

GUIDELINES FOR A LINGUISTIC ANNOTATION OF PREVERBED VERBS OF MOTION

Andrea Farina

Department of Digital Humanities, King's College London

Table of contents

Table of contents.....	1
1. Introduction.....	3
2. Theoretical premise.....	3
2.1 Motion events.....	3
2.1.1 Figure and Ground.....	3
2.1.2 Spatial relations	4
2.2 Motion verbs and verb classes.....	4
2.3 Preverbs	5
2.4 Actionality	6
2.4.1 Actional shifts.....	6
2.5 WordNet	7
2.5.1 Issues and weaknesses of WordNet.....	9
2.6 World Historical Gazetteer.....	11
2.6.1 Why are places and texts related?	12
3. Annotation overview.....	13
4. The annotation	15
4.1 Annotation layers	15
4.2 Annotation tagsets	19
5. Practical recommendations about the annotation	22
5.1 ACTIONALITY	22
5.2 EXPRESSED BY	23
5.3 FIGURE SYNSET.....	24
5.4 GROUND SYNSET	28
5.5 LITERAL MEANING.....	28
5.6 MORPHOLOGICAL FEATURES.....	30
5.7 PARTICIPANTS.....	31
5.8 PLACE	31
5.9 PREVERB	33
5.10 SEMPREV	33
5.11 SENTENCE.....	35

5.12 SPATIALITY	36
5.13 SYNSET	37
5.14 VERB CLASS	39
6. Other possible applications of this annotation scheme	42
6.1 Preverbed verbs of motion in (other) ancient or modern languages	42
6.2 Preverbed verbs not expressing motion in ancient or modern languages	47
6.3 Non-preverbed verbs (of motion or not) in ancient or modern languages	50
7. Conclusions and future work	50
8. References	50

1. Introduction

These guidelines contain the annotation scheme for an in-depth analysis of preverbed motion verbs. This has been designed as part of my PhD project, aiming to a quantitative cross-linguistic analysis on preverbed motion verbs in Ancient Greek and Latin.

The document is structured as follows. Section 2 briefly discusses some theoretical background necessary for the annotation: I describe motion events, motion verbs, preverbs, actionality, the WordNet architecture, and the World Historical Gazetteer. Section 3 includes a schematization of the annotation pipeline (Figure 5), describing the annotation an annotator is expected to perform. Each layer mentioned in Figure 5, as well as the tagsets created for this annotation scheme, are commented on in detail in Section 4. Section 5 includes relevant practical recommendations for the annotators, aiming to address the most common problems that they may face while performing the task. Advice is given by considering some instances of motion verbs and how they should be annotated. To provide comparable examples, some verb occurrences mentioned in Section 5 do not display a preverb when the focus of the subsection is not prevervation. Finally, section 6 contains other possible applications of this annotation scheme, providing sentences annotated in other languages as useful examples, and Section 7 includes conclusions and future work. References for this work follow in Section 8.

2. Theoretical premise

In this section, I briefly summarize the theoretical framework used for this annotation and describe the parameters in the annotation layers (Section 4.1) and tagsets (Section 4.2).

2.1 Motion events

2.1.1 Figure and Ground

Motion events have been extensively studied in literature, most notably by Talmy (1975; 1983; 1985; 2000). Talmy identifies two participants in a motion event, i.e. the Figure and the Ground.

- Figure: the Figure «is a moving or conceptually movable point whose path or site is conceived as a variable the particular value of which is the salient issue» (Talmy 1975: 419). In other words, the Figure is the entity that either concretely performs a motion in space (*the book* in *The book fell off the shelf*) or does not perform any motion, even though it could potentially perform one (*the book* in *The book lies on the shelf*). In the former case, the Figure is found with motion verbs, while in the latter case it is found with stative verbs.
- Ground: the Ground is a stationary entity which is used as a reference point for the motion/location of the Figure (*the shelf* in the two sentences above). In other words, the Ground is the entity used to characterize the Figure's motion or location. It could be the background of a moving entity (*the shelf* in *The book fell off the shelf*) or a generic reference point of a motion/stative event (*Harry* in *John is near Harry*).

In general, the Figure is the participant possessing prominent attention in the sentence, whereas the Ground is in the background and its only role in the motion event is to locate the Figure.

To clarify the difference between the two participants, Table 1 (from Talmy 2000: 183) specifies some properties of the Figure and the Ground.

Figure	Ground
--------	--------

• has unknown spatial (or temporal) properties to be determined	• acts as a reference entity, having known properties that can characterize the primary object's unknowns
• more movable	• more permanently located
• smaller	• larger
• geometrically simpler (often pointlike) in its treatment	• geometrically more complex in its treatment
• more recently on the scene / in awareness	• more familiar / expected
• of greater concern / relevance	• of lesser concern / relevance
• less immediately perceivable	• more immediately perceivable
• more salient, once perceived	• more backgrounded once Figure is perceived
• more dependent	• more independent

Table 1. Characteristics associated with the Figure and the Ground (Talmy 2000: 183).

2.1.2 Spatial relations

Lakoff (1987) identifies a tripartite scheme of spatial relations for motion events.

- Source: the point where the motion starts.
- Goal: the point where the motion ends.
- Path: the series of continuous locations through which an entity moves to get from the Source to the Goal.

These spatial relations are represented in Figure 1, with reference to the sentence *Tomas went from the station to the restaurant through Oxford Circus*.

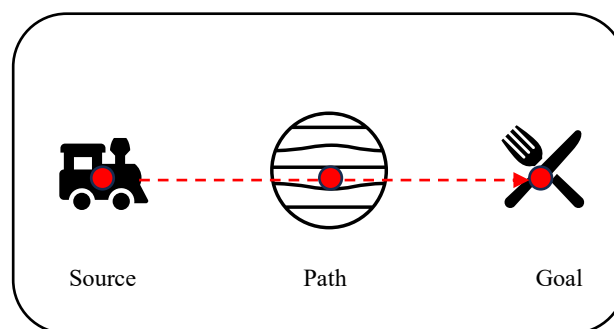


Figure 1. Source-Path-Goal schematization of the sentence *Tomas went from the station to the restaurant through Oxford Circus*. Tomas is the red dot, and the dotted line represents his motion from the station (Source) to the restaurant (Goal), passing through Oxford Circus (Path).

A fourth spatial relation (Location) occurs when the motion does not lead to a change of place (e.g. *The ant moves inside the box*) or the place in which a motion takes place is specified, being located in the background (e.g. *In London, Tomas went from the station to the restaurant through Oxford Circus*).

2.2 Motion verbs and verb classes

For the typology of motion verbs, I refer to VerbNet's¹ classification, which is based on the classification of English verbs by Levin (1993). In Levin's classification, motion verbs are

¹ <https://uvi.colorado.edu>.

assigned class 51. For the purposes of this document, here I only comment on the verb classes that are crucial for this annotation, i.e. motion verbs.

