Discourse cohesion in Xenophon’s On Horsemanship through Sketch Engine

Victoria Beatrice Fendel <victoria_dot_fendel_at_classics_dot_ox_dot_ac_dot_uk>, University of Oxford
Matthew T Ireland <mt20_at_cam_dot_ac_dot_uk>, Sidney Sussex College, University of Cambridge

The classical Greek historian Xenophon (5th / 4th c. BC) is best known for his literary works (Anabasis, Helleneia) 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 16.1418ff), weeps after Patroclus’ death (Iliad 17.426ff), and prophecies Achilles’ destiny (Iliad 19.352ff). 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 horsekeeping. Conversely, Xenophon’s On Horsemanship is of interest in the latter. It reflects attention to detail [Greer 2015] along with literacy in scientific discourse.

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 interim summary.

Despite this clear sectioning, Xenophon’s text ‘flows’, that is information is underpinned by the following three research questions which underpin the following sections:

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.

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.

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 / 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.

Morpho-syntactically, coherence and cohesion can be achieved by means of structuring particles (e.g. μετ’ – “on the one hand – on the other hand,” οὖν “thus,” ὡς “for”) [Bonifazi et al. 2016], conjunctions (e.g. ὅτι “but” and δὲn “because”), and pronouns, which must have a clear anaphoric or kataphoric link to the item they stand in for (e.g. ἦν μετὰ ἔργα). He is very well. where “he” can only be used once we have established who it refers to) [Halliday 1976] [Luraghi 2003] [Rysoval 2017]. Moreover, word-order choices play a role since Greek word order is information-structure driven [Celano 2013a] [Celano 2013b] [Dir 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 νῦν βάφθη δέ καί παραπληγήνωσα εἰς τὰς ἱστορίας κατὰ τὰν ἰστοὺς 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 vòpò is the topic component, an unmarked element follows [Beschi 2018, 181–182]. In Sophocles, Oedipus Colonus 1656 Μίλησα δ’ ὤμους κατὰ τὸν δούλον ὡς ἢ διακρίνειν ὡς ἢ διακρίνειν death.dat – yet. particle – what.dat – that.nom – destroy. aor.pass.3sg “Yet through which type of death did he die?” πόπος is the topic component, οἰμοὺς 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].

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. ἑρμηνεία “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.”

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
nom 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] [Will 2000] [Will 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 *inmáto te x mòqo* “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 *inmáto te x mòqo* there would be *inméc* “to ride” and *inmáfi* “to drive / to ride” (Xenophon, On Horsemanship 10.13), which are regular denominational formations (in- ‘cu’ and -diqalqapa) [van Emde Boas 2019, 274–275].

Support-verb constructions are characterized 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 MEWs [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:

\[
F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \times \text{true positives}}{2 \times \text{true positives} + \text{false positives} + \text{false negatives}}
\]

*Fp* = “true positive” (a support-verb construction the algorithm returns that is one)

*Fp*

*Fp* = “false positive” (a support-verb construction the algorithm thinks is one but is not)

*Fn* = “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://ttypo.uni-konstanz.de/parsemef/index.php/results/shared-task) for 18 modern languages, and very large and expandable corpora (e.g. [Scheible et al. 2013]). 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, it must be borne in mind that support-verb constructions are language-specific, that cross-linguistic studies can only point the direction (e.g. *eine Reise machen vs to take a trip.* Finally, contingency / 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 συνέκρισε προς αὐτὸν παραιτήματος agreements – with him* – making “making agreements with him.”)

Fully automated approaches rely on algorithms that classify the input data into output categories (here support-verb construction vs not support-verb construction). These are trained on a suitably prepared training corpus. Techniques of analysis combating 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]. [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] [McGilivray and Kilgarriff 2013]). 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” – ὁρᾶν present, ὅραμα / ὅρασιμον / ὅρασιμον future, ἔδω / ἔδωσα / ἔδωσα / ἔδωσα / ἔδωσα 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). 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]:

\[
\text{logDice} = 14 + \log_2 \frac{2f_{x y}}{f_x + f_y}
\]

\*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 [8] 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-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/1JEuEVe1tOyB03qVUVuVtPQFv9x_8?usp=sharing](https://colab.research.google.com/drive/1JEuEVe1tOyB03qVUVuVtPQFv9x_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](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:
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.

