Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference “Dialogue 2023”

June 14–16, 2023

Enhancing the semantic and conceptual description of Ancient Greek verbs in WordNet with VerbNet and FrameNet: a treebank-based study

Erica Biagetti
University of Pavia
Piazza del Lino 2, 27100 Pavia (PV), Italy
erica.biagetti@unipv.it

Luca Brigada Villa
University of Pavia/University of Bergamo
Piazza del Lino, 2, 27100 Pavia (PV), Italy
luca.brigadavilla@unibg.it

Chiara Zanchi
University of Pavia
Piazza del Lino 2, 27100 Pavia (PV), Italy
chiara.zanchi01@unipv.it

Silvia Luraghi
University of Pavia
Piazza del Lino 2, 27100 Pavia (PV), Italy
silvia.luraghi@unipv.it

Abstract

This paper presents a pilot study intended to enhance the semantic and conceptual description of Ancient Greek verbs in WordNet with information from two other resources, VerbNet and FrameNet, and to enrich a treebank of Ancient Greek texts with semantic information extracted from the three resources. We provided semantic annotation for verbs based on their morphosyntactic behavior, and performed a number of queries in order to extract occurrences from the Ancient Greek treebank that intended to match the different meanings of each verb. The manual check of the data extracted shows that, in spite of a limited number of mismatches, our queries yielded reliable results. The queries can be further refined in the future and complemented with a rule-based algorithm to map frame elements to dependency structure.

Keywords: WordNet; VerbNet; FrameNet; Ancient Greek verbs; dependency treebanks
DOI: 10.28995/2075-7182-2023-22-1009-1020

Улучшение семантического и концептуального описания древнегреческих глаголов в WordNet с помощью VerbNet и FrameNet: исследование на базе трибанка

Эрика Биаджетти
Университет Павии
Пьяцца дель Лино, 2,
27100 Павия (PV), Италия
erica.biagetti@unipv.it

Лука Бригада Вилья
Университет Бергамо /
Университет Павии
Пьяцца дель Лино, 2,
27100 Павия (PV), Италия
luca.brigadavilla@unibg.it

Кьяра Занки
Университет Павии
Пьяцца дель Лино, 2,
27100 Павия (PV), Италия
chiara.zanchi01@unipv.it

Сильвия Лураги
Университет Павии
Пьяцца дель Лино, 2,
27100 Павия (PV), Италия
silvia.luraghi@unipv.it

Аннотация

В статье предлагаются результаты предварительного исследования, целью которого является расширение семантического и концептуального описания древнегреческих глаголов в WordNet с использованием информации из FrameNet и VerbNet. Более того, настоящее исследование способствует пополнению трибанка древнегреческого языка семантической информацией из всех вышеупомянутых источников. Аннотирование глаголов было выполнено в соответствии с их морфосинтаксическим поведением; запросы, сделанные нами

в трибанке древнегреческого языка, предназначены для извлечения вхождений глаголов в соответствии с их различными значениями. Несмотря на незначительное число несоответствий результатов запросам, ручная проверка показывает, что полученные результаты являются надежными.

Ключевые слова: WordNet; VerbNet; FrameNet; древнегреческие глаголы; dependency treebanks

1 Introduction

WordNets (WN) are lexical databases that store meaning in a relational way. They comprise nodes for lemmas to which meanings are associated in the form of synsets, i.e. sets of synonymous words and phrases accompanied by brief definitions. Lemmas are connected to each other through lexical relations, whereas semantic relations establish connections among synsets.

The Princeton English WN, which serves as template for the creation of new WNs, contains limited morphosyntactic information for verb senses and semantic features of their participants [5], [6]. Attempts to enrich individual WNs with morphosyntactic and/or semantic information have been produced for several languages (e.g. [29], [27]). In fact, details on the morphosyntactic behavior of verbs and on the semantic properties of their dependents can be extracted from other resources, notably VerbNet and FrameNet (see Section 2). These resources can be jointly consulted, together with other ones, at the Unified Verb Index (https://uvi.colorado.edu). The reason for choosing these two resources over others was that VerbNet links syntactic and semantic patterns, which is ultimately the aim of this study, while FrameNet provides frames that, though being intended to describe English lexemes, are more easily adaptable to other languages. One effort towards integrating information from the three resources is described in Stayannova and Leseva [27]. The authors expanded the semantic and conceptual description of verbs in WordNet by combining information from FrameNet and VerbNet. They further supplemented the verbs' description with syntactic patterns extracted from the semantically annotated corpus BulSemCor ([15], [16]) in which all tokens are assigned a WordNet synset.