- ESCAPE-51.1: verbs denoting generic motion (e.g. *arrive, come, go*). No specification of manner of motion or implicit means of transport is given. Deixis is not a parameter in VerbNet's classification, so *come* and *go* are grouped together.
- LEAVE-51.2: verbs inherently denoting motion away from a place (e.g. *depart, leave, exit*).
- ROLL-51.3.1²: verbs implying rotation on an axis or non-voluntary motion (e.g. *drift, rotate, spin*).
- RUN-51.3.2: verbs inherently denoting faster (or slower) motion performed on foot (e.g. *jump, run, sprint*).
- VEHICLE-51.4.1: verbs denoting motion with a vehicle and whose lemma is the name of the vehicle itself (e.g. *ferry, skate, ski*).
- NONVEHICLE-51.4.2: verbs inherently denoting motion with a vehicle if the verb lemma is not the name of the vehicle itself (e.g. *drive, fly, sail*).
- WALTZ-51.5: verbs inherently implying dancing (e.g. *dance, tango, waltz*).
- CHASE-51.6: verbs inherently implying chasing (e.g. *chase, follow, track*).
- ACCOMPANY-51.7: verbs denoting an event with two participants, in which one accompanies the other (e.g. *accompany, escort, guide*).
- REACH-51.8³: verbs implying reaching a destination (e.g. *breast, hit, reach*).

Each of the classes above can have subclasses and sub-subclasses, i.e. more precise groupings of motion verbs. For instance, class ESCAPE-51.1 has one subclass ESCAPE-51.1-1, which in turn possesses three sub-subclasses, ESCAPE-51.1-1-1 (e.g. *depart, disembark, escape, exit, leave*), ESCAPE-51.1-1-2 (*enter, approach*), and ESCAPE-51.1-1-3 (*ascend, climb, cross, descend, pass*). From the examples in brackets, it seems that the three sub-subclasses of ESCAPE-51.1-1 are connected to motion away from (ESCAPE-51.1-1-1), to/towards (ESCAPE-51.1-1-2), or through/along a place (ESCAPE-51.1-1-3). For this reason, ESCAPE-51.1-1-1 seems to overlap with LEAVE-51.2, and ESCAPE-51.1-1-2 with REACH-51.8. The verbs that do not fit into any of these classifications occupy a higher position in the hierarchy, so they are generically assigned subclass ESCAPE-51.1-1 (e.g. *arrive, come, go*). The most general group, i.e. ESCAPE-51.1, includes three verbs which are probably not immediately associated with motion: *cut, get, and make it*. Usually, a verb only appears once in VerbNet, so it is only found in one class/subclass/sub-subclass. Nonetheless, one verb may appear twice, as for *approach*, found both in ESCAPE-51.1-1 and in its sub-subclass ESCAPE-51.1-1-2.

2.3 Preverbs

A preverb is a bound morpheme placed before a verbal base, which forms a semantic unit with this verbal base (cf. Booij and Van Kemenade 2003). In many cases, the meaning of the preverbed verb is compositional, resulting from the sum of the meaning of the verb and the meaning of the preverb, as in It. *circum-navigare* 'sail around', from *circum-* 'around' and *navigare* 'sail', or Ger. *ausgehen* 'go out', from *aus-* 'out' and *gehen* 'go'. In other cases, the meaning of the preverbed verb is not compositional, as in It. *sopravvivere* 'survive' (from Late Latin *supravivo*), from *sopra-* 'above' and *vivere* 'live'. Some languages – ancient and modern – allow for multiple preverbation (cf. Zanchi 2019; Farina 2021), i.e. verbs may possess more

² Overall, class 51.3 marks the so-called 'manner of motion verbs', i.e. motion verbs inherently referring to the way in which the Figure interacts with the Ground (Section 2.1.1).

³ This subclass was not present in Levin's (1993) classification and has been added in VerbNet.

than one preverb. For instance, this happens regularly in Ancient Greek as in *proekpléō* ‘sail off before’, resulting from *pro-* ‘before’, *ek-* ‘off from’, and *pléō* ‘sail’.

Semantically, a preverb can display different meanings (see Section 5.9). Consider, for example, *de-* in the following Italian verbs: *deportare* ‘deport’ from *de-* ‘away from’ and *portare* ‘carry’, *discendere* (*di-* < *de-*) ‘descend’ from *de-* ‘downwards’ and *scendere* ‘go down’, *detrarre* ‘detract’ from *de-* ‘away’ and *trarre* ‘take’, *decrescere* ‘decrease’ from *de-* ‘not’ and *crescere* ‘grow’.

2.4 Actionality

Actionality refers to the way in which an event is structured or presented in relation to time, and it has been referred to with different names, including lexical aspect and Aktionsart. To identify actional classes, Tatevosov (2002: 317) gives the following examples:

- (1) a. *John knows Russian.*
- b. *John walked in the garden.*
- c. *John ate an apple.*
- d. *John reached the summit.*

Table 2 shows a schematization of the time frames in which the events described in (1a)-(1d) take place. The parameters considered are (i) dynamicity, i.e. whether the situation is dynamic or stative; (ii) durativity, i.e. whether the situation is durative or instantaneous; (iii) telicity, i.e. whether the situation has an endpoint (telic) or not (atelic).

Sentence	Dynamic?	Durative?	Telic?
(1a) <i>John knows Russian.</i>	NO	YES	NO
(1b) <i>John walked in the garden.</i>	YES	YES	NO
(1c) <i>John ate an apple.</i>	YES	YES	YES
(1d) <i>John reached the summit.</i>	YES	NO	YES

Table 2. Schematization of sentences (1a)-(1d) according to the parameters of dynamicity, durativity, and telicity.

Many classifications have been proposed for verbal classes matching the parameters of Table 2. In literature, well-known examples are Vendler (1957), Kenny (1963), Dowty (1986), and Moens and Steedman (1988). Here, I focus on Vendler (1957), as I follow his classification for the annotation (see Section 4.2). Table 3 assigns Vendler’s (1957) labels to the classes in Table 2, providing examples from each class.

Vendler’s (1957) class	Dynamic?	Durative?	Telic?	Examples
State	NO	YES	NO	<i>love, know</i>
Activity	YES	YES	NO	<i>run, push a cart</i>
Accomplishment	YES	YES	YES	<i>run a mile, draw a circle</i>
Achievement	YES	NO	YES	<i>win a race, reach the top</i>

Table 3. Vendler’s (1957) actional classes and their characteristics with examples.

2.4.1 Actional shifts

Verbs are usually associated with one actional class. For instance, *love* usually denotes a state, as it usually denotes a non-dynamic, durative, and atelic situation. However, a verb can change actional class, undergoing a so-called actional shift (de Swart 1998; Zucchi 1998; Filip 1999)

or recategorization (Johanson 2000; Sasse 2002). Actional shifts may be caused by different lexical means.

1. The direct object (cf. Moens and Steedman 1988). Sometimes, a direct object may contribute to telicizing an atelic verb as in *I ate* (atelic, activity) and *I ate an apple* (telic, accomplishment). In other cases, it is the degree of specificity of the direct object that telicizes an event, as in *I wrote letters* (atelic, activity) and *I wrote the letter* (telic, accomplishment).
2. Temporal adverbials (cf. Dowty 1979). *For*-adverbials tend to occur with atelic situations, while *in*-adverbials occur with telic situations, as in *I wrote my essay for one hour* (atelic, activity) and *I wrote my essay in one hour* (telic, accomplishment).
3. Spatial adverbials (cf. Pustejovsky 1991). The presence of spatial relations such as the Goal (Section 2.1.2) may contribute to telicizing a verb, as in *I walked* (atelic, activity) and *I walked to the store* (telic, accomplishment).
4. Particles (cf. Brinton 1985). Some particles are markers of telicization, and their presence can turn an atelic situation into a telic situation, as in *I stood in the train* (atelic, activity) and *I stood up in the train* (telic, achievement).
5. Preverbs (cf. Booij and van Kemenade 2003). As for particles, some preverbs contribute to telicizing the situation expressed by the verb onto which they attach, as in Ger. *Otto trinkt Tee* ‘Otto drinks tea’ (atelic, activity) and *Otto trinkt Tee aus* ‘Otto drinks up tea’ (telic, accomplishment), from the verb *aus-trinken* ‘drink up’.