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]

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]

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.

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.

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:

```java
Example 3.
```

```java
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.

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
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 .ver 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]

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[19], a manually annotated sub-sample of the literary classical Attic corpus[18], and On Horsemanship.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of Words</th>
<th>Number of Unique Errors in error.txt</th>
<th>Percentage of Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corpus of text</td>
<td>492,620</td>
<td>451</td>
<td>0.09%</td>
</tr>
<tr>
<td>Test sample</td>
<td>117,783</td>
<td>56</td>
<td>0.05%</td>
</tr>
<tr>
<td>On Horsemanship</td>
<td>7,144</td>
<td>0</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 1. errors.txt File

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

<table>
<thead>
<tr>
<th>Lemma</th>
<th>Thesaurus Linguae Graecae</th>
<th>Sketch Engine Test Sample</th>
<th>LogDice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LogDice</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Co-occurrence with Predicative Noun</td>
<td>LogDice</td>
</tr>
<tr>
<td>δίηη (Predicative Noun)</td>
<td>114</td>
<td>n/a</td>
<td>113</td>
</tr>
<tr>
<td>λομβάζημα</td>
<td>257</td>
<td>28</td>
<td>11.27</td>
</tr>
<tr>
<td>δίδυμον</td>
<td>171</td>
<td>53</td>
<td>12.57</td>
</tr>
<tr>
<td>ἀπολείπημα</td>
<td>15</td>
<td>o</td>
<td>n/a</td>
</tr>
<tr>
<td>φέμηυ</td>
<td>92</td>
<td>3</td>
<td>8.898</td>
</tr>
<tr>
<td>συμψυχά (Predicative Noun)</td>
<td>55</td>
<td>n/a</td>
<td>55</td>
</tr>
<tr>
<td>ποιύκα</td>
<td>608</td>
<td>17</td>
<td>9.71</td>
</tr>
<tr>
<td>ἀνήμα</td>
<td>8</td>
<td>2</td>
<td>10.02</td>
</tr>
<tr>
<td>ἀρµήκη</td>
<td>34</td>
<td>2</td>
<td>9.52</td>
</tr>
<tr>
<td>ἀπηλου (Predicative Noun)</td>
<td>75</td>
<td>n/a</td>
<td>75</td>
</tr>
<tr>
<td>ἔχωυ</td>
<td>760</td>
<td>11</td>
<td>8.75</td>
</tr>
<tr>
<td>παράδωκμου</td>
<td>58</td>
<td>12</td>
<td>11.53</td>
</tr>
</tbody>
</table>

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.

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 that are misspelled, and (iv) any

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 recontextualised as predicative (e.g. I took a picture of him [Radiniski 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 roδικα “to do” [Baños 2015] [Cock 1981] [Fendel 2021] [Jiménez López 2011] [Jiménez López 2012] [Jiménez López 2016] [Jiménez López 2021] [Marinó 2010] [Zilliacus 1995] [Zilliacus 1987].

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

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 Raddke 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 pronominisation (e.g. I had a great idea...
In Xenophon's short treatise, three support-verb constructions recur, *ἵππασία* *μοιεύει* “to do horse exercise” including the topical keyword *ἵππασία*, along with *κυριεύρει παράρχει* “to give an indication” and *τιάριν ψάλλων* “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.

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

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

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

In Example 1, *ἵππασία* is combined with the support verb *μοιεύει* to mean “to do horse exercise / to exercise the horse.” The qualifying adjectives (*μακρὸς “long,” *βραχεῖας “short,” and *όμοιος “similar”) are spaced out across the sentence.

*Ὅρδες δ’ ἵππασιν καὶ τὰ άκρας μὲν ἐν ἄλλοις τόποις ἄλλος ἀκράς, ἄλλησ ἀκράς ἀκράς δὲ βραχεῖας τῆς ἱππασίας ποιοῦσα. ὁμοίεστα γὰρ καὶ ταῦτα τὰ ἱππασιώτατα ἐν τοῖς αὐτοῖς τοῖς τάπητοι οἷος ἔν τοῖς αὐτοῖς τοῖς τάπητοι ἔν τοῖς τάπητοι ἱππασιώτατοι ποιοῦσα.*