Building on the work by Stayannova and Leseva [27], in this paper, we present a pilot study for mapping semantic information onto Ancient Greek (AG) verbs contained in the Ancient Greek WordNet (AGWN) with morphosyntactic, semantic, and conceptual information provided by VerbNet and FrameNet. Notably, no semantically annotated corpus is available for AG. For this reason, we preliminarily enriched a selection of verbs in a morphosyntactically parsed corpus (treebank) of AG texts with semantic information from WordNet, VerbNet and FrameNet employing a semi-automatic methodology. The treebank contains various texts among which we selected two sub-corpora: the entire *Iliad* and *Odissey* (ca. 11th century BCE) and a selection of texts from Attic orators (5th century BCE). This is a convenience sample which, though unbalanced (232.340 vs. 97.142 tokens respectively, see Table 2), shows the potential of our approach for studying verbal valency and verb meanings across different periods (Homeric vs. Classical) and genres (epic poetry vs. prose).

The paper is organized as follows. In Section 2, we introduce the language resources employed in this study. After providing the basics of Ancient Greek verbal morphosyntax and illustrating our methodology (Section 3), in Section 4 we discuss the results. Section 5 concludes the paper and presents future work.

2 Resources employed in this study

2.1 The Ancient Greek WordNet

The AGWN belongs to a family of WNs for ancient Indo-European languages, an ongoing project jointly developed at the University of Pavia, the University of Exeter, the Center for Hellenic Studies at Harvard University, the Alpheios Project, and the Catholic University of the Sacred Hearth in Milan [2]. Besides Ancient Greek, it currently also comprises WNs for Sanskrit and Latin.

The Sanskrit, AG and Latin WNs were designed to be interoperable with each other and to enable cross-linguistic comparison of linguistic structures. To enhance compatibility, the developers maximize use of the synsets deriving from the Princeton English WN. As in other WNs, lemmas can be assigned multiple synsets, which indicates polysemy. In this family of WNs, however, polysemy is also framed within a cognitive linguistic approach (e.g. [17], [28]), which entails assuming that all senses of a word can be organized in a structured semantic network consisting of literal and non-literal senses.

For each synset, the three WNs provide information on periods and literary genres of attestation, and optionally *loci*, i.e. exemplifying occurrences referred to by author and work. In the AGWN, a word sense can be assigned to one of the following periods: Archaic (8th-6th c. BCE), Classical (5th c.-323 BCE), Hellenistic (323-31 BCE), Roman (31 BCE-290 CE). Labels for genres are Historiography, Liturgical, Miscellany, Novel, Oratory, Philosophy (epistles, treatise, dialogue), Poetry, Technical, Theater, Treatise. For several genres, additional subgenres can be identified; for example, Poetry can take epic, lyric, idyll, satire, epyllion or didactic as sub-label, whereas Oratory can be political, juridical or can take the form of a homily. Diachronic and stylistic metadata are meant to enable studies on semantic change over time and across literary genres and authors.

2.2 VerbNet

VerbNet (VN) [14] is a network of English verbs that links their syntactic and semantic patterns. The network is organized into verb classes extending Levin's classes [18], with the understanding that the syntactic form of a verb and its arguments inform its semantics. Levin's classes were then refined, and subclasses were added to achieve syntactic and semantic coherence among verb members of a class. Each verb class in VN is described by semantic roles and selectional preferences of its arguments, as well as by frames consisting of a syntactic description and a semantic representation with subevent structure based on the Dynamic Event Model [24], [25]. VN groups together verbs with identical sets of syntactic frames and semantic predicate structures. Each VN class contains a set of syntactic frames, depicting the possible surface realizations of the argument structure for constructions such as transitive, intransitive, resultative, and a large set of diathetic alternations. Semantic restrictions (such as animate, human, organization) are used to constrain the types of semantic roles taken by the arguments (Agent, Patient, Instrument, etc.). Each syntactic frame is then associated with explicit semantic information, expressed by semantic predicates such as 'contact', 'manner' or 'cause'. Table 1 shows a complete entry for a frame in VerbNet class Hit-18.1:

Class Hit-18.1			
Semantic roles and restrictions: Agent[+int_control], Patient[+concrete], Instrument[+concrete]			
Members: bang, bash, blast, bop, click, dash, knap, hit, smite …			
Frames:			
Name	Example	Syntax	Semantics
NP V NP PP.instrument	*Paula hit the ball with a stick.*	Agent V Patient {with} Instrument	¬ CONTACT(e1 , Patient , Instrument) DO(e2 , Agent) UTILIZE(e2 , Agent , Instrument) MANNER(e2 , Directedmotion , Instrument) CONTACT(e3 , Patient , Instrument) MANNER(e3 , Forceful , Instrument) CAUSE(e2 , e3)

Table 1: Entry for a frame in VerbNet class Hit-18.1-1

2.3 FrameNet

FrameNet (FN) [26] is a lexical database of contemporary English which maps meaning to form through the theory of Frame Semantics [7]-[12]. The theory assumes that people understand the meaning of words largely by virtue of the larger conceptual structures, i.e. frames, which they evoke. FrameNet defines frames and their elements, and provides annotated sentences to show how frame elements fit syntactically around the lexical unit that evokes the frame. In the simplest case, the frame-evoking word is a verb, and the frame elements are its syntactic dependents. In (1), the Hit_target frame is evoked by the verb *hit*; the frame elements are the Agent instantiated by the pronoun *he,* the Target expressed by *the bull's-eye,* and the Instrument instantiated by the prepositional phrase *with his first arrow*:

(1) It looks like [$_{\text{Agent}}$ he] HIT [$_{\text{Target}}$ the bull's-eye] [$_{\text{Instrument}}$ with his first arrow]!