However, actional shifts may also be caused by grammatical aspect (cf. Comrie 1976), i.e. the grammatical category connected to the internal temporal development of a situation. In terms of grammatical aspect, the main aspectual opposition lies between the imperfective and the perfective aspect. The former represents a situation focusing on its development, regardless of its temporal starting and ending point. The latter represents a situation as a single complete whole, regardless of its internal development. An example of actional shift depending on grammatical aspect in Italian is provided below in (2a) and (2b), from Bertinetto (1986: 103).

- (2) a. *Gilberto calzava un paio di scarpe nere.*
 Gilberto put_on:IMPF.3SG a pair of shoes black
 ‘Gilberto was wearing a pair of black shoes.’
- b. *Gilberto calzò un paio di scarpe nere.*
 Gilberto put_on:PST.3SG a pair of shoes black
 ‘Gilberto put on a pair of black shoes.’

The examples in (2a) and (2b) show that the Italian verb *calzare* can possess two different meanings and denote different classes. The imperfect *calzava* in (2a) has an imperfective aspect and denotes an atelic situation (state). Conversely, the past *calzò* has a perfective aspect and denotes a telic situation (achievement).

2.5 WordNet

WordNet (WN) is a lexical database of English developed at Princeton University⁴ (Miller et al. 1990; Fellbaum 1998; Miller and Fellbaum 2007). It contains English verbs, nouns, adjectives, and adverbs. Each lemma is assigned one or more definitions, depending on the number of meanings a lemma has. These definitions are given in the form of a set of cognitive

⁴ <https://wordnet.princeton.edu>.

synonyms, so-called synsets, expressing different concepts. The term “cognitive synonyms” is rooted in the cognitive linguistic theory that underlies the design of WN. In the context of WN, cognitive synonyms within a synset are words that share similar meanings and can be used interchangeably in certain contexts. This reflects the cognitive linguistic principle that words are grouped based on conceptual relationships and mental representations. The synset gloss, which provides the definition for the lemma, is typically followed by one or more illustrative examples to enhance the understanding of the word's meaning.

Synsets are not isolated entities. Instead, they form a network of interconnected concepts, establishing lexical and semantic relations between each other. The hierarchical structure in which synsets are organized (Figure 2) mirrors the cognitive organization of concepts in the human mind (see e.g. Collins and Quillian 1969). The inclusion of semantic relations such as hypernymy and hyponymy, i.e. super-subordinate relations, aligns with the cognitive linguistic framework, emphasizing the cognitive processes involved in language comprehension and categorization.

- **S: (v) run** (move fast by using one's feet, with one foot off the ground at any given time) *"Don't run--you'll be out of breath"; "The children ran to the store"*
 - [direct troponym](#) / [full troponym](#)
 - [verb group](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (v) travel rapidly, speed, hurry, zip** (move very fast) *"The runner zipped past us at breakneck speed"*
 - [direct troponym](#) / [full troponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (v) travel, go, move, locomote** (change location; move, travel, or proceed, also metaphorically) *"How fast does your new car go?"; "We travelled from Rome to Naples by bus"; "The policemen went from door to door looking for the suspect"; "The soldiers moved towards the city in an attempt to take it before night fell"; "news travelled fast"*
 - [phrasal verb](#)
 - [derivationally related form](#)
 - [sentence frame](#)
 - [phrasal verb](#)
 - [derivationally related form](#)
 - [sentence frame](#)

Figure 2. Hierarchical structure for the first entry of the lemma *run* in Princeton WordNet.

Here, the verb *run* is defined by synset move fast by using one's feet, with one foot off the ground at any given time, whose hypernym, for instance, is move very fast, connected to lemmas *travel rapidly*, *speed*, *hurry*, and *zip*. In turn, the latter synset has the hypernym change location; move, travel, or proceed, also metaphorically, connected to generic verbs of motion such as *travel*, *go*, *move*, and *locomote*. This synset represents the highest level of the hierarchical chain based on hypernymy, as there are no other hypernyms above it.

Building onto Princeton WN (PWN), other WNs arose for both modern and ancient languages within the MultiWordNet (MWN) project⁵ (Pianta et al. 2002), initiated by Fondazione Bruno Kessler – at that time called Istituto Trentino di Cultura. The MWN project aimed to create a lexical network where different lemmas in different languages could be grouped together and compared based on the synset(s) to which they were connected. This not only helped with semasiological (see review in Marongiu et al. 2023: 2-4) and onomasiological analyses (e.g. recently Farina et al. 2023b), but it also paved the way for new NLP tasks such as information retrieval (e.g. Padave 2014; Basu et al. 2018; Ngo et al. 2018), semantic tagging (e.g. Haynes 2001; Andreevskaia and Bergler 2006; Cole and Gwizdka 2008), word sense disambiguation (e.g. McCarthy 2006; Siemiński 2011; AlMousa et al. 2022), and machine

⁵ <https://multiwordnet.fbk.eu/english/home.php>.

translation (e.g. Salm et al. 2010; Chakravarthi et al. 2019). Modern languages covered in MWN include Italian, Spanish, Portuguese, Hebrew, and Romanian, while ancient languages only include Latin. Although all these languages are accessible via MWN, the original project only included the Italian WN, aligned with PWN (Ciravegna et al. 1994). In the Italian WN, lemmas and word senses are interconnected through lexical and semantic relations (cf. Princeton WN above). However, compared to PWN new features are added, as MWN also provides correspondences between Italian and English lexical concepts and groupings based on semantic fields. WNs for other languages have been developed at the following universities: Universitat Politècnica de Catalunya (Spanish and Catalan WN), University of Lisbon (Portuguese WN), University of Haifa (Hebrew WN), University of Iasi (Romanian WN), University of Verona (Latin WN). All the WNs in MWN were originally created with semi-automatic procedures, bootstrapping from bilingual dictionaries (e.g. Collins English/Italian dictionary for the Italian WN; see below in detail for Latin).