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 (*μακρὸς, βραχεῖας*) 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 [cf. *κυριεύρει παράρχει* “to give an indication”] highlighting what is to be avoided when training a horse.

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. 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.

In Example 2, *ἵππασία* is combined with the support verb ἤρια 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 ὁμοίωσις “movement.”

Yesterday, I suddenly had it on the train to London. Finally, it allows for the noun to be used recuriously in the discourse rather than the support-verb construction at all times (see *ἵππασία* “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).[23] This allows for the bracketing of pieces of information, thus assigning them unequivocally to the support-verb construction, e.g. Lysias, Speech 3.22 συνήθειας *πολλὰς αὐχενίας* 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 [Sannom Vlas 2009] [Storer 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 belong to / to have explains the meaning of the support-verb construction [Hermann 2020, 58–61]; in to take a picture, we need to recontextualize the meaning of the noun to refer to the process resulting in the object rather than the concrete object [Radinsky 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] [Shinifux 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. Rather, lexical affinity between the verb and the noun governs the creation of support-verb constructions.

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

Example 1. *ἵππασία* is combined with the support verb *μοιεύει* to mean “to do horse exercise / to exercise the horse.” The qualifying adjectives (*μακρὸς “long,” *βραχεῖας “short,” and *όμοιος “similar”) are spaced out across the sentence.

*Ὅρδες δ’ ἵππασιν καὶ τὰ άκρας μὲν ἐν ἄλλοις τόποις ἄλλος ἀκράς, ἄλλησ ἀκράς ἀκράς δὲ βραχεῖας τῆς ἱππασίας ποιοῦσα. ὁμοίεστα γὰρ καὶ ταῦτα τὰ ἱππασιώτατα ἐν τοῖς αὐτοῖς τοῖς τάπητοι οἷος ἔν τοῖς αὐτοῖς τοῖς τάπητοι ἔν τοῖς τάπητοι ἱππασιώτατοι ποιοῦσα.*

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 (*μακρὸς, βραχεῖας*) 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 [cf. *κυριεύρει παράρχει* “to give an indication”] highlighting what is to be avoided when training a horse.

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. 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.

In Example 2, *ἵππασία* is combined with the support verb ἤρια 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 ὁμοίωσις “movement.”
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].

Similar condensation appears with the one-off structures ἔχω / ἔθρυμα 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 adom” 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.

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 παρέχω λαμβάνω recurs three times in quick succession.

Example 3. Xenophon, On Horsemanship 3.7

Unlike ἔστι παρέχω, παρέχω λαμβάνω seems to assume a highly specific meaning, possibly acquired through lexicalisation.[29] παρέχω 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 “to examine / to test”) focuses the reader’s attention on 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 ἔστιν ἄνωθεν “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”) reflects Xenophon’s interest in the scientific method. However, unlike παρέχω λαμβάνω “to put to the test” and more like ἔστι παρέχω “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 (πάνω σαφῶ “very clear” and καθαρὰ “strong”).

Example 4. Xenophon, On Horsemanship 1.1

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 5. Xenophon, On Horsemanship 3.11

The aspect indicated is specified in the form of an objective genitive in Example 4 and Example 5 (ἡς ψυχῆς “of the disposition” and ψυχῆς καρτερᾶς “of the steadfast character”). The simplex verb παρέχω λαμβάνω “to judge from signs” differs in meaning; the active παρέχω λαμβάνω “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 ἔξουσιαν παρέχω “to give the power to” in section 6.9, a related simplex verb only develops in later Greek.

