Rhetoric, Politics I–III.

[2] They accommodate only the "full, auxiliary and modal verb" options [Scheible et al. 2013, 4]. Support verbs are neither full verbs nor auxiliaries, in that they retain a reduced argument grid [Butt 1997, 145f] [Cinque 2004, 172, n.31] [Loporcaro 2022, 216]. In the further uses, they only accommodate verb-prepositional phrase structures (cf. [Kamber 2008]). [Cap et al. 2015] however expand their approach to support-verb constructions in the form of verb-object structures.

[3] Classical and modern Greek differ significantly, such that e.g. the PARSEME shared task modern Greek corpus is not of use as training data.

[4] Automated tools often had to restrict either the number and type of verbs or nouns, thus limiting the range of support-verb constructions from the outset and rarely detecting candidates such as reconceptualised concrete nouns, e.g. in *to take a photo* [Radimský 2011], and verbs of realisation, e.g. *on a donné <imposé, infligé, collé, filé> à Jean une amende de 30 euros* [Mel'čuk 2004].

[5] The exact definition of support-verb constructions differs between researchers, not just approaches, such that the comparison of approaches and results is often difficult.

[6] Support-verb constructions are identified by means of the word-association measure of the log-likelihood. The log-likelihood is calculated based on item frequency.

[7] Concordances are vertical tables that put the attestations of the word form or lemma queried for in a corpus one underneath the other with the context to the left and the right. N-grams are sequences of lemmata or word forms that always appear in exactly the same order (e.g. English *in spite of* which forms a compound preposition would be a strong n-gram in a corpus of modern English).

[8] This distinguishes them from fully automated tools.

[9] See footnote 1.

[10] TEI describes a family of standards for electronically representing texts. More details on the TEI format can be found under https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html and https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html.

[11] Perseus contains tags regarding (i) verbal morphology, (ii) nominal morphology, (iii) part of speech, (iv) dialect, (v) genre, (vi) word formation, (vii) semantics, and the tag "rare" regarding the attestation of items. The tags found in the classical Attic literary corpus were grouped manually into the above categories. Xenophon's *On Horsemanship* did not show any tags that would have necessitated the creation of a new category.

[12] Note that Perseus does not currently have access to pre-defined multi-word-expression structures in the form of "listing words with spaces" or listing "idiomatic expressions" [Constant et al. 2017, 844].

[13] BetaCode does not restrict the use of characters that are reserved within a URI(reserved characters within a URI are defined in RFC 3986 and include forward slashes, back slashes, and apostrophes, which are common in BetaCode to encode acute accents, grave accents, and smooth breathings respectively).

[14] We owe thanks go to Barbara McGillivray (Turing Institute London) for letting us see the configuration file for her Latin Sketch Engine corpus for comparison.

[15] See footnote 1.

[16] Thucydides, Histories V; Xenophon, Anabasis I–IV; Isocrates, Speech 4; Lysias, Speeches 1, 3, 7, 12, 14, 19, 22, 30–32; Plato, Gorgias; Aristotle, Politics I.

[17] Lemmatised as ἀπολιμπάνω

[18] The preposition ἄνευ 'without' and the adverb ἄνω 'above' can be parsed as forms of ἀνίημι.

[19] ἄπειμι 'I will go' shares several forms with ἀφίημι.

[20] This is unlike in internal-object structures, such as English *to run a race*, in which the verb is the semantic head and the noun qualifies this head [van Emde Boas 2019, 364–365]. [Pompei 2006] argues for such structures to univerbate in the form of noun incorporation (e.g. οἰκοδομέω from οἶκος "house" + δομέω "to build"). The same cannot happen for support-verb constructions given their differing internal structure.

[21] The link between active, causative, and passive support-verb-construction patterns in the sense of the causative and passive as derivations of the active is captured in the idea of prototypes by [Gross 1998], [Mel'čuk 1996] [Mel'čuk 2004] [Kamber 2008].

[22] For a formal decision tree, which illustrates the many decisions to be made: https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.0/?page=060_Specific_tests_-_categorize_VMWEs/020_Light-verb_constructions (accessed 28 June 2021).

[23] Diachronically, there is the potential for univerbation under certain circumstances, e.g. *λόγον ποιέομαι* into later *λογοποιέομαι* "to report" [Creissels 2016] [Lehmann 2020] [Rosén 2020].

