

# Toward the creation of WordNets for ancient Indo-European languages

**Erica Biagetti**

University of Pavia / Bergamo

Pavia / Bergamo, Italy

`erica.biagetti01@universitadipavia.it`

**Chiara Zanchi**

University of Pavia

Pavia, Italy

`chiara.zanchi01@unipv.it`

**William M. Short**

University of Exeter

Exeter, United Kingdom

`W.Short@exeter.ac.uk`

## Abstract

This paper presents the work in progress toward the creation of a family of WordNets for Sanskrit, Ancient Greek, and Latin. Building on previous attempts in the field, we elaborate these efforts bridging together WordNet relational semantics with theories of meaning from Cognitive Linguistics. We discuss some of the innovations we have introduced to the WordNet architecture, to better capture the polysemy of words, as well as Indo-European language family-specific features. We conclude the paper framing our work within the larger picture of resources available for ancient languages and showing that WordNet-backed search tools have the potential to re-define the kinds of questions that can be asked of ancient language corpora.

## 1 Introduction

This paper presents the work in progress toward the creation of a family of WordNets for ancient Indo-European (IE) languages, namely Sanskrit (Skt.), Ancient Greek (AG), and Latin (Lat.). This ongoing project is being jointly developed by an international team of scholars at the University of Exeter, the University of Pavia, the Center for Hellenic Studies at Harvard University, and the

Alpheios Project, spearheaded by William M. Short. The design, as well as the specific content, of these WordNets builds on several previous (but, as far as we know, now defunct) attempts in the field (for AG, Bizzoni et al., 2014, Boschetti, 2019; for Lat., Minozzi, 2009),<sup>1</sup> extending and elaborating this work in certain critical respects (in particular, by bringing the framework under theories of meaning from Cognitive Linguistics). Crucially, these WordNets share the same data organization and exploit of the same pool of sense designations (synsets), enabling comparison of linguistic – above all semantic – structures cross-linguistically through the use of a common set of definitional elements.

In this paper, we discuss some of the innovations we have introduced to the WordNet architecture, to better capture the polysemy of words (including their figurative metaphorical and metonymic uses) as well as IE language family-specific features. We finally frame our family of WordNets in the wider picture of linguistic resources available for ancient languages.

## 2 Representing meaning in ancient language WordNets

Like previous WordNets (Fellbaum, 1998), our ancient language WordNets are lexical databases in which meaning is stored in a relational way. WordNets comprise nodes for lemmas to which meanings are associated in the form of synsets,

---

<sup>1</sup> The previous Ancient Greek and Latin WordNets can be partly consulted here: [http://www.languagelibrary.eu/new\\_ewnui/](http://www.languagelibrary.eu/new_ewnui/). Boschetti (2019) references to an ongoing project on the Homeric lexicon, named Homeric Greek WordNet, which at the time,

when the paper was published, was being developed at ILC, CNR, Pisa. Currently, the website <https://cophilab.ilc.cnr.it/hgwnWeb/> requires a username and password to be accessed, and detailed information on the status of the project does not seem to be available online.

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.

Different lemmas can share one or more synset(s), which means that they are (partly) synonymous. Other semantic relations are typically tagged in WordNets, which mostly interconnect synsets associated with lemmas of the same part of speech: for example, the HYPONYMY-HYPERNYMY relation connects nouns to nouns (e.g. AG *ikhthûs* ‘fish’ and *zôion* ‘animal’), the ENTAILS relation connects verbs to verbs (e.g. AG *plêō* (ACT) ‘sail’ and *kinéomai* (M/P) ‘be in motion’), etc. (for similarities and differences between the traditional and our set of relations, see Section 3.4). Like in previous WordNets, our set of semantic relations fails to capture semantic solidarity due to belonging in the same Frame (Fillmore *et al.*, 2003) or semantic field (Fellbaum, 1998: 10 *tennis problem*). Thus, for example, no semantic relation links the AG words in (1):

- (1) *ikhthûs* ‘fish’, *thálassa* ‘sea’, *naûs* ‘ship’,  
*naútēs* ‘sailor’, *plêō* ‘sail’

However, *naútēs* is morphologically derived from *naûs*, which is annotated among lexical relations.