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 τῶν νευτέρων 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]

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).

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.
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 ἀλον ἢ ἄρτος “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 (κρίθη), κρίθης is throughout Greek literature a word that is rare and appears exclusively in treatises on horsemanship, as searches in the Thesaurus Linguae Graecae show. Κρίθης seems etymologically related to κρίνω “barley” / κρίθω “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, κρίθης 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 -φαία, such as διαλέγω / διάλεγες “argumentation,” περιέλθω / περιέλχομαι “circumlocution,” καταλαμβάνω / καταλαμβάνω “comprehension,” κρούω / κρούσις “attempts at deception” in Aristophanes, Clouds 316–318, all related to rhetoric [Galdi 2018, 242] [Langslow 2000] [Willis 2003, 677].

### 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] Επειδή διό τι συμβῆναι ἡμῖν πολὺν χρόνον ἵππεων οἰκεῖων ἑμένου ἵππης γεγονέναι,

[objective of the work] Βουλήσατε καὶ τοὺς γεώργους τῶν φίλων δηλούσα ἢν νομίζομεν αὐτὸν ἄρθρον ἵππης προσφέρεται.

[past research] συνέγραψε μὲν οὖν καὶ Σίμων περὶ ἵππης, ὡς καὶ ταῦτα τὸ Ἐλευσινὸν καθήγησιν ἰππών χαλκόων ανέθηκε καὶ ἐν τῷ βόθρῳ τὰ ἐκσκοτεῖσθαι ἔξετροισσάν.

[procedure] ἡμᾶς γιὰ μένοι τοὺς συνεχοῦς ταῦτα γνώνες έκεινα, οὐκ ἔχουσαν ἐκ τῶν ἴππων, ἀλλὰ πολὺ άνω παραδόθηκαν αὐτὰ τὰς φίλους, νομίζοντες δὲς αμποτηρεῖα ἐνά ἐκ θάλασσας καὶ ταῦτα ἐμὲ ἐγὼν ἴππης αὐτῷ· καὶ οὖν δὲ παράλει ἡμᾶς παρασχεῖται δηλούσα.

[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 Ehon 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.

### Incremental Results

Throughout On Horsemanship, Xenophon facilitates it for the reader to tie pieces of the discourse together by means of intermediaries 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 (διὸ λέγω “to say twice”) is necessary rather than redundant: μὲν γὰρ ἐνεξεπέδειξα, περιέπερα έκκεντρον μὲν δούλου ό ἴππος ταῦτα ποιεῖν· νῦν δὲ διδάσκασι φανὲν χρήσαν τοὺς ἐκατοντάκας καὶ γράφων ὡς δε διδάσασι. “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 παρέχουμα “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. [Shein 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 -φαία 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 ποιεῖμα. 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,” ἔμω “to have,” λαμβάνω “to take / receive,” παρέχω “to provide,” ποιεῖμα “to do,” πιθάμω “to put,” φέρω “to bring,” γιγάντω “to get,” and χρημάζω “to use.”[36] All the candidates except for ποιεῖμα form strong collocations with a small number of nouns, which can receive a support-verb construction reading [Pasquer et al. 2018, 2583]. By contrast, παρέχω enters into a large number of loose combinations with nouns in -φαία (deverbal formations) and -αι (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 -αι (deverbal formations), e.g. Mulomedicinae Chloris, 26 simili modo sanguinis detractionem in eis facies, sic ut in proritus demonstravi “in the same way you will do the extraction of blood in them, as I have described before” (vs detrahō “to extract”) [Galdi 2018, 239–240, 242].[37] παρέχω support-verb constructions almost seem like a default systemic means to integrate the newly-created technical terms [cf. [Langer 2004]]. Section 3 examined topical ἵππων ποιεῖμα “to do horsemanship / horse exercise.”