[24] In Greek, anaphora of a contextually salient support verb and / or predicative noun is permissible even without pronominalisation (e.g. Plato, *Gorgias* 479d *καὶ ἀεὶ τὸν ἀδικοῦντα τοῦ ἀδικουμένου ἀθλιώτερον εἶναι καὶ τὸν μὴ διδόντα δίκην τοῦ διδόντος* "and that the wrongdoer is always more wretched than the wronged and the unpunished than the punished").

[25] "The syntactic distance between two components is defined as the number of elements in the syntactic dependency chain between these two components, regardless of the direction of the dependencies and excluding the components themselves" [Pasquer et al. 2018, 2589].

[26] For later tendencies with *facere* "to do" in Latin, see however [Galdi 2018].

[27] The comparative *ἁμισέστερα* explicitly links the two instances of the support-verb construction. The type of link is not left to contextual inference.

[28] Support-verb constructions in Greek can be modified by both adverbs and adjectives. While the modification by adverbs is not constrained, the modification by adjectives is constrained as analysability and compositionality of the support-verb construction is required in many cases, except if the adjective is a fixed component of the structure (e.g. German *erste Hilfe leisten*).

[29] [Lipka 1981, 120]: "*Unter Lexikalisierung verstehe ich die Erscheinung, daß einmal gebildete komplexe Lexeme bei häufigem Gebrauch dazu tendieren, eine einzige lexikalische Einheit mit spezifischem Inhalt zu werden. Durch die Lexikalisierung geht der Syntagmacharakter in mehr oder weniger starkem Maße verloren*" [Lipka 1992]. Examples are compounds such as *housewife* [Hopper and Traugott 2003, 53] and *Feierabend*. The meaning of the compound is in both cases a lexical unit with a meaning different from the meaning of its parts.

[30] The same may be true of *φόβον παρέχω* in section 6.15: *οἱ δὲ πληγαῖς ἀναγκάζοντες ἔτι πλείω φόβον παρέχουσιν· οἴονται γὰρ οἱ ἵπποι, ὅταν τι χαλεπὸν πάσχωσιν ἐν τῷ τοιούτῳ, καὶ τούτου τὰ ὑποπτευόμενα αἴτια εἶναι.* "Those forced with beating show even greater fear. For the horses believe, when they suffer something terrible in a certain situation, that the thing which they are suspicious of is responsible (for that too)." *φόβον παρέχω* is here not causative ("to scare") but intransitive ("to show fear / to be afraid") [Ittzés 2007] [Ittzés 2015] [Machounis 2004] [Marini 2018].

[31] An anonymous reviewer pointed out correctly that the anaphoric pronoun *ταῦτα* functions as the object and *τεκμήρια* thus moves closer to a predicative element in Example 5. This is true and structures like this are otherwise omitted from the data collection. However, firstly, Greek support-verb constructions are on occasion found with an accusative object (e.g. [Ittzés 2007; but also [Lowe 2017]), and secondly, one could equally argue that *ταῦτα* is appositional or even parenthetical, thus functioning outside the sentence grammar [Koev 2022], or that the predicative relation works the other way round, i.e. with *ταῦτα* as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[32] Pindar, Olympia 6.73; Pindar, Nemea 6.8; Aeschylus, Prometheus vinctus 605. The fourth instance appears in Aratus, Phaenomena 1.18. Aratus' work is a didactic poem and dates slightly later (4th / 3rd c. BC).

[33] [Greer 2015] aligns Xenophon's treatise with modern approaches to horsemanship. While differences certainly exist, which may partly be due to the intended purpose of Xenophon's horses, his focus on the care and the behaviour of horses is noticeable (cf. [McGreevy 2012] [McIlwraith and Rollin 2011]).

[34] Apart from Xenophon, we find it in (i) the grammarian Aristophanes (3rd / 2nd c. BC), (ii) the rhetorician Pollux (2nd c. AD), (iii) in the compilation of the Hippiatrica (9th c. AD), and finally in the (iv) the encyclopaedia of the Suda (10th c. AD). 17 instances are attested in total.

[35] At the very end, in section 12.14, he reiterates almost like a disclaimer that his notes only pertain to the private person, whereas the cavalry leader is directed to another book.

[36] See footnote 1.