Like in other WordNets, lemmas can be assigned multiple synsets, which indicates polysemy. We have decided to frame our lexicographic work within a cognitive linguistic approach (e.g. Lakoff and Johnson, 1980; Tyler and Evans, 2003; on Cognitive Linguistics applied to the study of ancient languages, see Mocchiari and Short, 2019) and thus have embraced a principled view of polysemy. This entails (a) avoiding exaggerating the number of distinct senses associated to a lemma; (b) assuming that all senses of a lemma can be organized in a structured semantic network. Roughly, literal senses are detected based on their early attestation, concreteness, and predominance in the network (Tyler and Evans, 2003: 45-50), whereas non-literal senses are derived from literal ones through the cognitive processes of metaphor and metonymy. For example, in the Princeton WordNet, three senses are associated with the adjective *salty*, reported in (2):

- (2) a. containing or filled with salt;  
 b. one of the four basic taste sensations; like the taste of sea water;

c. engagingly stimulating or provocative.

The sense in (2)a is the basic one, as *salty* morphologically derives from the noun *salt*. The sense in (2)b can be derived from (2)a via a metonymic process: a word denoting a state is employed to denote the physical sensation that such state triggers. The metaphoric meaning in (2)c can be connected with (2)a via the metonymic sense in (2)b: *being salty* is as positively or negatively engaging for the palate as *being stimulating/provocative* is for the spirit. The difference between cognitive metonymy and metaphor is that, with the former, the senses associated with the polysemous word belong to the same conceptual domain, whereas with the latter two senses belonging to different conceptual domains are mapped to one another.

Crucially, in our WordNets, we are implementing this principled view of polysemy by asking annotators to avoid multiplying the number of synsets associated to lemmas and to tag only senses that clearly do not emerge from context. Moreover, our annotators are required to maximize the usage of the synsets deriving from the Princeton WordNet for English, in order to enhance the compatibility of our WordNets with existing ones and to establish a common base of sense definitions. Finally, while tagging senses of lemma entries, our annotators are asked to distinguish among synsets that correspond to literal, metonymic, and metaphoric meanings.

For example, 16 synsets are currently associated to the AG word for ‘salt’, *hâls*, and classified into three groups, viz. literal (4 synsets), metonymic (4 synsets) and metaphoric (8 synsets) senses, exemplified in (3)a, (3)b, and (3)c, respectively.

- (3) a. literal sense ‘salt’  
 n#05846273 | white crystalline form of especially sodium chloride used to season and preserve food  
 b. metonymic sense ‘body of salty water’  
 n#10771040 | water containing salts  
 c. metaphoric sense ‘wit’  
 n#05075890 | a message whose ingenuity or verbal skill or incongruity has the power to evoke laughter

As discussed above, the difference between metonymy and metaphor relies in being vs. not-being part of the same conceptual domain. Clearly, the senses of ‘salt’ and ‘body of salty water’ both pertain to the domain of SEA; by contrast, the senses of ‘salt’ and ‘wit’ belong to

two different domains (cf. the meanings of *salty*, remembered above in this section).

As we are dealing with corpus languages that enjoy centuries of attestation and a long tradition of studies, each of the identified literal, metonymic, and metaphoric synsets will be tagged for its periodization(s), literary genre(s), and optionally *loci*, i.e., exemplifying attestations referred to by author(s) and work(s). Thus, for example, the senses in (3)a-c are enriched with the following diachronic and stylistic metadata:

Sense	Period	Genre	Loci
(3)a	Archaic (8 <sup>th</sup> -6 <sup>th</sup> BCE)	poetry epic historiography narrative	Il.9.214, Od.11.123 Ar.Ach.835 Hdt.4.53
(3)b	Hellenistic (323-31 BCE)	-	Call.Fr.50
(3)c	Roman (31 BCE-290 CE)	philosophy treatise	Plut.2.685 Plut.2.854

Table 1: Diachronic and stylistic metadata associated with the sense of *háls* in (3)a-c

We expect this information to be extremely useful for philologists, lexical typologists, and historical linguists interested in semantic change. On the one hand, as our WordNets will also include etymological information (Section 3.1), users will be able to investigate whether Skt., AG, and Lat. cognate words lexicalize comparable arrays of concepts (see Section 4). On the other hand, users will be able to track whether and how word meanings change over time and vary across literary genres and authors.