Note that FN frame elements are frame-specific semantic roles. Consequently, they are more fine-grained than VN semantic roles. For example, the Telling frame evoked by speech verbs has Speaker, Addressee and Message as core elements, whereas VerbNet has the more generic Agent, Recipient and Theme as semantic roles of the roughly corresponding transfer_mesg-37.1.1 VN class; similarly, the Awareness frame evoked by verbs of cognition like *know*, *comprehend*, and *conceive* takes the core elements Cognizer and Content, instead of the more generic semantic roles Experiencer and Stimulus.

2.4 Dependency treebanks

Treebank	Tokens	Text	Author
aeschines-1-1-50-bu1.conllu	4353	Against Timarchus	Aeschines
aeschines-1-51-100-bu1.conllu	3879		
aeschines-1-101-150-bu1.conllu	3975		
aeschines-1-151-196-bu1.conllu	3710		
antiphon-1-bu2.conllu	2034	Against the Stepmother for Poisoning	Antiphon
antiphon-2-bu2.conllu	2908	First Tetralogy	
antiphon-5-bu2.conllu	7411	On the Murder of Herodes	
antiphon-6-bu2.conllu	3989	On the Chorus Boy	
dem-59-neaira-2019.conllu	10457	Apollodorus Against Neaera	Demosthenes
demosthenes-18-1-50-bu2.conllu	4141	On the Crown	
demosthenes-18-51-100-bu1.conllu	4046		
demosthenes-18-101-150-bu2.conllu	4292		
demosthenes-18-151-200-bu2.conllu	4272		
demosthenes-18-201-275-bu1.conllu	5665		
demosthenes-18-276-324-bu1.conllu	3877		
demosthenes-1-bu1.conllu	2181	First Olynthiac	
demosthenes-4-phil1-bu1.conllu	3926	Philippic 1	
lysias-1-bu1.conllu	2826	On the Murder of Eratosthenes	Lysias
lysias-12-bu1.conllu	5614	Against Eratosthenes	
lysias-13-bu1.conllu	5629	Against Agoratus	
lysias-14-bu1.conllu	2786	Against Alcibiades 1	
lysias-15.conllu	685	Against Alcibiades 2	
lysias-19-bu1.conllu	3593	On the Property of Aristophanes	
lysias-23-bu1.conllu	893	Against Pancleon	
tlg0012.tlg001.perseus-grc1.1-6.tb.conllu	31966	Iliad	Homer
tlg0012.tlg001.perseus-grc1.7-12.tb.conllu	30011		
tlg0012.tlg001.perseus-grc1.13-18.tb.conllu	35600		
tlg0012.tlg001.perseus-grc1.19-24.tb.conllu	31550		
tlg0012.tlg002.perseus-grc1.1-6.tb.conllu	26559	Odyssey	
tlg0012.tlg002.perseus-grc1.7-12.tb.conllu	27144		
tlg0012.tlg002.perseus-grc1.13-18.tb.conllu	26661		
tlg0012.tlg002.perseus-grc1.19-24.tb.conllu	24633		

Table 2: Treebanks used for our study

Dependency treebanks are collections of morphosyntactically annotated texts, representing sentences as trees where each word corresponds to a node. The treebanks[1] we considered for our work (Table 2) consist in the conversion made by Francesco Mambrini of a series of treebanks of AG texts to the Universal Dependencies (UD) format, which provides a standard set of annotation guidelines for part-of-speech tags, morphological features, and syntactic dependencies.[2]

3 Methods

The queries presented in Section 3.2 are in line with Levin's [18] work and VN and rely on the assumption that certain morphosyntactic configurations in which specific Ancient Greek verbs occur (syntactic frames in VN terminology), which can be easily extracted from treebanks, are reliable predictors for verb senses in context. This means that, once the occurrences instantiating a syntactic frame are extracted, the annotation of the verbal nodes they contain can be easily enriched with semantic information from WN, VN and FN.

3.1 Basics of Ancient Greek verbal morphosyntax

In this paper, we take into account two aspects of verbal morphosyntax to which AG verb senses are sensitive: voice and argument structure. The verbal lemmas included in this case study (Table 3) are known to be good representatives of how voice and argument structure affect verb senses.