With respect to historical languages, the Latin WN (Minozzi 2009) was also created by bootstrapping Latin-English pairs from bilingual dictionaries. Therefore, if a word w_E was the English translation of a Latin word w_L , then the synset s assigned to w_E was also assigned to w_L . The multilingual nature of MWN also allowed to use Latin-Italian bilingual dictionaries together with Latin-English dictionaries. In this case, if a Latin word w_L was the translation of the English word w_E and of the Italian word w_I , meaning that w_E and w_I had the same synset s , then the Latin word w_L was attributed s (Minozzi 2017). These processes sometimes led to noisy and anachronistic results, especially for more recent technological advances not present at the time of the Romans. The Latin WN has been checked for historical accuracy and manually corrected at the Catholic University of the Sacred Heart (Franzini et al. 2019) within the project LiLa: Linking Latin (Passarotti and Mambrini 2021). Within LiLa, the Latin WN has also been mapped to version 3.0 of PWN, whereas Minozzi had mapped it to PWN 1.6. More recently, an Ancient Greek WN was planned (Sausa 2012) and then created in 2014 (Bizzoni et al. 2014), adopting the same techniques as the Latin WN, i.e. mapping onto Greek-English bilingual dictionaries, and then exploiting Greek-Italian parallel texts and translations (Bizzoni et al. 2015). Ambitious projects towards the creation of new WNs for historical languages (Biagetti et al. 2021) are in progress, and they involve Latin, Ancient Greek, and Sanskrit. The new Latin WN aims to expand the Latin WN by Fondazione Bruno Kessler (see above), adding over 70,000 words and covering a time span from archaic to medieval Latin. It has also been included within the LiLa Project, and it is linked to the *Lexicon Translativum Latinum* (Fedriani et al. 2020), specifically designed for the study of Latin metaphors. These new WNs for historical languages are not complete yet and many lemmas still need to be checked, cleaned, and annotated. However, it is possible to query each raw version via its own API⁶. These new WNs are still based on PWN, but they aim to add the annotation of prepositions, as well as other linguistic information about lemmas (etymology and different textual/dialectal forms) and their meaning (e.g. they distinguish among literal, metaphorical, and metonymic senses of a word).

2.5.1 Issues and weaknesses of WordNet

Despite their pivotal role in linguistics and NLP (see Section 2.5 above), WNs are not perfect resources. The most common criticism that WNs have received in literature concerns their granularity (see Section 2.5.1.1). Here, I stress some other problematic aspects of WNs synsets creation and attribution.

⁶ Latin WN: <https://latinwordnet.exeter.ac.uk/api>; Ancient Greek WN: <https://greekwordnet.chs.harvard.edu/api>; Sanskrit WN: <https://sanskritwordnet.unipv.it/api>.

Within the synset network, some glosses do not convey the meaning of the lemma to which they are assigned or are empty (e.g. v#N0002603 None). For some of them, it is not possible to understand the POS of the lemma from the gloss itself. For instance, Lat. *desero* is attributed synset v#01761339 as of an organization, a country or an army. This synset only suggests that the word is used in the context of organizations, countries, or armies, but nothing in the gloss points towards a verbal meaning and no information about this meaning is actually contained in the gloss.

As for the attribution of synsets, some of them appear to be missed within the annotation of some words. Consider, for instance, the Italian adjectives *marino* ‘marine’ and *marittimo* ‘maritime’. Despite their semantic similarity, their meaning slightly differs, and the two adjectives are usually not interchangeable (e.g. *sale marino* ‘sea salt’, but **sale marittimo*; *commercio marittimo* ‘maritime trade’, but **commercio marino*). Nonetheless, both *marino* and *marittimo* are attributed synset a#02671223 relating to or involving ships or shipping or navigation or seamen. Synsets a#02670038 of or relating to the sea and a#00126777 native to or inhabiting the sea, attributed to Eng. *marine*, would probably better fit It. *marino*.

2.5.1.1 The issue of granularity

Some issues can be noticed concerning the creation of synsets and their granularity. Sometimes, two identical synsets have been generated, e.g. v#01350293 rise up and v#01839580 rise up \ "The building rose before them\ ". Apart from the addition of an English example, synset 01839580 does not change compared to 01350293, as they are both glossed with ‘rise up’. In other cases, two identical synsets with different ID exist within the synset network and the examples do not seem to point towards different meanings, as for v#01821151 go beyond: \ "Their loyalty exceeds their national bonds\ " and v#01820991 go beyond: \ "She exceeded out⁷ expectations\ ". Unlike the example with ‘rise up’, where one gloss does not possess an example, here both synsets possess a sentence used to clarify the meaning of the gloss. Nonetheless, there does not seem to be a distinction between the two synsets, as the meaning of ‘exceed’ in the two glosses appears to be extremely similar. Finally, consider the following synsets: n#02716224 a fortified place and n#03431121 a strongly fortified place. The only difference between the two synsets is the presence of the adverb ‘strongly’ in 03431121. However, it is extremely difficult to grasp the difference between a ‘fortified place’ and a ‘strongly fortified place’, especially considering that no examples are given after the gloss.

Due to these granularity issues, weaknesses have been stressed when WNs are used for word sense disambiguation tasks (Edmonds and Kilgarriff 2002; Ng et al. 2003; Palmer et al. 2004; McCarthy 2006; Snow et al. 2007). As sense distinctions contained in the WNs have proved to be sometimes unclear and of difficult recognition (see above), the synset choice may become extremely difficult (Edmonds and Kilgarriff 2002; Kilgarriff 2002). During their experiment in SENSEVAL-3, Snyder and Palmer (2004) used the English WN and calculated the Inter-Annotator Agreement (IAA) rate with human annotators, which was 72.5%. Verbs had the lowest IAA (67.8%; see above for issues on the granularity of verbs), while adjectives had the highest IAA (78.5%). It has been recognized that this relatively low IAA derives from the granularity of the WN sense inventory (Ng et al. 1999; Snow et al. 2007). When choosing a coarse-grained WN sense inventory, IAA increases (Ng et al. 1999; Navigli 2006).

⁷ Sic.

The granularity issues described above have been considered while developing the annotation scheme outlined here, as synsets for both nouns and verbs are involved in the annotation (see Section 4 and Sections 5.3, 5.4, 5.8, and 5.13). How I dealt with granularity is specifically explained in Sections 5.3, 5.5, and 5.13. As for the issues connected to problematic synsets mentioned at the beginning of this section, I have proceeded as follows. Leaving aside empty synsets as they do not carry any semantic information, I tried to avoid using ambiguous synsets, whenever possible. Nonetheless, sometimes a given word in context had to be associated with one of them, as with the case of ‘exceed’ above (v#01821151 go beyond: `\`"Their loyalty exceeds their national bonds\`\`" / v#01820991 go beyond: `\`"She exceeded out expectations\`\`"). In such cases, I first checked which synset was currently assigned to the corresponding lemma in the Ancient Greek or Latin WN using the API (see fn. 6), if any, and then chose that synset. If none or both synsets were assigned to that lemma, I selected the one whose meaning looked closer to the word in context, also based on the examples contained in the gloss, if any. It must be stressed, however, that such problematic cases were extremely rare in my annotation, and they did not impact the annotation scheme and/or its results.

2.6 World Historical Gazetteer

The World Historical Gazetteer (WHG)⁸ project (Manning and Mostern 2015; Manning 2015; Mostern 2017) is one of the most comprehensive historical gazetteers available online. On the one hand, it contains 60,000 world places annotated at different chronological levels, starting from the 16th century CE. On the other hand, it is interconnected to more than 141,000 historical places already annotated in other books/online resources and covering other time periods, starting from early antiquity. For instance, with respect to ancient places, more than 10,000 places mentioned in the DK Atlas of World History (Black 1999) and more than 20,000 places contained in the Pleiades gazetteer (Simon et al. 2016; Elliott 2021) are included in the WHG. Within the WHG, places are interconnected to one another throughout their own history. This takes into consideration, for instance, that a place may have changed its status (e.g. from ‘city’ to ‘state’) or its name, or that it may be known, or have been known, with different names through history (e.g. *Tiber* or *Tevere* to refer to the same river in central Italy flowing through the city of Rome). By including different types of geographical information in the same entry, drawn from different resources and time periods, the WHG aims to create a complex geographical network. Each place in the WHG has been annotated on different levels, which are not necessarily included in each entry: (i) place name and the datasets (i.e. other geographical resources) that mention the place; (ii) its linguistic variants (e.g. *Tiber/Tevere*); (iii) its type (e.g. ‘river’, ‘city’, ‘continent’); (iv) links to the datasets themselves; (v) periodization(s), useful to understand the different chronological stage(s) of a place; (vi) other locations to which the place might be related. Moreover, each entry in the WHG has been assigned a unique WHG identifier, and the place can also be visualized on a map. For instance, 25 results are provided when searching for the city of Athens, spread across different countries (Greece, Canada, the United States of America). The Greek city of Athens itself is linked to six different resources (GeoNames, Getty TGN, Euratlas Cities, Pleiades, Old World Trade, DK Atlas of World History), sometimes with different periodizations (Figure 3).