[37] The other context in which Galdi's Latin and our Greek "to do" support-verb constructions are common is multilingual contexts (e.g. translations), which pose their own problems [Bakker 2003, 132] [Myers-Scotton 2002, 134–139] [Reintges 2001] [Ronan 2012, 231] [Rutherford 2010, 203].

[38] https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.0/?page=060_Specific_tests_-_categorize_VMWEs/020_Light-verb_constructions (accessed 01 July 2022).

[39] The subject of *ἱππασίαν ἔχω* is the horse, thus riding-related skill is here in the sense of accepting and working with the rider.

Works Cited

Adams 2013 Adams, J. (2013). *Social variation and the Latin language.* Cambridge: Cambridge University Press.

Adams and Ehorn 2015 Adams, S., & Ehorn, S. (2015). "What is a composite citation? An introduction", in Adams S. & Ehorn S. (eds.) *Composite citations in antiquity.* Vol. 1, *Jewish, Graeco-Roman, and early Christian uses* [electronic resource], The Library of New Testament Studies, Vol. 525, pp. 1–16. New York: Bloomsbury T&T Clark.

Aerts 1965 Aerts, W. (1965). *Periphrastica: An investigation into the use of εἶναι and ἔχειν as auxiliaries or pseudo-auxiliaries in Greek from Homer up to the present day.* Publications (Universiteit van Amsterdam, Byzantijns-Nieuwgrieks Seminarium), Vol. 2. Amsterdam: AMHakkert.

Anderson 1961 Anderson, J. (1961). *Ancient Greek horsemanship.* Berkeley: University of California Press.

Bakker 2003 Bakker, P. (2003). "Mixed languages as autonomous systems", in Bakker P. & Matras Y. (eds.) *The mixed language debate: Theoretical and empirical advances, Trends in linguistics.* Studies and Monographs, Vol. 145, pp. 113–56. Berlin: Mouton de Gruyter.

Baños 2015 Baños, J. (2015). "Bellum gerere y proelium facere: Sobre las construcciones con verbo soporte en latín (y en griego)", in Muñoz Garcia de Iturrospe M. & Carrasco Reija L. (eds.) *Miscellanea Latina*, pp. 227–34. Madrid: Sociedad de Estudios Latinos.

Bentein 2013 Bentein, K. (2013). "Register and the diachrony of post-Classical and early Byzantine Greek", in *Revue belge de Philologie et d'Histoire*, 91/1: 5–44. DOI: 10.3406/rbph.2013.8407

Bentein 2016 Bentein, K. (2016). *Verbal periphrasis in ancient Greek: Have- and be- constructions.* Oxford: Oxford University Press.

Beschi 2018 Beschi, F. (2018). "The ancient Greek sentence left periphery: A study on Homer", *Journal of Greek Linguistics*, 18/2: 172–210. DOI: 10.1163/15699846-01802003

Biber and Conrad 2009 Biber, D., & Conrad, S. (2009). *Register, genre, and style.* Cambridge textbooks in linguistics. Cambridge: Cambridge University Press.

Bonifazi et al. 2016 Bonifazi, A., Drummen, A., & Kreij, M. de. (2016). *Particles in ancient Greek discourse: Exploring particle use across genres.* Hellenic Studies Series, Vol. 74. Washington, D.C.: Center for Hellenic Studies.

Bowern 2008 Bowern, C. (2008). "The diachrony of complex predicates", *Diachronica*, 25/2: 161–85.

Bowersock and Marchant 2014 Bowersock, G., & Marchant, E. (2014). *Xenophon. Hiero. Agesilaus. Constitution of the Lacedaemonians. Ways and Means. Cavalry Commander. Art of Horsemanship. On Hunting. Constitution of the Athenians.* Cambridge, MA: Harvard University Press.

Butt 1995 Butt, M. (1995). *The structure of complex predicates in Urdu.* Dissertations in Linguistics. Stanford: CSLI Publications.

Butt 1997 Butt, M. (1997). "Complex predicates in Urdu", in Alsina A., Bresnan J., & Sells P. (eds.) *Complex predicates*, pp. 107–49. Stanford: CSLI Publications.

Butt 2010 Butt, M. (2010). "The light verb jungle: Still hacking away", in Amberger M., Baker B., & Harvey M. (eds.) *Complex predicates: Cross-linguistic perspectives on event structure*, pp. 48–78. Cambridge: Cambridge University Press.