AG verbs distinguish three values for voice: active, middle, and passive (Act, Mid, Pass in Table 3). The middle and the passive voices are morphologically distinct only in some tenses (aorist and future), while others feature a mediopassive that instantiates both the middle and the passive diatheses (MidPass). The active/middle distinction can encode the (anti)causative alternation [13] in AG: *rhḗgnumi*(Act) 'break(tr.)' (as in *Rose broke the vase*) vs. *rhḗgnumai*(MidPass) 'break(intr.)' (as in *The vase broke*) [22] (queries 2a-d in Table 3 below). The same alternation is featured by some experiential verbs, such as *mimnḗskō*(Act) 'remind' vs. *mimnḗskomai*(MidPass) 'remember' and *kholóō*(Act) 'anger someone' and *kholóomai*(MidPass) 'be/get angry' [20] (queries 1a-c and 7a-b). Notably, not all experiential verbs behave in the same way in this respect, as shown by *lanthánō*(Act) 'escape notice' and *lanthánomai*(MidPass) 'forget' (queries 9a-b). In Homeric Greek (HG), morphologically passive verb forms sometimes do not express the passive diathesis (e.g. *The vase was broken [by Rose]*); rather, they are semantically equivalent to middles [1]. Voice distinctions often express meanings not related with either the (anti)causative alternation or with active/passive diatheses: this is the case of *phrázō*(Act) 'show, tell' vs. *phrázomai*(MidPass) 'plan, devise, realize' (queries 3a-c). In the case of *phobéō* the difference between the meanings of Pass and MidPass is not easy to tell given the verb's semantics: *phobéō*(Act) HG 'be set on flight' ClG 'be frightened' vs. *phobéomai*(MidPass) HG 'flee' ClG 'fear' (queries 8a-d). In both corpora the Pass can occur in contexts in which an agent is overtly realized or implied but this does not always need to be the case; in addition, in HG the active can occur with passive agent PPs and function as lexical passive; see [21].

AG nouns feature five cases: nominative, genitive, dative, accusative, and vocative. Subjects of transitive and intransitive verbs are encoded in the nominative. Second arguments of two-place verbs (e.g. *hit*) most frequently (74% of two-place verbs in HG; see [22]) feature the accusative case (NomAcc construction), but can also occur in the genitive and dative cases (NomGen and NomDat constructions). The dative case is used for third arguments of three-place verbs of transfer (e.g. *give*; NomAccDat construction). Semantic distinctions brought about by voice may be paired with changes in argument structure: *phrázō*(Act) 'show, tell' is a knowledge transfer verb, which takes the NomAccDat construction (the dative third argument is optional); *phrázomai*(MidPass) 'plan, devise, realize' is a cognition verb, featuring the NomAcc construction. The behavior of *orégō/orégomai* is partially similar (queries 6a-e): *orégō*(Act) 'give, assign' takes the NomAccDat construction with an optional dative third argument. However, *orégomai*(MidPass) is more complex: in Homeric Greek, it features the NomAcc and the NomGen construction with changes in meaning ('hit' vs. 'try to hit'; cf. [20]). In Post-Homeric Greek, only the NomGen construction remains with *orégomai*, but it displays a meaning shift, 'desire,

[1] https://github.com/francescomambrini/katholou.

[2] https://universaldependencies.org/.

wish', hence moving to the experiential domain [20]. Thus, changes in argument structure can result in different senses of the same verb; this is also the case of *oîda* 'know, be skilled', queries 4a-c, and *mnáomai* 'have in mind, woo', queries 5a-b. Another verb that shows a shift in argument structure is *punthánomai*, which can take both the NomAcc and the NomGen construction in HG, but is limited to the NomAcc construction in the ClG corpus we selected (queries 10a-b).

3.2 Extraction of the patterns from the treebanks

In order to infer synsets, verb classes and frames based on syntactic patterns, we chose ten AG lemmas (*mimnḗskō*, *rhḗgnumi*, *phrázō*, *oîda*, *mnáomai*, *orégō*, *kholóō*, *phobéō*, *lanthánō* and *punthánomai*) and developed 30 queries (Table 3) to extract their corresponding syntactic patterns from the treebank.[3] This process was facilitated by the tool UDeasy [3]. The ten verbs were chosen to exemplify a variety of different morphosyntactic behaviors (cf. Section 3.1). Regarding syntactic conditions, we imposed a criterion on all queries that required the direct object (obj) and indirect object (iobj) nodes, if present, to be directly dependent on the verb node.