### 3 Family-specific attributes and relations

#### 3.1 Annotation of lemmas

As anticipated in Section 2, etymological information completes the diachronic picture. Etymological information for each database entry is hierarchically structured and consists of:

- ETYMOLOGY proper: e.g. PIE *\*pleu-* ‘float’ for AG *pléō* ‘sail’ and Skt. *plu-* ‘float, swim’;
- ETYMON: a discrete form in the history of a word’s etymological development (e.g. Lat. *pulmo* < AG *pleumon* < AG *pneumon* ‘lung’ < PIE *\*pleu-* ‘float’);

- MORPHEME: a discrete element within the etymon (e.g. *\*-ti-* in Skt. *plu-ti-* ‘flood’, AG *plú-si-s* ‘washing’).

Each of the three levels of etymological information is stored as a separate entry in the database, which allows lemmas to be linked via their etymological constituents at many different levels (root, stem, morphemes, etc.).

For AG, a dedicated field gives information on dialectal variants (e.g. Attic *plóus* ‘sailing, voyage’, Ionic *plóos* ‘id.’).

Unlike in other WordNets, each lemma is provided with morphological information in our databases. Beside specification of the part of speech, morphological information is stored in three fields:

- MORPHO: we employ a modified version of the tagging schema developed for the *Perseus Digital Library*<sup>2</sup> for encoding morphological properties of tokens. The schema consists of a ten-place character string, where each place corresponds to a grammatical category (e.g. AG *limén* ‘harbor’ n-s---mn3n).
- MORPHOLOGY: this field consists of a subfield PRINCIPAL PARTS, where relevant parts of the paradigm are listed, and a subfield PROSODY providing vowel length when relevant. For instance, AG *háls* ‘salt’ has a principal part *halós*, which corresponds to its genitive form. Prosody, instead, is provided in cases such as Lat. *occīdo* ‘to strike down’ as distinct from *occīdo* ‘to fall; die’.
- FORM TOKENS: it consists of a token with its morphological tag, specifying whether this is ‘irregular’ and/or ‘alternative’. Since irregular forms may be case- or number-specific, this field constitutes an exception to the exclusion of inflected word forms from our WordNet. One instance is again represented by AG *háls* ‘salt’ with its two dative plural forms *halsí* and *hálasi*: the latter, being built on a different stem, is annotated as an alternative form (Form n-p---md3-, Token *hálasi*, Alternative).

Table 2 displays the annotation associated to the AG lemma *háls* ‘salt’:

Field	Subfield	Value
Etymology	–	PIE <i>*séh2l-</i> ‘salt’
Lemma	–	<i>háls</i>

<sup>2</sup> <http://www.perseus.tufts.edu>.

POS	_	Noun
Morpho	_	n-s---mn3-
Morphology	Prin. Parts	<i>halós</i>
	Prosody	_
Form Tokens	Form	n-p---md3-
	Token	<i>hálasi</i>
	Alternative	✓

Table 2: Lemma annotation for AG *háls*.