Thirdly, support-verb constructions serve to tie pieces of information together by means of 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 συνεκαθήμενος πρὸς αὐτῶν παραδόσεις “to reach agreements with him”), establishing long-distance relationships through pronominisation, and/or anaphora of one component (e.g. Plato, Gorgias 47bd καὶ ἐὰν τὸν ἀδικοῦντα ἀδικοῦντα εἶναι καὶ τὸν μὴ δίδουσα δικίον τὸν δίδους “and that the wrongdoer is always more wretched than the wronged and the unpunished than the punished”), and the repetition of keywords (cf. ἵππων 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 μὴν 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 μοιάζω “to do horsemanship” / horse exercise,” μετέχω λαμβάνω “to put to the test,” and τοιχομενον πορεύομαι “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.

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.

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. ὄνομα with ὄνομα and ὄνομα). 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).

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 μοιάζω “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 ἱππασία μοιάζω 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). ἱππασία in ἱππασία ἔγινε “to have a 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] μετέχω 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. τοιχομενον πορεύομαι “to give an indication of” fills a gap in the lexicon since τοιχομενον (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.

Extra-linguistically, all three recurring support-verb constructions reflect an aspect of scientific writing: ἱππασίαν μοιάζω 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 μοιάζω “to do” [Savary et al. 2018, 128], thus broadcasting the author’s methodical working: μετέχω 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): τοιχομενον πορεύομαι 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. 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] [Tulín 2016]). In Xenophon’s On Horsemanship, they seem to reflect specificity of expression [Storror 2009] as well as markedness for register [Fendel 2020], rather than be related to the colloquial sphere or to stylistic choices [Zílliacus 1956] [Zílliacus 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].

**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.*

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ἄγω [9]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>2 δίδωμι [6]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>3 φέρω [6]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>4 τίθημι [1]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>5 πάσχω [6]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>6 κοιμίζω [9]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>7 κτάο [4]</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>8 λαμβάνω [12]</td>
<td>πείραν λαμβάνω “to put to the test”</td>
<td>3.7 ἔτει ἐκ τῆς πολεμιστήριας ἱππον ὑποθέτωμα ὑπερθανεῖ. Λιποτέντα πείραν ἀνάμειν ἀνόμων ἀξίων καὶ ὁ πάλμως πείραν λαμβάνει. Εἰς τῇ ταΐᾳ, τῷ πάροικῳ τἀνθρώπων, τεχνὴ ὑπερθανείναι, ἐν ὅρμοις ἀνορθόν, ἐν ὁρίζων καθάλλεσθαι· καὶ πρός ἀναντες ὑπερθανεῖ. 7.7 ὄμοις ἕκται 7.19 ὅτι ὀνειρεύσαι ἐλαύνουσα περαν λαμβάνειν.</td>
</tr>
<tr>
<td>9 τυχέω [9]</td>
<td>ἔφασιν</td>
<td>11.5 ὃμεις ἀνεῖ</td>
</tr>
</tbody>
</table>
10 δέχομαι
κόσμον δέχομαι
“to get decorated / to get adored”
12.2 τούτω γάρ ἀμα κόσμον τε παρέξεαι καὶ, ἢν οὖν δὲν ἐγραμμένον ἢ, δέξεταί ὅπως ἀναβάσθη τὸ πρόσωπον μέχρι τῆς ῥύνος.

11 χράομαι
ἐπιτελεὶς χράομαι
“to use weapons / to fight”
8.10 ἔτει δὲ δεῖ ἐν παντοῖο τοῦ χιώνον τὸν ἱππίον ἀνά κρατοῦ εἴλαντον ἔπεμον μὲ καὶ ἀν τῷ ἱπποῦ τοῦ ἄλος καὶ καλὸς δωρεᾶθα χρῆσθαι, ὅπου μὲν ἡ χιών ἐποίητε καὶ θηρία, ἀρμονεῖς ἢ ἐν θηρεῖας μικρός τῆς ἱππικῆς.

12 ποιέω
ἱππασίας ποιέω
“to do horse exercise”
8.9 Ὠθέω δὲ ἔχει καὶ τὸ ἄλλο τὸ ἄλλο τῷ τῆς ἱππικῆς ἀνάγκης ἐχθρία τοῦ τῶν ἑκομίσθη τοῦ ἄλλοτε, ἡμεῖς ἦς τοῖς ἱπποῖς καὶ ἢπο ἔχει τὸ ἱππότης.

13 παρέχω[13]
τεκμήρια παρέχω
“to give an indication”
1.1 τῆς γὰρ ψυχῆς οὐ πάντως σαφῆ τεκμήρια παρέχει ὁ μῆτις ἀναβατινόμονος.

14 ἔχω[37]
σχῆμα ἔχω ἢ ἄλλο ἄλλο ἄλλο
“to have a shape / form”
1.8 καὶ βιβλιοθήκη δὲ ἀδικία ἢ δύναμιν οὐκ ὅτι οὐκ ΣΧΗΜΑ ἔχει καὶ ἢ πάντως ὅτι τοιαύτης ἐγν.

μέντοι τὸ κρατοῦν τῶν διδασκαλίων νομίμων, ὅπως ἢ δὲ λέγωμεν, ἢν ἐν πάνι παρέχεται ἡ, ἢν ἐν πάνι παρέχεται 

πάλιν ἔχει τοὐσ ἀθυ 

παρέχεται ἡ, ἔχει τοῦ ἀναβάσθη κατὰ γνώμην ἐντελῶς ἡ ἀπαλλάθης, καὶ εἰ.