We then provided semantic annotation with data from WN, VN and FN eliciting meanings from Liddel/Scott dictionary.[4] Annotation was performed by adding the relevant synsets, verb classes and frames to the MISC field of the CoNNL-U file, the standard format followed by UD treebanks to store annotation at sentence and token level. In the case of verbs that could in principle feature three different voice values, i.e. *mimnḗskō*, *rhḗgnumi*, *phrázō*, *orégō*, *kholóō*, *phobéō*, and *lanthánō*, we predicted that Pass instantiated the passive diathesis and accordingly we used the same semantic annotation for the Pass and the Act.[5] This prediction turned out to be correct in the case of *mimnḗskō*, but not in the case of *phrázō* and *kholóō*, which feature two and 11 Pass forms respectively with middle meaning (we corrected the annotation in Table 4 after the manual check). For *rhḗgnumi*, *orégō* and *lanthánō* we did not find any Pass forms. Concerning *phobéō*, as noted above, the verb's semantics blurs the difference between MidPass 'be afraid' and Pass 'be frightened'. Table 4 shows the semantic annotation aligning the three resources and the occurrences extracted through our queries from the two sub-corpora (HG and Classical Greek, ClG).

[3] Note that the treebank contains texts in Greek script; to facilitate reading, in this article we provide transcriptions of AG words in Latin alphabet.

[4] Henry George Liddell, Robert Scott, A Greek-English Lexicon (https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a1999.04.0057)

[5] The verbs *mnáomai* and *oîda* do not have Pass forms in AG.

id	verb lemma	node	conditions	texts
1a	μιμνήσκω (mimnḗskō)	verb obj iobj	verb: Voice=Act	all
1b			verb: Voice=MidPass|Mid	
1c			verb: Voice=Pass	
2a	ῥήγνυμι (rhḗgnumi)	verb obj iobj	verb: Voice=Act obj: Case=Acc	all
2b		verb obj iobj	verb: Voice=Pass	
2c			verb: Voice=MidPass|Mid	
2d			verb: Voice=Act obj: Case≠Acc	
3a	φράζω (phrázō)	verb obj iobj	verb: Voice=Act	all
3b			verb: Voice=MidPass|Mid	
3c			verb: Voice=Pass	
4a	οἶδα (oîda)	verb obj iobj	obj: Case=Acc	all
4b			obj: Case=Gen	
4c		verb	no obj	
5a	μνάομαι (mnáomai)	verb obj iobj	obj: Case=Gen	all
5b			obj: Case=Acc	
6a	ὀρέγω (orégō)	verb obj iobj	verb: Voice=Act	all
6b		verb obj iobj	verb: Voice=MidPass|Mid obj: Case=Acc	hom
6c			verb: Voice=MidPass|Mid obj: Case=Gen	
6d			verb: Voice=MidPass|Mid	posthom
6e		verb obj iobj	verb: Voice=Pass	all
7a	χολόω (kholóō)	verb obj iobj	verb: Voice=Act	all
7b			verb: Voice= MidPass|Mid|Pass	
8a	φοβέω (phobéō)	verb obj iobj	verb: Voice=Act	hom
8b			verb: Voice= MidPass|Mid|Pass	
8c			verb: Voice=Act	posthom
8d			verb: Voice= MidPass|Mid|Pass	
9a	λανθάνω (lanthánō)	verb obj iobj	verb: Voice=Act	all
9b			verb: Voice= MidPass|Mid|Pass	
10a	πυνθάνομαι (punthánomai)	verb obj iobj	obj: Case=Acc|Gen|no obj	hom
10b		verb obj iobj	obj: Acc iobj: Gen	posthom

Table 3: Conditions of the 30 queries
(underlined nodes are obligatory, non-underlined nodes are optional)

query	verb lemma	WordNet	VerbNet	FrameNet	HG	ClG
1a	μιμνήσκω (mimnḗskō)	v#00410666| v#00412253| v#00413589| v#01609936| v#00414046| v#00491159	characterize-29.2-1-1| care-88.1	Evoking	10	
1b		v#00412961	characterize-29.2-1-1	Memory	100	30
1c		v#00412961	tell-37.2	Evoking		3
2a	ῥήγνυμι (rhḗgnumi)	v#00231588	break-45.1	Cause_to_fragment	29	
2b		v#00231588	break-45.1	Cause_to_fragment		
2c		v#00231062	break-45.1	Breaking_apart	16	
2d		v#00231588	break-45.1	Breaking_apart	4	
3a	φράζω (phrázō)	v#00639889	transfer_mesg-37.1.1-1-1	Telling	22	11
3b		v#01134235| v#00493867	intend-61.2-1-1| comprehend-87.2-1-1-1	Making_arrangements| Coming_to_believe	83	
3c		v#01134235| v#00493867	intend-61.2-1-1| comprehend-87.2-1-1-1	Making_arrangements| Coming_to_believe	2	
4a	οἶδα (oîda)	v#00401762	comprehend-87.2-1-1-1	Awareness	151	80
4b		v#00402497	not found	Expertise	24	
4c		not found	not found	not found	119	125
5a	μνάομαι (mnáomai)	v#00491159	care-88.1	Cause_emotion|Awareness	3	
5b		v#01727931	compel-59.1-1	Forming_relationships	14	
6a	ὀρέγω (orégō)	v#01583087	conduct-111.1	Giving	18	
6b		v#00841893	hit-18.1-1	Hit_target	3	
6c		v#01153576	not found	Aiming	3	
6d		v#01245362	want-32.1	Desire		2
6e		v#01583087	conduct-111.1	Giving		
7a	χολόω (kholóō)	v#01789790	amuse-31.1	Stimulate_emotion	4	
7b		v#01790925	marvel-31.3	Experiencer_focused_emotion	65	
8a	φοβέω (phobéō)	v#01111362	banish-10.2	Beat_opponent (rout) Removing (expel)	15	
8b		v#02079709	escape-51.1	Fleeing	17	
8c		v#01783082	amuse-31.1	Stimulate_emotion		
8d		v#01784021	marvel-31.3	Experiencer_focused_emotion		17
9a	λανθάνω (lanthánō)	v#02078305 v#01419809 |	avoid-52	Elusive_goal	55	17
9b		v#00611721 v#00610654 v#00616520 v#00614532	not found	Abandonment| Remembering_experience| Remembering_information | Remembering_to_do	31	
10a	πυνθάνομαι (punthánomai)	v#00899241| v#00592510| v#00600349	inquire-37.1.2| comprehend-87-2| learn-14-2-1	Questioning| Coming_to_believe| Hearsay	69	
10b		v#00899241	inquire-37.1.2	Questioning		38