Butt and Lahiri 2013 Butt, M., & Lahiri, A. (2013). "Diachronic pertinacity of light verbs", *Lingua*, 135: 7–29. DOI: 10.1016/j.lingua.2012.11.006

Butt et al. 2013 Butt, M., Geuder, W., van Riemsdijk, H., & Corver, N. (2013). "On the (semi)lexical status of light verbs", in *Semi-lexical categories, Studies in generative grammar [SGG]*, Vol. 59, pp. 323–70. Berlin; Boston: Mouton De Gruyter. DOI: 10.1515/9783110874006.323

Cap et al. 2015 Cap, F., Nirmal, M., Weller, M., & Schulte im Walde, S. (2015). "How to account for idiomatic German support verb constructions in statistical machine translation." *Proceedings of the 11th Workshop on Multiword Expressions*, pp. 19–28. Denver, Colorado: Association for Computational Linguistics. DOI: 10.3115/v1/W15-0903

Celano 2013a Celano, G. (2013a). "Argument-focus and predicate-focus structure in ancient Greek", *Studies in Language*, 37/2: 241–66.

Celano 2013b Celano G. (2013b). "Word order", in Giannakis, G. (ed.) *Encyclopedia of ancient Greek language and linguistics.* Leiden; Boston: Brill.

Cinque 2004 Cinque, G. (2004). "Restructuring and functional structure", in *Structures and beyond: The cartography of syntactic structures*, Vol. 3, pp. 132–91. Oxford: Oxford University Press.

Cock 1981 Cock, A. (1981). "ΠΟΙΕΙΣΘΑΙ: Ποιein. Sur les critères déterminant le choix entre l'actif Ποιein et le moyen ΠΟΙΕΙΣΘΑΙ," *Mnemosyne*, 34/1/2: 1–62.

Constant et al. 2017 Constant, M., Eryiğit, G., Monti, J., van der Plas, L., Ramisch, C., Rosner, M., & Todirascu, A. (2017). "Multiword expression processing: A survey", Computational Linguistics, 43/4: 837–92. DOI: 10.1162/COLI_a_00302

Cordeiro and Candito 2019 Cordeiro, S. R., & Candito, M. (2019). "Syntax-based identification of light-verb constructions", *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa)*, September 30 - October 2, Turku, Finland, pp. 97–104. Turku: Linköping Electronic Conference Proceedings.

Creissels 2016 Creissels, D. (2016). "Univerbation of light verb compounds and the obligatory coding principle", in Nash L. & Samvelian P. (eds.) *Approaches to complex predicates, syntax and semantics,* Vol. 41, pp. 46–69. Leiden ; Boston: Brill.

Didakowski and Radtke 2020 Didakowski, J., & Radtke, N. (2020). "Verwendung der deutschen stützverbgefüge mit adjektiven und ihre ermittlung mithilfe des DWDS-Wortprofils", in De Knop S. & Hermann M. (eds.) *Funktionsverbgefüge im fokus*, pp. 101–36. Berlin; Boston: Mouton De Gruyter.

Dik 1995 Dik, H. (1995). *Word order in ancient Greek: A pragmatic account of word order variation in Herodotus*, Amsterdam Studies in Classical Philology, Vol. 5. Leiden; Boston: Brill.

Doucet and Ahonen-Myka 2004 Doucet, A., & Ahonen-Myka, H. (2004). "Non-contiguous word sequences for information retrieval", in *Second ACL workshop on multiword expressions: Integrating processing*, July 2004, pp. 88–95. Barcelona: ACL Anthology. DOI: 10.3115/1613186.1613198

Durant 1994 Durant, J. (1994). "What is scientific literacy?", *European Review,* 2/1: 83–9.

Fendel 2020 Fendel, V. (2020). "Taking stock of support-verb constructions in journalistic French", *Xanthos: A Journal of Foreign Literatures and Languages*, 2: 13–44.

Fendel 2021 Fendel, V. (2021). "Greek in Egypt or Egyptian Greek? Syntactic regionalisms (IV CE)", in Bentein K. & Janse M. (eds.) *Varieties of post-Classical and Byzantine Greek, Trends in linguistics*, Studies in Monographs, Vol. 331, pp. 115–40. Berlin: Mouton De Gruyter.

Finnegan 2011 Finnegan, R. (2011). *Why do we quote?: The culture and history of quotation.* Cambridge: Open Book Publishers.