The components themselves μ to univerbrate in the form of noun incorporation [19]. The preposition Lemmatised as Thucydides, Histories V; Xenophon, Anabasis I–IV; Isocrates, See footnote 1. BetaCode does not restrict the use of characters that are common in BetaCode to encode. See footnote 1. 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. Note that Perseus does not currently have found doc/en/html/HD.html 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/TEI.html. 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. 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]. 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). 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. See footnote 1. Table 3. Abbreviations

<table>
<thead>
<tr>
<th>NOM</th>
<th>nominative case (subject case in classical Greek)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>accusative case (object case in classical Greek)</td>
</tr>
<tr>
<td>DAT</td>
<td>dative case (indirect object case in classical Greek)</td>
</tr>
<tr>
<td>OED</td>
<td>Oxford English Dictionary (<a href="http://www.oed.com">www.oed.com</a>)</td>
</tr>
</tbody>
</table>

Table 4. Data

.data files for Xenophon’s On Horsemanship: https://gist.github.com/MatthewIreland/81e7b4653a3812fa2c02741ba21e34

Credits/funding

Required credit for Perseus: Texts provided by Perseus Digital Library, with funding from The Annenberg CPB/Project. Original versions available for viewing and downloading at http://www.perseus.tufts.edu/hopper'. This study has been funded by the Leverhulme Trust within the project "Giving gifts and doing favours: Unlocking Greek support-verb constructions" at the University of Oxford, UK.

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 [Kamber 2008]. Support verbs are neither full verbs nor auxiliaries, in that they retain a reduced argument grid [Butt 1997, 1459] [Cinque 2004, 172, n.31] [Lopušnaro 2022, 216]. In the further uses, they only accomodate 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 donón <implor, inglit, collé, filé> à Jean une amende de 30 euros [Mel’tuk 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 exactly the same order (e.g. English in which each forms a compound preposition would be a strong n-gram in a corpus of modern English).
[8] This distinguishes them from fully automated tools.
[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.
[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 ἓ ὁμολογήσω
[18] The preposition ἄνω “without” and the adverb ἄνω “above” can be parsed as forms of ὅνυπα.
[19] ὅνυπω “I will go” shares several forms with ἅτακτομοι.
[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]. [Pompeii 2006] argues for such structures to unverbate in the form of noun incorporation (e.g. οἶκος ὁμοιοῦμεν ἄνω “house” + ἄνω “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’tuk 1996] (Mel’tuk 2004) [Kamber 2008].
[22] For a formal decision tree, which illustrates the many decisions to be made: https://parsemefr.lis-lab.fr/parseme-fr-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. ἀφίνομαι into later ἀφίνομαι “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 475d καὶ ἀνείπων δὲ οἰκοδομὴν τοῖς ἀνείπων ἄνω καὶ τοῖς ἀθλιώτεροι καὶ τοῖς ἀδικοῦνταί ἀπὸ τῶν δεινῶν ἄνω καὶ τῶν δυσγάργαλίς “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].
At the very end, in the Suda.

Apart from Xenophon, we find it the behaviour of horses is noticeable (cf. [10th c. AD]). 17 instances are attested in total.

The comparative 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 no support is adpositional or even parenthetical, thus functioning outside the sentence grammar [Kloev 2002], or that the predictive relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.

[208x662]Ittzés 2015

[314x662]

[310x219]Press.

[311x628]predicative relation works the other way round, i.e. with no support as the predicative element. The parallel with Example 8 makes us include the passage here, yet with the caveats just outlined.