Table 4: Semantic annotation and extracted occurrences

4 Results

After extracting the data, we manually checked them in order to see whether the meanings we selected actually matched the occurrences of each specific query: this turned out to be most frequently correct. In Table 5 we calculated Precision. Queries for which Precision is not relevant are those that turned out not to yield results after checking annotation errors. We could not calculate Recall, as the corpus we used was not semantically annotated.

Several interesting observations emerge from the data. In the first place, only two queries (*9a* and *10a*) extracted occurrences that did not show the expected meaning, as Precision for query *2d* is <1 due to an annotation error. Query *9a* concerned the verb *lanthánō* 'escape notice'. This verb instantiates the (anti)causative alternation through voice only exceptionally, the MidPass and in HG the Pass meaning 'forget'. Hence, query *9a* did not yield the expected results in a single occurrence (*Il.* 15.60), in which the active has the causative meaning 'make forget', which was not predicted. With query *10a* we extracted forms of *punthánomai* in HG and found four occurrences which did not match the expected meaning but rather expressed the meaning 'perceive (a sound)'.

With the verb *mimnḗskō* the Act/MidPass distinction encodes the anticausative alternation. Interestingly, no Act forms occur in ClG, but Pass forms encode the passive diathesis for the meaning 'remind'. Hence the same semantic annotation applies to the verb in both corpora for the Act (HG) and the Pass (ClG), in line with the preliminary choice we made. It needs to be remarked that the number of occurrences extracted may be higher than the actual number of occurrences of verbs: in the case of query *1a*, for example, this is because *mimnḗskō* (Act) is a three-place verb featuring the NomAccGen construction, but in some occurrences only the second or the third argument are overtly realized, without a change of meaning. For this reason, our query did not specify the case of the co-occurring argument(s) and how many arguments we were looking for; consequently in passages in which both arguments were overtly realized we extracted the same occurrence twice (e.g. *mēdé me toútōn mímnēsk'* 'do not remind me (*me* Acc) of these things (*toútōn* Gen)', *Od.* 14.168-169, yielded two results). Similarly, with *kholóō* the Act/MidPass distinction also encodes the anticausative alternation. This verb only occurs in HG, most frequently in non-Act voice meaning 'be/become angry'. This includes Pass forms, as remarked in sec. 3.1 and 3.2 (see further below). It can occur with a dative second argument that indicates the stimulus, with a genitive indicating the reason or without a second argument, with no semantic differences, hence we performed a single query (7b). Act forms with causative meaning occur four times (7a).

For the verb *rhḗgnumi* we performed two separate queries for active forms depending on whether the second argument was realized (*2a*) or not (*2d*). In this second case, we expected to find cases of P-lability, which are known from post-HG; however, all passages featured referential null objects (NO) easily recoverable from the preceding context. Due to the extent to which referential NOs occur under different syntactic and discourse conditions [19], this type of query does not seem to be suitable to automatically retrieve labile verb forms. In our sample, this verb only occurs in HG; however, according to reference works it also occurs in Attic prose writers, albeit infrequently, due to frequent occurrence of prefixed verbs such as *diarrḗgnumi*, *ekrḗgnumi*, *katarrḗgnumi*, all featuring similar meanings. We also performed two different queries for *oîda* depending on whether the second argument was overtly realized or not (*4a* and *4b* vs *4c*). In this case, the motivation was that we wanted to see whether the verb could have some discourse function when occurring without a direct object. A survey of the occurrences shows that this is not the case, and that occurrences without an overt direct object may either feature referential NOs or subordinate clauses. In HG, occurrences of NOs may show the semantics associated to both query *4a* and *4b* (the latter query did not yield any result in ClG): this indicates that more queries should be added to separate occurrences of NOs from occurrences with subordinate clauses, as the latter feature the semantics associated to query *4a*. This query is the only one which, in spite of the unbalance of the two sub-corpora, yielded a higher number of results for ClG. The reason might lie in the expansion of subordination from HG to ClG [4] or in differences between textual genres.