⁸ <https://whgazetteer.org>.

[return to search results](#)

Athens

ATTESTATIONS TO DATE

Athens dataset: GeoNames cities (more) issue? 🔗 pid: 2487444 📍
Variants: <i>ATH</i> ; <i>Afina</i> ; <i>Afini</i> ; <i>Afiny</i> ...73 more
Types: <i>populated place</i>
Links: 5 linked records gn:264371 ; loc:n79018143 ; viaf:131280462 ; wd:Q1524 ; tgn:7001393 ;
Athinai dataset: Getty TGN (partial) issue? 🔗 pid: 4867311 📍
Variants: <i>Āten</i> ; <i>Aθῆνα</i> ; <i>τὸ κλεινὸν ἄστυ</i> ; 25... more
Types: <i>inhabited place</i> ; <i>city</i>
Links: 1 linked records tgn:7001393 ;
Related: <i>within</i> Periféria Protevousis, Ellás ; Europe ; World
Athen dataset: EurAtlas Cities issue? 🔗 pid: 6368608 📍
Variants: <i>Athenae</i> ; <i>Athinai</i> ; <i>Atina</i> ; <i>Setines</i> ...3 more
Types: <i>city</i>
When: [1900, 2000]; [1300, 1400]; [200, 1200]; [1500, 1800]; [1600, 1600]; [0, 100]
Atene dataset: Pleiades (partial) issue? 🔗 pid: 6460262 📍
Variants: <i>Atēnē</i>
Types: <i>fort</i> ; <i>tower</i>
When: [-550, 300]
Athinai dataset: Old World Trade issue? 🔗 pid: 6375428 📍
Variants: <i>Afini</i> ; <i>Atene</i> ; <i>Athen</i> ; <i>Athenae</i> ...6 more
Types: <i>current settlement</i>
Links: 10 linked records bnf:11936474k ; gn:264371 ; gnd:4003366-1 ; loc:n79018143 ; pl:579885 ; tgn:7001393 ; viaf:131280462 ; wd:Q1524 ; wp:Athens ; wp:Category:Athens, Greece ;
When: [1300, 1600]
Athens dataset: DK Atlas of World History issue? 🔗 pid: 81642 📍
Variants: <i>Athens</i>
Types: <i>state</i>
Links: 6 linked records dbp:Athens ; gn:264371 ; loc:n79018143 ; tgn:7001393 ; viaf:131280462 ; wd:Q1524 ;
When: [1204, 1300]; [1200, 1400]

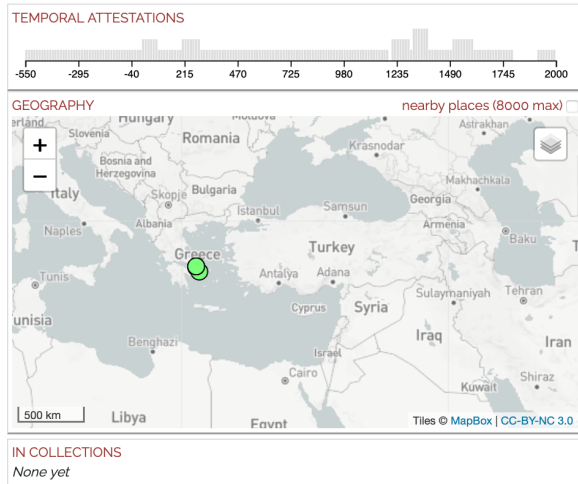


Figure 3. Results for ‘Athens’ (Greece) in the WHG. The left side of the figure shows six different results, each of them directing to a different resource (GeoNames, Getty TGN, EurAtlas Cities, Pleiades, Old World Trade, DK Atlas of World History). The right side of the figure shows the place on a map.

2.6.1 Why are places and texts related?

Textual material and geographical studies are strictly intertwined, especially for the study of the ancient world (e.g. Romm 1994; Purves 2002; 2010). For this reason, after the digitalization of ancient texts and technological advances in geo-spatial fields, recent projects have started to create digital resources for a new spatial understanding of the ancient world. For instance, based on GIS (Geographic Information System), Google Earth, and the Narrative TimeMap, the HESTIA (Herodotus Encoded Space- Text-Imaging Archive) Project⁹ (Barker et al. 2010) studies the ancient places mentioned in Herodotus’s *Histories* (5 cent. BCE). Google Ancient Places (GAP) (Isaksen et al. 2012) identifies historical places in the Google Books corpus and associates them with locations in the real world. More recently, the Pelagios Project (Simon et al. 2012; Barker et al. 2016b; Simon et al. 2016; Kahn et al. 2021; Vitale et al. 2021) seeks to develop a linked open data network encompassing geographical information pertaining to the ancient world. It combines different geographical resources, including the WHG and Pleiades. Pelagios has also created new tools for the visualization of geospatial data drawn from text, such as Peripleo¹⁰ (Barker et al. 2016a), where users can visualize geospatial data alongside with the text (Figure 4).

⁹ <https://hestia.open.ac.uk>.

¹⁰ <http://pelagios.org/peripleo-lite/>.