### 3.2 Lexical databases

Previous WordNets comprise lemmas belonging to open class parts of speech only, that is, nouns (N), adjectives (A), verbs (V), and adverbs (Adv). In our WordNets, a new part of speech was added, that of prepositions (P), for a number of reasons. First, because of the importance these elements hold in the grammar systems of IE languages. Following the literature on AG (e.g. Chantraine, 1953), we take *preposition* as a catch-all term for a class of uninflected morphemes that feature high semantic and syntactic flexibility in IE, functioning either as local adverbs, adpositions or preverbs (Reinöhl and Casaretto, 2018; Zanchi, 2019: Ch. 3). Second, prepositions are originally associated with concrete meanings, which constitute the starting point for developing more abstract meanings thanks to the cognitive mechanisms of metonymy and metaphor (Section 2). Therefore, they are of particular importance in Cognitive Linguistics, as they constitute a privileged viewpoint for studying how discrete senses associated to a lemma organize in a structured network. Finally, including prepositions in WordNet allows us to study the semantic interaction between simplex and compound verbs. A compound verb such as AG *ap-eípon* ‘deny’ illustrates the points above: in combination with the communication verb *eípon* ‘say’, the preverb *apó-* ‘away’ gains an abstract meaning and expresses refusal, making the meaning of the compound verb non-compositional (Zanchi 2019: 67).

Sometimes, prepositions occur in *multi-word units* (Fellbaum, 2015), such as Lat. *sub divo* ‘in the open air’. In our WordNets, the lemma list will include such multi-word units that show a word-like distribution and feature some degree of semantic idiomaticity and of structural fixedness (on multi-word expressions, see also Masini, 2019 with references). Other examples are Lat. *res publica* ‘state, republic’ and AG *thalássia érga* ‘navigation’.

### 3.3 Lexical relations

In WordNet, lexical relations include both morphological relations, such as derivation and composition, and the semantic relation of antonymy. The reason for including antonymy among lexical relations is that, in a word association test, two antonyms are always given as the most common response one of the other (Deese, 1964; 1965); therefore, *heavy/light* are antonyms, but *weighty/light* are not, and antonymy is defined as a semantic relation between words rather than synsets (Miller, 1998: 48). However, since we cannot rely on speakers’ judgments, we have decided to split the antonymy relation into a lexical (i.e. morphological) and a semantic relation. Morphological antonyms are lemma pairs, where one of the antonyms is derived from the other through the privative prefix *a-*: Skt. [*a-mitra-* ‘non-friend, enemy’] IS PRIVATIVE OF [*mitra-* ‘friend’]. Note that lexical antonymy is asymmetric: if we take the base as a starting point, we get [*mitra-* ‘friend’] HAS PRIVATIVE [*a-mitra-* ‘non-friend, enemy’].

In order to represent the rich derivational morphology of IE languages, we have decided to extend the set of lexical relations as follows:

- Derivation: asymmetric relation holding between a base and a word derived from it either by conversion (Skt. *nāga-* A ‘serpentine’ > *nāga-* N ‘a kind of serpent’) or by affixation: AG [*makró-tēs* ‘length’] IS DERIVED FROM [*makrós* ‘long’]. The inverse relation is IS RELATED TO: [*makrós*] IS RELATED TO [*makró-tēs*].
- Parasynthesis: asymmetric relation holding between a base and a word derived from it by simultaneous conversion and affixation: AG [*ánoos* A ‘without understanding’] IS PARASYNTHETIC OF [*nóos* N ‘mind’]. The inverse relation is HAS PARASYNTHETON: [*nóos*] HAS PARASYNTHETON [*ánoos*].
- Composition: asymmetric, many-to-many relation holding between a compound word and its constituents: Skt. [*rāja-putra-* ‘a king’s son, prince’] IS COMPOSED OF [*rāja-* ‘king’], [*rāja-putra-*] IS COMPOSED OF [*putra-* ‘son’]. The inverse relation is COMPOSES: [*rāja-*] COMPOSES [*rāja-putra-*].
- Inclusion: asymmetric many-to-many relation holding between a multi-word unit and its parts: AG [*thalássia érga* ‘navigation’] INCLUDES [*thalássios* ‘related to the sea’], [*thalássia érga*] INCLUDES [*érgon* ‘work’].

The inverse relation is IS INCLUDED IN: [*thalássios*] IS INCLUDED IN [*thalássia érga*], [*érgon*] IS INCLUDED IN [*thalássia érga*];

- e. Participle: asymmetric relation holding between a participle and its base verb: Skt. [*sát-* ‘true’] IS PARTICIPLE OF [*as-* ‘be’].