As anticipated in Sec. 3.2, we corrected the sense we had selected for Pass forms of *phrázō* and *kholóō* based on the results of the query. Indeed, as we remarked in Sec. 3.1, Pass forms frequently have middle meaning in HG. Remarkably, Pass forms of *mimnḗskō* that only appear in post-HG actually encode the passive diathesis, in line with known developments of the voice system, whereby the passive diathesis is increasingly connected with Pass voice (and partly with MidPass) after HG [21]. This suggests that

the annotation should be better differentiated based on language stages and known facts about the diachrony of individual verbs.

Query	Extracted results	Results with expected meaning	Precision
1a	6	6	1
1b	99	99	1
1c	3	3	1
2a	24	24	1
2b	0	0	not relevant
2c	16	16	1
2d	5	4	0,800
3a	22	22	1
3b	83	83	1
3c	0	0	not relevant
4a	149	149	1
4b	22	22	1
4c	116	116	1
5a	3	3	1
5b	14	14	1
6a	18	18	1
6b	3	3	1
6c	3	3	1
6d	2	2	1
6e	0	0	not relevant
7a	4	4	1
7b	65	65	1
8a	15	15	1
8b	17	17	1
8c	17	17	1
9a	55	54	0,981
9b	31	30	1
10a	69	65	0,942
10b	44	44	1

Table 5: Precision index

With *mnáomai*, *orégō* and *phobéō* the queries yielded results that reflected the semantic annotation we provided in Table 4. In particular, *mnáomai* never occurs in Attic, but, outside poetry, is limited to Ionic and occurs in Herodotus (5th century BCE). It features the NomAcc construction and the meaning 'sue for, solicit', which results from semantic extension of the meaning 'woo' (query *5b*) and is compatible with the VN class compel-59.1 that we selected for this query (it does not contain this meaning, but it contains the related meaning 'persuade'). In the case of *orégō* it must be stressed that the queries we performed were more fine-grained than for other verbs, being based not only on morphosyntactic features but also on language stages. This is also the case for *phobéō*, a verb that underwent a semantic shift after HG from 'flee/be put on flight' to 'frighten/fear'. Notably, while both Act and MidPass are equally frequent in HG (*8a*, *8b*), in our sample of ClG only MidPass occurs (*8d*).

5 Conclusion

The manual check of the data extracted through the queries we designed based on the morphosyntactic behavior of a number of Ancient Greek verbs shows that the morphosyntactic features chosen were to a high extent reliable predictors of the associated meanings, as also shown by the Precision values. In future work some of the queries must be refined, e.g. specifying the form of subordinators for verbs likely to take complement clauses such as *oîda* or extended with more fine-grained distinctions based on known developments of AG morphosyntax. Furthermore, we plan to implement the rule-based algorithm described in [23] to automatically add role-dependency links to event participants i.e. tags that indicate how a frame element is expressed syntactically with respect to the node that corresponds to the frame trigger.

Finally, we plan to connect resources by exploiting Linked Open Data principles. The Ancient Greek WordNet identifies each lemma through a dedicated URI; as a next step, we plan to apply the same URIs to the corresponding lemmas in treebank, so that each lemma is uniquely identifiable and findable.

Acknowledgements

We wish to express our gratitude toward the anonymous reviewers, who greatly contributed to improving the quality of this paper. This chapter results from joint work of the authors. For Italian academic purposes, Erica Biagetti is responsible for Section 2, Chiara Zanchi for Section 3.1, Luca Brigada Villa for Section 3.2, and Silvia Luraghi for Section 4. Sections 1 and 5 were written jointly by the authors.

References

[1] Allan Rutger J. The middle voice in Ancient Greek: a study in polysemy. — Amsterdam: JC Gieben, 2003.
[2] Biagetti Erica, Zanchi Chiara, Short William M. Toward the creation of WordNets for ancient Indo-European languages // Proceedings of the 11th Global Wordnet Conference. — Pretoria, South Africa, 2021. — P. 258–266.
[3] Brigada Villa Luca. UDeasy: a Tool for Querying Treebanks in CoNLL-U Format // Proceedings of the Workshop on Challenges in the Management of Large Corpora (CMLC-10). — Marseille, France, 2022. — P. 16–19.
[4] Cristofaro Sonia. Aspetti sintattici e semantici delle frasi completive in greco antico. — Firenze: La Nuova Italia Ed., 1996.
[5] Fellbaum Christiane (ed.). WordNet: An electronic lexical database. — Cambridge, MA: MIT Press, 1998.
[6] Fellbaum Christiane. WordNet, The Encyclopedia of Applied Linguistics. — Hoboken, NJ: Wiley Online Library, 2012. https://doi.org/10.1002/9781405198431.wbeal1285
[7] Fillmore Charles J. Frame semantics and the nature of language // Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech. — New York, NY, 1976. — P. 20-32.
[8] Fillmore Charles J. Scenes-and-frames semantics, Linguistic Structures Processing. — Amsterdam: North Holland Publishing, 1977a.
[9] Fillmore Charles J. The need for a frame semantics in linguistics, Statistical Methods in Linguistics. — Stockholm: Scriptor, 1977b.
[10] Fillmore Charles J. Frame semantics, Linguistics in the morning calm. — Seoul: Hanshin, 1982. — P. 111-37.
[11] Fillmore Charles J. Frames and the semantics of understanding, Quaderni di Semantica — Bologna, Italy: Il Mulino, 1985. — Vol. 6, P. 222-254.