Foley 2007 Foley, W. (2007). "A typology of information packaging in the clause", in Shopen T. (ed.) *Language typology and syntactic description: Volume 1: Clause structure*, 2nd ed., pp. 325–61. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511619427.006

Frenda 2017 Frenda, A. (2017). "Non-conventional arguments: Finite and non-finite verbal complementation in Sicilian", in Nolan B. & Diedrichsen E. (eds) *Argument realisation in complex predicates and complex events: Verb-verb constructions at the syntax-semantics interface, Studies in language companion series,* Vol. 180, pp. 117–36. Amsterdam; Philadelphia: John Benjamins.

Galdi 2018 Galdi, G. (2018). "On the use of facio as support verb in late and Merovingian Latin", *Journal of Latin Linguistics*, 17/2: 231–57.

Greer 2015 Greer, A. (2015). *Xenophon and the ancient Greek cavalry horse: An equestrian perspective* (PhD thesis). National University of Ireland, Galway, Galway.

Gross 1998 Gross, M. (1998). "La fonction sémantique des verbes supports", *Travaux de Linguistique : Revue Internationale de Linguistique Française,* 37/1: 25–46.

Halliday 1976 Halliday, M. (1976). *Cohesion in English*, in *English language series,* Vol. 9. London: Longman.

Hermann 2020 Hermann, M. (2020). "Über funktionsverbgefüge und verbale mehrwortverbindungen. Eine analyse am Beispiel von stellen." De Knop S. & Hermann M. (eds) *Funktionsverbgefüge im Fokus*, pp. 39–74. Berlin; Boston: Mouton De Gruyter.

Hopper and Traugott 2003 Hopper, P., & Traugott, E. (2003). "Grammaticalization", in *Cambridge textbooks in linguistics,* 2nd ed. Cambridge: Cambridge University Press.

Howell and Brossard 2021 Howell, E., & Brossard, D. (2021). *(Mis)informed about what? What it means to be a science-literate citizen in a digital world*, PNAS, 118: 1–8.

Hutchinson 2017 Hutchinson, G. (2017). "Repetition, range, and attention: The Iliad", in Tsagalis C. & Markantonatos A. (eds.) *The winnowing oar: New perspectives in Homeric studies*, pp. 145–70. Berlin; Boston: Mouton de Gruyter.

Ittzés 2007 Ittzés, M. (2007). "Remarks on the periphrastic constructions with the verb 'to make, to do' in Sanskrit, Greek and Latin", in Csaba D. (ed.) *Indian languages and texts Through the ages*. Essays of Hungarian Indologists in Honour of Prof. Csaba Töttössy, pp. 1–40. New Delhi: Manohar.

Ittzés 2015 Ittzés, M. (2015). "Light verb constructions in Vedic", *Manas: Tradition and modernity in Indian culture*, 2/1: 1–19.

Jackson 2016 Jackson, R. (2016). *The pragmatics of repetition, emphasis, and intensification* (PhD thesis). University of Salford, Salford.

Jiménez López 2011 Jiménez López, M. (2011). "El uso de ποιεῖσθαι en Lisias: Construcciones con verbo soporte", *Lingüística en la Red*, 9: 1–20.

Jiménez López 2012 Jiménez López, M. (2012). "Construcciones con verbo soporte, Verbo simple y nombre predicativo: Un ejemplo en griego antiguo / Support verb constructions, Simple verb and predicative noun: An example in ancient Greek", *Minerva*, 25: 83–105.

Jiménez López 2016 Jiménez López, M. (2016). "On support verb constructions in ancient Greek", *Archivio Glottologico Italiano*, 51/2: 180–204.

Jiménez López 2021 Jiménez López, María. (2021). "Γίγνομαι as the lexical passive of the support verb ποιέω in ancient Greek", in Giannakis G., Conti L., de la Villa J., & Fornieles R. (eds.) *Synchrony and diachrony of ancient Greek: Language, linguistics and philology*, pp. 227–40. Berlin; Boston: Mouton de Gruyter.

Kamber 2008 Kamber, A. (2008). *Funktionsverbgefüge — Empirisch: Eine korpusbasierte Untersuchung zu den nominalen Prädikaten des Deutschen*, Reihe Germanistische *Linguistik*, Vol. 281. Tübingen: Max Niemeyer.