Table 3 summarizes newly added lexical relations:

Rel.	Label	Inverse
Anton.	IS PRIVATIVE OF	HAS PRIVATIVE
Der.	IS DERIVED FROM	IS RELATED TO
Paras.	IS PARASYNTHETIC OF	HAS PARASYNTHON
Comp.	IS COMPOSED OF	COMPOSES
Incl.	INCLUDES	IS INCLUDED IN
Part.	IS PARTICIPLE OF	HAS PARTICIPLE

Table 3: Family-specific Lexical Relations.

### 3.4 Semantic relations

Semantic relations constitute the core of WordNet architecture. In order to ensure compatibility of our WordNets with the existing ones, we tried to stick to the established set as closely as possible. However, some differences must be mentioned:

- a. Semantic antonymy: contrary to morphological antonymy (Section 3.3), and to antonymy in other WordNets, semantic antonymy holds between synsets. Thus, semantic antonymy does not link e.g. AG *kalós* ‘good’ and *kakós* ‘bad’ themselves, but rather the synsets to which they belong; contrary to morphological antonymy, semantic antonymy is a symmetric relation: {n#01963712 “of moral excellence”} HAS ANTONYM {n#01078381 “having undesirable or negative qualities”}.
- b. Similar to / Also see: in other WordNets, the relation IS SIMILAR TO links satellite synsets to one of the antonyms in a cluster of adjectives; ALSO SEE, instead, links the half cluster to another half cluster related to it. Since semantic antonymy links synsets in our WordNets, we avoid using both relations and employ IS NEAREST TO as a catch-all relation for similar synsets: {n#01893072 “a young pig”} IS NEAREST TO {n#01892895 “domestic swine”} for AG *khoîros* and *sûs*.

- c. Verbal sense group: symmetric relation linking verbs related by aspectual, voice- or valency-related properties: {v#00399347 “become conscious of”} VERBAL SENSE GROUP {v#00401762 “possess knowledge or information about”} for AG *gignôskō* (PRS) ‘perceive, know’ and *oîda* (PF) ‘know’.<sup>3</sup>
- d. Qualifies event as: asymmetric relation holding between an adverb and an adjective: {r#00162139 “for an extended time or at a distant time:”} QUALIFIES EVENT AS {a#01380813 “being or indicating a relatively great or greater than average duration or passage of time or a duration as specified”} for AG *makrán* ‘at length’ and *makrós* ‘long’; the inverse relation is QUALIFIES ENTITY AS.

Table 4 summarizes family-specific semantic relations:

Rel.	Label	Inverse
Anton.	HAS ANTONYM	HAS ANTONYM
Near.	IS NEAREST TO	IS NEAREST TO
Verb. Sense Group	VERBAL SENSE GROUP	VERBAL SENSE GROUP
Qual. event as	QUALIFIES EVENT AS	QUALIFIES ENTITY AS

Table 4: Family-specific Semantic Relations.

## 4 Integrating ancient language WordNets with existing resources

The Skt., AG, and Lat. WordNets have been designed to be fully interoperable, as well as integrated into the larger ecosystem of digital lexical and textual resources for ancient languages. What is more, they make available a standard API (application programming interface) permitting any user, or computer application, to programmatically access their lexical and semantic content in a consistent manner, regardless of language (or simultaneously for all languages). For example, it would be trivial to discover the words in Skt., AG, and Lat. that correspond to the meaning ‘a short stabbing weapon’ (i.e., a dagger) – represented by synset n#02542418 – simply by querying the endpoint */api/synsets/n/02542418/lemmas/* at the address of the relevant WordNet. More sophisticated queries could take advantage of the rich semantic, morphological, etymological, and figurative data

<sup>3</sup> In IE studies, *oîda* (PF) ‘know’ is said to be a defective form, which thus enters paradigms of other verbal roots.

that, while characterizing specific structures of a given language, are encoded through a set of language-independent (as it were, ‘etic’) elements. In fact, because they share certain linguistic structures (including etymological primitives) at a fundamental level, the Sanskrit, Greek and Latin WordNets represents the first systematic attempt in classical language lexicography to deliver a basis for comparative semantic research (Section 2).