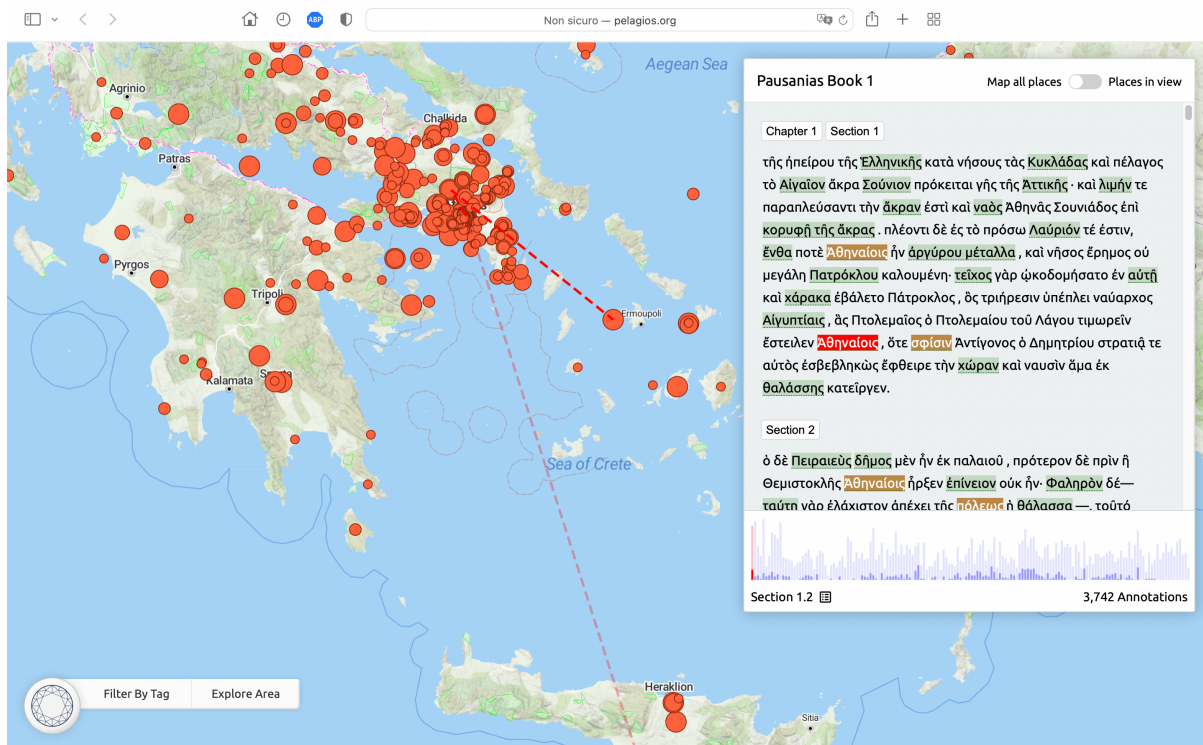


Figure 4. A screenshot from Periplo. The map marks all the places mentioned in Pausanias, *Periegesis*. On the right side, the places occurring in the text are highlighted in different colors.

Both HESTIA and Pelagios implicitly connect geographical locations and motion. In the Greek works they consider, several places are mentioned since people *move* across different locations. For this reason, given that the annotation scheme described here is specifically designed for verbs of motion, I also introduce an extra-linguistic annotation of places (see Section 5.8).

3. Annotation overview

After a theoretical introduction in Section 2, in Figure 5 I give a schematic representation of the annotation pipeline. This is the recommended order for the annotation, but an annotator may decide to start from other parameters. The annotation is composed of 20 layers, all mentioned in Figure 5 and Table 4 (Section 4.1). Six of them are not mandatory, as they cannot occur within a sentence: in this case, there would be nothing to annotate for them. In Figure 5, a lighter background corresponds to non-obligatoriness. If nothing is found for that layer, the annotator can proceed with the following layer. Numbers in Figure 5 match those in the first column of Table 4, where layers are listed in alphabetical order. Correspondences between the theoretical part outlined in the previous section of this work and the annotation layers are as follows: (i) Figure and Ground (Section 2.1.1) are relevant for layers 11 (PARTICIPANTS), 4 (FIGURE SYNSET) and 5 (GROUND SYNSET); (ii) spatial relations (Section 2.1.2) are relevant for layers 16 (SPACE) and 17 (SPATIALITY); (iii) motion verbs and verb classes (Section 2.2) are relevant for layer 19 (VERB CLASS); (iv) preverbs (Section 2.3) are relevant for layers 13 (PREVERB) and 14 (SEMPREV); (v) actionality (Section 2.4) is relevant for layer 1 (ACTIONALITY); (vi) WordNet (Section 2.5) is relevant for layers 4 (FIGURE SYNSET), 5 (GROUND SYNSET), and 18 (SYNSET); (vii) World Historical Gazetteer (Section 2.6) is relevant for layer 12 (PLACE).

A detailed explanation of the layers and the tagsets used for this annotation is provided in Section 4. Section 5 contains other relevant and more practical information to perform the annotation.



Figure 5. Pipeline for the linguistic annotation of preverbed motion verbs. Numbers match those in the first column of Table 4, where layers are listed in alphabetical order.

4. The annotation

4.1 Annotation layers

The annotation is performed using 20 layers, which include linguistic information at different levels. Three layers (LEMMA, MORPHOLOGICAL FEATURES, PART OF SPEECH) were already present in INCEpTION, when launching the tool for the first time, while all the others have been specifically designed and added for this annotation (see Figure 6). Table 4 lists layers in alphabetical order and contains different columns:

- #: a number that I assign to a layer in this document, proceeding in alphabetical order, useful for internal references between layers.
- Layer name: the name of the layer in the resource (see Figure 6).
- Type: the type of the layer within the INCEpTION tool. There are three types of layers:
 - Span layers: they attach onto one or more tokens and carry information assigned to the token(s).
 - Relation layers: they connect two span annotations. In Table 4, in the case of a relation layer, the number of the layer on which it acts is given in brackets. For instance, '[acting] on 11.' for layer 2. DEPENDENCY means that the relation layer DEPENDENCY will connect two span layers already annotated with PART OF SPEECH (layer 11. in Table 4).
 - Chain layers: they are used to form chains among span annotations.
- Where: the type of token possessing the layer (e.g. verb token, token expressing a spatial relation, the whole sentence).
- Function: what the layer identifies and annotates on the token (e.g. the layer LEMMA is used to annotate the lemma of a token).
- Obligatory: whether a given layer is mandatory (YES) or optional (NO). A layer can be optional as the tokens onto which it is attached are not necessary expressed in the sentence. For instance, it is not compulsory to express spatial relations with a motion verb since it can occur without such specifications (e.g. Eng. *The ship is sailing fast thanks to the strong wind*).

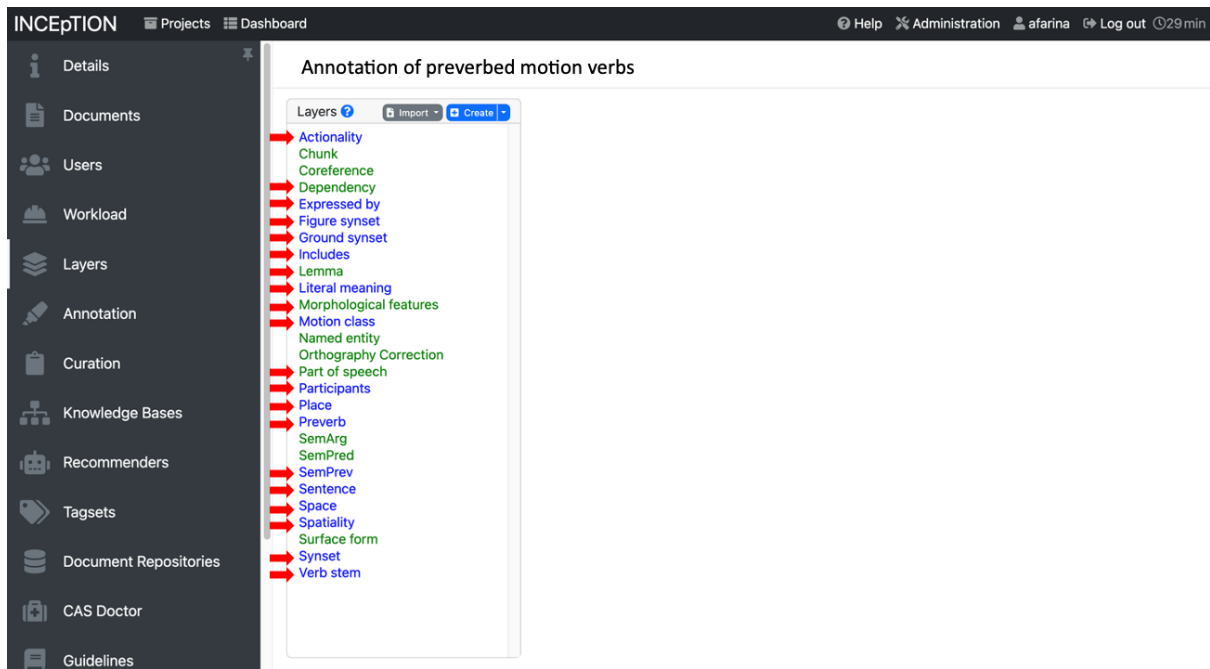


Figure 6. Layers used for this annotation in INCEpTION are pointed by a red arrow. Layers in green are included in INCEpTION when launching the tool. Layers in blue have been specifically designed for this annotation.