Kilgarriff et al. 2014 Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., et al. (2014). "The sketch engine: Ten years on", *Lexicography*, 1/1: 7–36. DOI: 10.1007/s40607-014-0009-9

Koev 2022 Koev, T. (2022). *Parenthetical meaning*, in *Oxford studies in semantics and pragmatics,* Vol. 14. Oxford: Oxford University Press.

Lakoff and Johnson 1980 Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago; London: University of Chicago Press.

Langer 2004 Langer, S. (2004). "A linguistic test battery for support verb constructions", *Lingvisticæ Investigationes*, 27/2: 171–84. DOI: 10.1075/li.27.2.03lan

Langslow 2000 Langslow, D. (2000). *Medical Latin in the Roman Empire.* Oxford Classical Monographs. Oxford: Oxford University Press.

Ledgeway and Vincent 2022 Ledgeway, A., & Vincent, N. (2022). "Periphrasis and inflexion", in Ledgeway A., Smith J. C., & Vincent N. (eds.) *Periphrasis and inflexion in diachrony: A view from romance*, Oxford studies in diachronic and historical linguistics, pp. 11–60. Oxford: Oxford University Press.

Lehmann 2020 Lehmann, C. (2020). "Univerbation", *Folia Linguistica Historica*, 41/1: 205–52. DOI: 10.1515/flih-2020-0007

Lipka 1981 Lipka, L. (1981). "Zur lexikalisierung im Deutschen und Englischen", in Lipka L. & Günther H. (eds.) *Wortbildung, Wege der forschung*, Vol. 564, pp. 119–32. Darmstadt: Wissenschaftliche Buchgesellschaft.

Lipka 1992 Lipka, L. (1992). "Lexicalization and institutionalization in English and German", *Linguistica Pragensia*, 1: 1–13.

Loporcaro 2022 Loporcaro, M. (2022). "The morphological nature of person-driven auxiliation", in Ledgeway A., Smith J. C., & Vincent N. (eds.) *Periphrasis and inflexion in diachrony: A view from romance*, Oxford Studies in Diachronic and Historical Linguistics, pp. 213–37. Oxford: Oxford University Press.

Lowe 2017 Lowe, J. (2017). *Transitive nouns and adjectives: Evidence from early Indo-Aryan*. Oxford Studies in Diachronic and Historical Linguistics. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198793571.001.0001

Luraghi 2003 Luraghi, S. (2003). "Definite referential null objects in Ancient Greek", *Indogermanische Forschungen*, 108: 167–94. DOI: 10.1515/9783110243482.167

Machounis 2004 Machounis, P. (2004). "Nominalizations of English neutral verbs", in Leclère C., Laporte É., Piot M., & Silberztein M. (eds.) *Lexique, syntaxe et lexique-grammaire / Syntax, lexis & lexicon-grammar: Papers in honour of Maurice Gross, Lingvisticae Investigationes Supplementa*, Vol. 24, pp. 413–21. Amsterdam: John Benjamins.

Maienschein 1998 Maienschein, J. (1998). "Scientific literacy", Science, 281: 917. DOI: 10.1126/science.281.5379.917

Maiko 2020 Maiko, T. (2020). "What can you give in Italian that you can't give in Russian? A contrastive study of constructions with the light verbs dare in Italian and davat'/dat' in Russian", in Szerszunowicz J. & Awier M. (eds.) *Reproducible multiword expressions from a theoretical and empirical perspective*, pp. 33–53. Białystok: University of Bialystok Publishing House.

Marini 2010 Marini, E. (2010). "L'antipassivo in greco antico: ποιείσθαι come verbo supporto in Aristotele", *Journal of Latin Linguistics*, 11/1: 147–80. DOI: 10.1515/joll.2010.11.1.147

Marini 2018 Marini, E. (2018). "La fonction support et ses facettes: Facere [+support] {+causatif] dans le type sacra facere", in Spevak O. & Bodelot C. (eds.) *Les constructions à verbe support en latin, Cahier du Laboratoire de Recherche sur la Langage,* Vol. 7, pp. 129–47. Clermont-Ferrand: Presses Universitaires Blaise Pascal.

Mastronarde 2013 Mastronarde, D. (2013). "Ancient Greek tutorials (atticgreek.org) as complementary content for use with Introduction to Attic Greek", 2nd ed. Berkeley, CA: University of California Press.