[12] Fillmore Charles J., Baker Collin F. Frame Semantics for Text Understanding // Proceedings of WordNet and Other Lexical Resources Workshop. — Pittsburgh, PA, 2001.
[13] Haspelmath Martin. More on the typology of inchoative/causative verb alternations, Causatives and transitivity. — Amsterdam: John Benjamins, 1993. — P. 87–120.
[14] Kipper-Schuler Karin. VerbNet: A broad-coverage, comprehensive verb lexicon (Ph. D. thesis). — Philadelphia, PA: University of Pennsylvania, 2006.
[15] Koeva Svetla, Leseva Svetlozara, Rizov Borislav, Tarpomanova Ekaterina, Dimitrova Tsvetana, Kukova Hristina, and Todorova Maria. 2011. Design and development of the Bulgarian Sense-Annotated Corpus. In Proceedings of the Third International Corpus Linguistics Conference (CILC), 7-9 April 2011, Valencia, Spain, pages 143–150. Universitat Politecnica de Valencia.
[16] Koeva Svetla, Leseva Svetlozara, Tarpomanova Ekaterina, Rizov Borislav, Dimitrova Tsvetana, and Kukova Hristina. 2010. Bulgarian Sense-Annotated Corpus – results and achievement. In Proceedings of the 7th International Conference of Formal Approaches to South Slavic and Balkan Languages (FASSBL-7), pages 41–49.
[17] Lakoff George, Johnson Mark. Conceptual metaphor in everyday language, The journal of Philosophy. — New York, NY: The journal of Philosophy, 1980. — Vol. 8(77), P. 453-486.
[18] Levin Beth. English verb classes and alternations. — Chicago: Chicago University Press, 1993.
[19] Luraghi Silvia. On the meaning of prepositions and cases. — Amsterdam: John Benjamins, 2003.
[20] Luraghi Silvia. Experiential verbs in Homeric Greek. — Leiden: Brill, 2020.
[21] Luraghi Silvia. Beyond affectedness – Partitive objects and degrees of agenthood in Ancient Greek, Linguistic Variation. — Amsterdam: John Benjamins, 2023. — Vol. 23(1), P. 95–123
[22] Luraghi Silvia, Mertyris Dionysios. 2021. Basic valency in diachrony: from Ancient to Modern Greek, Valency over time. — Berlin: De Gruyter, 2021. — P. 169-208.
[23] Minnema Gosse, Gemelli Sara, Zanchi Chiara, Caselli Tommaso, Nissim Malvina. SocioFillmore: A Tool for Discovering Perspectives // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. — Dublin, Ireland: Association for Computational Linguistics, 2022. — P. 240–250.
[24] Pustejovsky James, Moszkowicz Jessica. The qualitative spatial dynamics of motion in language, Spatial Cognition & Computation. — Milton Park, UK: Taylor & Francis, 2011 — Vol. 11(1), P. 15-44.
[25] Pustejovsky James. Dynamic event structure and habitat theory // Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013). — Pisa, Italy, 2013. — P. 1-10.
[26] Ruppenhofer Josef, Ellsworth Michael, Petruck Miriam R. L., Johnson Christopher R., Baker Collin F., Scheffczyk. FrameNet II: Extended Theory and Practice. — https://framenet.icsi.berkeley.edu/fndrupal/the_book, 2016.
[27] Stoyanova Ivelina, Leseva Svetlozara. Expanding the Conceptual Description of Verbs in WordNet with Semantic and Syntactic Information // Proceedings of the 12th Global WordNet Conference. — Donostia/San Sebastián, Spain, 2023.
[28] Tyler Andrea, Evans Vyvyan. The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition. — Cambridge, UK: Cambridge University Press, 2003.
[29] Zanchi Chiara, Luraghi Silvia, Biagetti Erica 2021. Linking the Ancient Greek WordNet to the Homeric Dependency Lexicon. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialog 2021” [Komp’yuternaya Lingvistika i Intellektual’nye Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2021”]. — P. 729-737.