Beyond interoperability, the architecture of the Skt., AG and Lat. WordNets aims to facilitate their integration with other lexical and textual resources. The Lat. WordNet, for instance, is now being aligned with the ERC-funded *Linking Latin* project (<https://lila-erc.eu>), which aims to standardize different resources around a single set of lemma-based URIs. This will enable users to easily tie together information available from disparate lexical and textual resources by guaranteeing the correct identification of lemmas (e.g., in the case of ambiguous word forms). Similarly, the Sanskrit WordNet is tightly integrated with the *Digital Corpus of Sanskrit* (<http://www.sanskritlinguistics.org/dcs/index.php>), which will allow users of this corpus to query semantic data utilizing pre-existing identification tags. The morphological encoding schema is compatible with the quasi-standard system used in most annotated corpora of Greek and Latin, adding two further fields to provide greater specificity in lexical categorization (see Section 3.1). This is meant to enable scholars to inject semantic information, along with syntactic information, into natural language processing pipelines for the first time. At the same time, this means that other NLP tools already available for the ancient languages can automatically and immediately take advantage of the WordNet data to improve their functionality, accuracy, and scope.

The Sanskrit, Greek and Latin WordNets are, finally, designed to work hand-in-hand with electronic corpora of semantically annotated texts – what we call “sembanks” on the model of syntactic “treebanks”. The creation of the WordNets, on one hand, and of sembanks, on the other, in fact constitute two prongs of a single effort to bring research on ancient language semantics under computational approaches. For this reason, efforts are currently underway to produce a robust but flexible XML schema, following standards established by the Text Encoding Initiative (<http://tei-c.org>) for use in annotating texts with WordNet constructs (above

all, synsets) in order to capture the senses of words or larger textual units, as they occur in specific contexts. This schema incorporates the concept of a *semtagm* as a semantically meaningful unit consisting of one or more tokens and that of a *reading*, representing one discrete possible interpretation of the given *semtagm*. So, for example, the mark-up of the first sentence of the preface of Cato’s *De Agri Cultura* would consist of a sequence of *semtagm* elements, whose values correspond straightforwardly to definite synsets: *est* = v#01775163, ‘have an existence, be extant’; *interdum* = r#00020741, ‘on certain occasions’; *praestare* = v#01246259, ‘value more highly’; *mercaturis* = n#00707408, ‘the commercial exchange of goods and services’ and so on.

Because the ancient language WordNets also include information about the figurative senses of words and the conceptual structures that underpin these senses (Section 2), it is further possible to annotate the figurative senses of words. For example, in the following annotation of Ovid’s *Metamorphoses* 13.11, the synset glossed “a hostile meeting of opposing military forces in the course of a war” (n#00610417) has been encoded as the contextual sense of *Mars*, which is indicated as a metonymical usage of the god’s name and includes a designation of the conceptual metonymy that motivates this interpretation:

```
<semtagm n="73"
urn="latinLit:phi0959.phi006.perseus-lat1"
cite="13.11:11:7">
  <token n="1" form="Marte" lemma="Mars"
    uri="50193" morpho="n-s---mb3-">
    <reading n="1" synset="n#00610417"
      figure="#" mapping="247" />
  </token>
</semtagm>
```

This annotation schema is designed, moreover, to help capture the polysemy that tends to characterize word usage in literary contexts – due to textual problems arising from the process of transmission, intentional or unintentional lexical ambiguities, or genuine disagreements of interpretation in critical analysis – by permitting annotators to tag lexical or phrasal tokens with multiple sense designations. Thus, for example, the famous ambiguity of Catullus’s *puđicitiam matris indicet ore*, where *os* can be interpreted either as ‘face’ or (more specifically) as ‘mouth’ and again by metonymy, “speech”, is represented by two *reading* elements within a single *semtagm*, to simultaneously encode synsets n#03683012, ‘outward or visible aspect of a person or thing’

and n#05319899, ‘communication by word of mouth’ as this word’s possible interpretations.