#	Layer name	Type	Where	Function	Obligatory
1.	ACTIONALITY	Span	On the verb token	It identifies the actionality of each verbal form within a given sentence, using the Vendlerian actional classification (Section 2.4).	YES
2.	DEPENDENCY	Relation (on 11.)	Between two tokens that have been attributed a POS (layer 10)	It identifies the syntactic relation between two tokens within a given sentence.	NO
3.	EXPRESSED BY	Span	On token expressing a spatial relation (in case of a PP, on the noun)	It identifies the way in which a spatial relation is expressed (e.g. with an adverb, a PP, a noun without preposition).	NO
4.	FIGURE SYNSET	Span	On the verb token	It identifies the meaning of the Figure (Section 2.1.1) using one or more synsets from the WordNet.	YES
5.	GROUND SYNSET	Span	On the verb token	It identifies the meaning of the Ground (Section 2.1.1) using one or more synsets from the WordNet.	NO

6.	INCLUDES	Relation (on 15.)	Between two tokens that have been attributed a SENTENCE layer (layer 15)	It connects the verb token to the whole sentence in which the verb occurs.	YES
7.	LEMMA	Span	On all tokens that possess at least one other annotated layer	It identifies the lemma of a given token.	YES
8.	LITERAL MEANING	Span	On the verb token	It identifies the meaning (literal/non-literal) of a verb token in a given sentence.	YES
9.	MORPHOLOGICAL FEATURES	Span	On the verb token	It provides the morphological analysis of the verb token in a given sentence.	YES
10.	PART OF SPEECH	Span	On all tokens that possess at least one other annotated layer	It identifies the part of speech of a token.	YES
11.	PARTICIPANTS	Relation (on 18.)	Between two tokens that have been attributed a SYNSET (layer 18), one token being the verb token, and the other token being a noun denoting the Figure or the Ground (see column “Function” on the right)	It identifies the entities – mostly people or places – participating in a motion event, i.e. its Figure and Ground (Section 2.1.1).	NO
12.	PLACE	Span	On the name of a place expressing a spatial relation (in	It identifies the name of the place that maintains a spatial relation with the verb token.	NO

			case of a PP, on the noun)		
13.	PREVERB	Span	On the verb token	It identifies the preverb of the preverbed form.	YES
14.	SEMPREV	Span	On the verb token	It identifies the meaning of the preverb in a given context (Section 2.3).	YES
15.	SENTENCE	Span	On the verb token and on the whole sentence	It marks the verb form and the whole sentence, which will be connected by layer INCLUDES.	YES
16.	SPACE	Span	On the verb token and on token expressing a spatial relation (in case of a PP, on the noun)	It marks a verb token and its spatial relation(s), if present (Section 2.1.2).	NO
17.	SPATIALITY	Relation (on 16.)	Between two tokens that have been attributed a SPACE layer (layer 16)	It connects the verb token to the PP or nominal form which expresses a spatial relation (Section 2.1.2).	NO
18.	SYNSET	Span	On the verb token and on nominal tokens possessing a layer PARTICIPANT (layer 11) or SPACE (layer 16)	It identifies the meaning of the verb token and the nominal tokens somehow related to the verb token itself.	YES
19.	VERB CLASS	Span	On the verb token	It identifies the verb class of the verb token (Section 2.2).	YES
20.	VERB STEM	Span	On the verb token	It identifies the verb stem in which the verb form is inflected.	YES

Table 4. Annotation layers used for the annotation of preverbed motion verbs.

Figure 7 shows an example of annotation where almost all the layers discussed in Table 4 have been used.

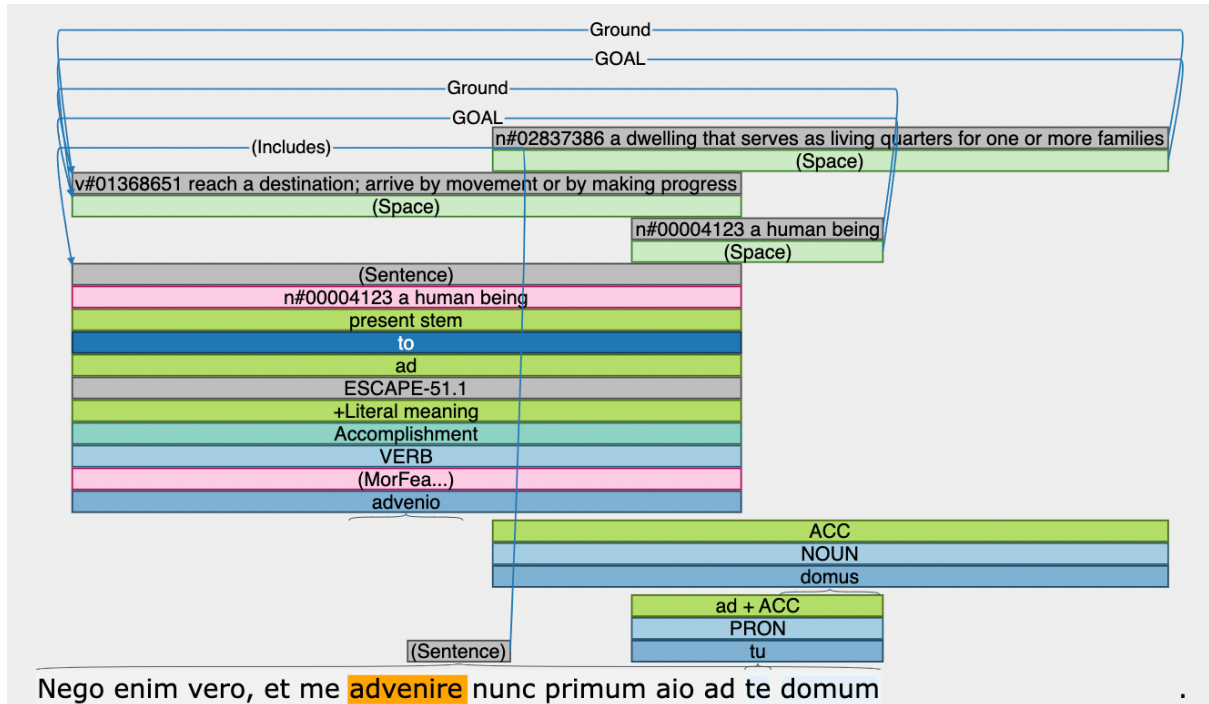


Figure 7. Example of annotation where almost all the layers discussed in Table 4 have been used. See below in the text for a detailed description of the layers.