McGillivray and Kilgarriff 2013 McGillivray, B., & Kilgarriff, A. (2013). "Tools for historical corpus research, and a corpus of Latin", in Bennett P., Durrell M., Scheible S., & Whitt R. (eds.) *New methods in historical corpora, korpuslinguistik und interdisiyiplinäre perspektiven auf sprache*, Vol. 3, pp. 247–57. Tübingen: Narr.

McGreevy 2012 McGreevy, P. (2012). *Equine behavior: A guide for veterinarians and equine scientists*, 2nd ed. Edinburgh: Saunders Elsevier.

McIlwraith and Rollin 2011 McIlwraith, C. W., & Rollin, B. (2011). *Equine welfare*. UFAW Animal Welfare Series. Oxford: Wiley-Blackwell.

Mel'čuk 1996 Mel'čuk, I. (1996). "Lexical functions: A tool for the description of lexical relations in a lexicon", in Wanner L. (ed.) *Lexical Functions in lexicography and natural language processing*, pp. 37–102. Amsterdam: John Benjamins.

Mel'čuk 2004 Mel'čuk, I. (2004). "Verbes supports sans peine", *Lingvisticæ investigationes*, 27/2: 203–17. DOI: 10.1075/li.27.2.05mel

Myers-Scotton 2002 Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Nagy et al. 2013 Nagy, I., Vincze, V., & Farkas, R. (2013). "Full-coverage identification of English light verb constructions." *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 329–37. Presented at the IJCNLP 2013, October, Nagoya, Japan: Asian Federation of Natural Language Processing.

Nunberg et al. 1994 Nunberg, G., Sag, I., & Wasow, T. (1994). "Idioms", *Language*, 70: 491–538.

Pasquer 2017 Pasquer, C. (2017). "Expressions polylexicales verbales : Étude de la variabilité en corpus (Verbal MWEs : A corpus-based study of variability)", in Flamein H. & Parmentier Y. (eds.) *Actes des 24ème conférence sur le traitement automatique des langues naturelles*. 19es Rencontres jeunes Chercheurs en Informatique pour le TAL (RECITAL 2017), Orléans (France), pp. 161–74. ATALA.

Pasquer et al. 2018 Pasquer, C., Savary, A., Ramisch, C., & Antoine, J.-Y. (2018). "If you've seen some, you've seen them all: Identifying variants of multiword expressions." *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 2582–94. Presented at the COLING 2018, Santa Fe, New Mexico: Association for Computational Linguistics.

Pasquer et al. 2020 Pasquer, C., Savary, A., Ramisch, C., & Antoine, J. (2020). "Verbal multiword expression identification: Do we need a sledgehammer to crack a nut?". The 28th *International Conference on Computational Linguistics* (COLING-20). Barcelona, Spain: ACL Anthology.

Pompei 2006 Pompei, A. (2006). "Tracce di incorporazione in greco antico", in Cuzzolin P. & Napoli M. (eds.) *Fonologia e tipologia lessicale nella storia della lingua greca*. Atti del Incontro Internazionale di Linguistica Greca, Bergamo, 15–17 September 2005, pp. 216–37. Milan: FrancoAngeli.

Radimský 2011 Radimský, J. (2011). "Noms prédicatifs, noms de résultat et noms concrets dans les constructions à verbe support", *Lingvisticæ Investigationes*, 34/2: 204–27. DOI: 10.1075/li.34.2.02rad

Reintges 2001 Reintges, C. (2001). "Code-mixing strategies in Coptic Egyptian", in Goldwasser O. & Sweeney D. (eds.) *Structuring Egyptian syntax: A tribute to Sarah Israelit- Groll, Lingua Aegyptia*, Vol. 9, pp. 193–237. Hamburg: Widmaier.

Ronan 2012 Ronan, P. (2012). *Make peace and take victory: Support verb constructions in old English in comparison with old Irish*, in *North-western European language evolution supplement*, Vol. 24. Odense: University Press of Southern Denmark.

Rosén 2020 Rosén, H. (2020). "Composite predicates in the layers of Latin", *Journal of Latin Linguistics*, 19/2: 231–79. DOI: 10.1515/joll-2020-2009

Rutherford 2010 Rutherford, I. (2010). "Bilingualism in Roman Egypt? Exploring the archive of Phatres of Narmuthis", in Evans T. & Obbink D. (eds.) *The language of the papyri*, pp. 198–207. Oxford: Oxford University Press.