When combined with a next-generation corpus search tool like Cylleneus,<sup>4</sup> WordNet-based semantically annotated texts will enable users to query ancient texts on the basis not only of their morphological and syntactic properties, but also of their semantic properties – that is, on the basis of the meanings of words as well as of the kinds of grammatical constructions in which they appear. For example, someone interested in ancient “courage” would easily be able to find occurrences of this concept in Sanskrit, Greek, or Latin literature, simply by searching for a specific synset or some higher-order semantic category (semfield) – without needing to conduct separate searches for each lemma. This would make identifying semantic intertextualities, for instance – the ways in which one text creates new meanings by reworking the themes and ideas (not merely the verbal elements) of other texts – almost trivial. More generally, whereas current corpus search methodologies require painstaking and time-consuming “brute force” searching in order to identify patterns of usage, by abstracting away from the lexicon and thus permitting efficient queries of whole semantic fields (in conjunction with morphosyntactic queries), WordNet-backed search tools have the potential to redefine the kinds of questions that can be asked of ancient language corpora.

## Acknowledgements

The innovations we introduced in the architecture of our WordNets partly implement suggestions made by students at the University of Pavia, who volunteered to annotate the data. We are extremely grateful to them for their time and insightful comments and to Silvia Luraghi for leading the Pavia research unit in this project.

The present paper results from intense collaboration of the three authors. For academic purposes, Erica Biagetti is responsible of Section 3, William M. Short of Section 4, and Chiara Zanchi of Sections 1 and 2.

## References

Federico, Boschetti. 2019. Semantic Analysis and Thematic Annotation. In: Monica Berti (ed.), *Digital Classical Philology*. De Gruyter, Berlin, DEU.

Pierre Chantraine. 1953. *Grammaire homérique*, Tome 2: Syntaxe. Klincksieck, Paris, FR.

Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini and Gregory R. Crane. 2014. The Making of Ancient Greek WordNet. In: Nicoletta Calzolari et al. (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC, vol. 2014)*, Reykjavik, Iceland, may 2014, 1140-1147. Accessed online at <https://www.aclweb.org/anthology/L14-1054/>.

James Deese. 1964. The associative structure of some common English adjectives. *Journal of Verbal Learning and Verbal Behavior* 3(5):347-357.

James Deese. 1965. *The Structure of Associations in Language and Thought*. John Hopkins Press, Baltimore, MD.

Christiane Fellbaum (ed.). 1998. *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA.

Christiane Fellbaum. 2015. The treatment of multi-word units in lexicography. In: Philip Durkin (ed.), *The Oxford handbook of lexicography*. OUP, Oxford, UK, 411-424.

Charles. J. Fillmore, Christopher Johnson, and Miriam R. L. Petruck. 2003. Background to Framenet. *International Journal of Lexicography*, 16(3):235-250.

Francesca Masini. 2019. *Multi-Word Expressions and Morphology*. Accessed online at <https://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-611>.

Katherine J. Miller. 1998. Modifiers in WordNet. In: Christiane Fellbaum (ed.), *WordNet: An electronic lexical database*. MIT Press, Cambridge, MA, 47-67.

Stefano Minozzi. 2009. The Latin WordNet Project. In: Peter Anreiter and Manfred Kienpointner (eds.), *Latin Linguistics Today. Akten des 15. Internationalen Kolloquiums zur Lateinischen Linguistik*, Innsbrucker Beiträge zur Sprachwissenschaft 137, Innsbruck, AUT, 707-716.

Egle Mocciano and William M. Short. 2019. *Toward a Cognitive Classical Linguistics. The Embodied Basis of Constructions in Greek and Latin*. De Gruyter, Berlin, DEU.

Uta Reinöhl and Antje Casaretto. 2018. When grammaticalization does not occur—Prosody-syntax mismatches in Indo-Aryan. *Diachronica* 35(2):238-276.

---

<sup>4</sup> <http://github.com/cylleneus/cylleneus>.

Andrea Tyler and Vyvyan Evans. 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meanings and Cognition*. CUP, Cambridge, UK.

Chiara Zanchi. 2019. *Multiple preverbs in ancient Indo-European languages*. Narr, Tübingen.