Rychlý 2008 Rychlý, P. (2008). "A lexicographer-friendly association score", in Sojka P. & Horak A. (eds.) *Proceedings of recent advances in Slavonic natural language processing*, RASLAN 2008, pp. 6–9. Brno: Masaryk University.

Rysová 2017 Rysová, K. (2017). "Possibilities of text coherence analysis in the Prague Dependecy Treebank", in Menzel K., Lapshinova-Koltunski E., & Kunz K. (eds.) *New perspectives on cohesion and coherence: Implications for translation and multilingual natural language processing*, Vol. 6, pp. 35–48. Berlin: Language Science Press.

Sag et al. 2002 Sag, I., Baldwin, T., Bond, F., & Copestake, A. (2002). "Multiword expressions: A pain in the neck for NLP". Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing, pp. 1–15. Berlin; Heidelberg: Springer.

Sanroman Vilas 2009 Sanroman Vilas, B. (2009). "Diferencias semanticas entre construcciones con verbo de apoyo y sus correlatos verbales simples", *Estudios de lingüística*. Universidad de Alicante, 23: 289–314.

Savary et al. 2018 Savary, A., Candito, M., Mititelu, V., Bejček, E., Cap, F., Čéplö, S., Cordeiro, S., et al. (2018). "PARSEME multilingual corpus of verbal multiword expressions", in Markantonatou S., Ramisch C., Savary A., & Vincze V. (eds.) *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop*, pp. 87–147. Berlin: Language Science Press.

Scheible et al. 2013 Scheible, S., Schulte im Walde, S., Weller, M., & Kisselew, M. (2013). "A compact but linguistically detailed database for German verb subcategorisation relying on dependency parses from Web Corpora: Tool, guidelines and resource." *Proceedings of the 8th Web as Corpus Workshop (WAC-8) @Corpus Linguistics 2013*. Presented at the 8th Web as Corpus Workshop (WAC-8).

Sheinfux et al. 2019 Sheinfux, L., Greshler, T., Melnik, N., & Winter, S. (2019). "Verbal multiword expressions: Idiomaticity and flexibility", in Parmentier Y. & Waszczuk J. (eds.) *Representing and parsing of multiword expressions, Phraseology and multiword expressions*, Vol. 3, pp. 35–68. Berlin: Language Science Press.

Storrer 2009 Storrer, A. (2009). "Corpus-based investigations on German support verb constructions", in Fellbaum C. (ed.) *Idioms and collocations: Corpus-based linguistic and lexicographic studies, Research in corpus and discourse*, pp. 164–87. London: Continuum.

Tutin 2016 Tutin, A. (2016). "Comparing morphological and syntactic variations of support verb constructions and verbal full phrasemes in French: A corpus based study." *PARSEME COST Action. Relieving the pain in the neck in natural language processing: 7th final general meeting. Dubrovnil, Croatia.*

Unwin et al. 2007 Unwin, L., Hughes, J., & Jewson, N. (2007). *Communities of practice: Critical perspectives*. London: Routledge.

Webster 2019 Webster, J. (2019). "Key terms in the SFL model", in Thompson G., Bowcher W., Fontaine L., & Schönthal D. (eds.) *The Cambridge handbook of systemic functional linguistics*, pp. 35–54. Cambridge: Cambridge University Press.

Willi 2003 Willi, A. (2003). *The languages of Aristophanes: Aspects of linguistic variation in classical Attic Greek*. Oxford Classical Monographs. Oxford: Oxford University Press.

Willi 2010 Willi, A. (2010). "Register variation", in Bakker E. (ed.) *A companion to the ancient Greek language*, pp. 295–310. Chicester; Malden: John Wiley & Sons.

Zilliacus 1956 Zilliacus, H. (1956). "Zur Umschreibung des Verbums in spätgriechischen Urkunden", *Eranos*, 54: 160–6.

Zilliacus 1967 Zilliacus, H. (1967). "Zur Abundanz der spätgriechischen Gebrauchssprache", *Commentationes humanarum litterarum*, Vol. 41/2. Helsinki: Societas Scientiarum Fennica.

van Emde Boas 2019 van Emde Boas, E., Rijksbaron, A., Huitink, L., & de Bakker, M. (2019). *Cambridge grammar of classical Greek*. Cambridge: Cambridge University Press.