Here, the verb token is *advenire*, and it has been assigned the following layers (from the bottom to the top): LEMMA, MORPHOLOGICAL FEATURES, PART OF SPEECH, ACTIONALITY, LITERAL MEANING, VERB CLASS, PREVERB, SEMPREV, SYNSET, VERB STEM, FIGURE SYNSET, SENTENCE, SPACE. The token *te*, part of the PP *ad te* expressing the Goal, has been assigned the following layers (from the bottom to the top): LEMMA, PART OF SPEECH, SYNSET. The entire PP *ad te* has been assigned the following layers (from the bottom to the top): EXPRESSED BY, SPACE. The token *domum* has been assigned the following layers (from the bottom to the top): LEMMA, PART OF SPEECH, SYNSET, EXPRESSED BY, SPACE. The tokens *advenire* and *ad te* are interconnected through the relation layer SPATIALITY, and the same holds for *advenire* and *domum*. The sentence *Nego enim vero, et me advenire nunc primum aio ad te domum* has been assigned the layer SENTENCE, connected to *advenire* through the layer INCLUDES.

4.2 Annotation tagsets

Some of the layers listed in Section 4.1 are based on specific tagsets, for a total of 11. They are described in Table 5. Two tagsets, i.e. UD UNIVERSAL DEPENDENCIES (v.2) and UD UNIVERSAL POS TAGS (v2), were already present in INCEPTION when launching the tool. They contain standard tagsets from the Universal Dependencies¹¹, which have not been adapted for this annotation – whenever possible, standard tagsets gave not been changed. All other tagsets were specifically designed and added for this annotation.

The columns of Table 5 describe the following parameters:

- **#**: a number that I assign to a tagset in this document, proceeding in alphabetical order, useful for internal references between tagsets.
- **Tagset name**: the name of the tagset in the resource (see Figure 8).
- **Connected to layer(s)**: which layer/layers uses/use the tagset.

¹¹ <https://universaldependencies.org>.

- **Tags:** the tags included in the tagset, put between single quotation marks. If the number of tags is higher than five, I generically describe the tags of the tagset, providing some examples.
- **Number of tags per layer:** whether one or more tags are allowed per each layer.
- **Closed:** whether the tagset is closed (YES), i.e. no more tags can be added to the tagset, or open (NO), i.e. the tagset is not decided *a priori*, and more tags can be added.

If a layer does not appear in the third column of Table 5, this means that it does not use any tagset for its annotation. Therefore, the values of the annotation are manually typed by the annotator. This is because these layers, among which, for instance, LEMMA and MORPHOLOGICAL FEATURES, cannot be based on pre-set tagsets. It would be extremely inconvenient to create a tagset with all the lemmas of a language or all the morphological features of a token, whereas it is easier to type them manually.

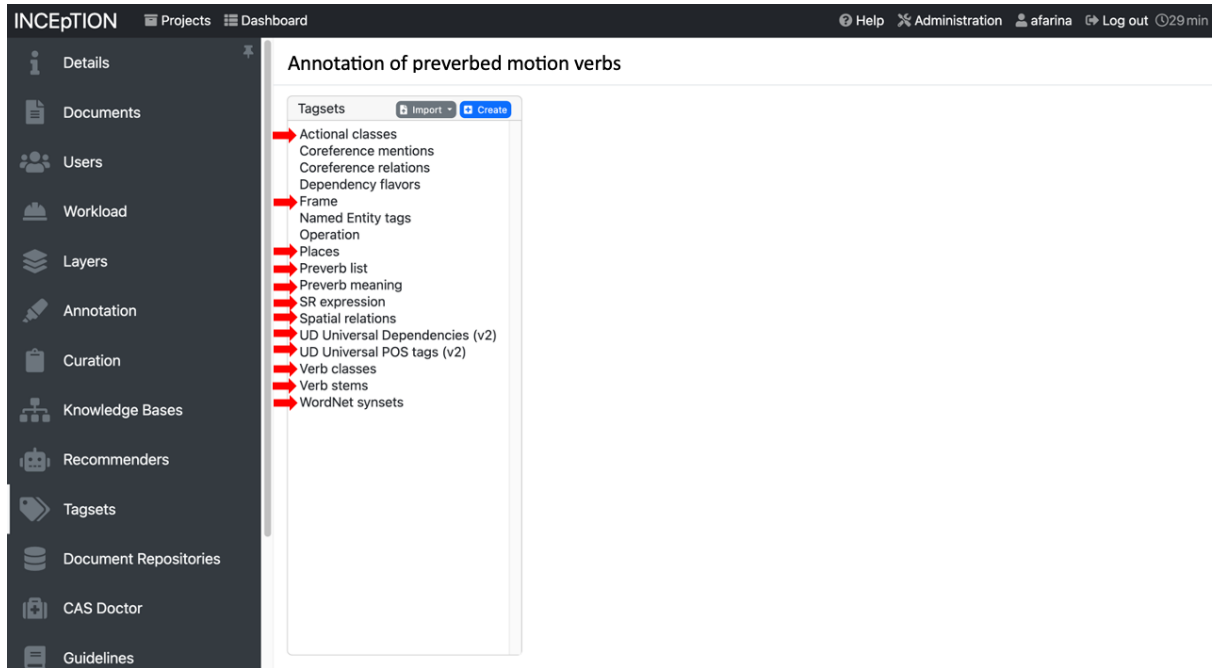


Figure 8. Tagsets used for this annotation in INCEpTION are pointed by a red arrow. Tagsets that are not marked with an arrow are included in INCEpTION when launching the tool. This also holds for the tagsets UD UNIVERSAL DEPENDENCIES (v2) and UD UNIVERSAL POS TAGS (v2), used for this annotation. All other tagsets have been created specifically for this annotation.

#	Tagset name	Connected layer(s) to	Tags	Number of tags per layer	Closed
1.	ACTIONAL CLASSES	ACTIONALITY	‘Accomplishment’ ‘Achievement’ ‘Activity’ ‘State’	Only one tag can be chosen.	YES
2.	FRAME	PARTICIPANTS	‘Figure’ ‘Ground’	Only one tag can be chosen.	YES
3.	PLACES	PLACE	List of places preceded by WHG identifier	Only one tag can be chosen.	NO
4.	PREVERB LIST	PREVERB	List of Ancient Greek and Latin	Only one tag can be chosen.	YES

			preverbs (e.g. ‘ab’, ‘ad’, ‘epí’, ‘ek’).		
5.	PREVERB MEANING	SEMPREV	List of preverb meanings selected a priori but expandible (e.g. ‘to’, ‘from’, ‘completely’, ‘together’).	More than one tag might be chosen, as sometimes a preverb encodes more than one meaning in context.	NO
6.	SR (= SPATIAL RELATION) EXPRESSION	EXPRESSED BY	List of morphological ways in which a spatial relation can be expressed (e.g. ‘ab + ABL’, ‘epí + GEN’, ‘ACC’, ‘adverb’).	Only one tag can be chosen.	NO
7.	SPATIAL RELATIONS	SPATIALITY	‘Goal’ ‘Location’ ‘Path’ ‘Source’	Only one tag can be chosen.	YES
8.	UD UNIVERSAL DEPENDENCIES (v. 2)	DEPENDENCY	List of 37 tags denoting syntactic relations among tokens within one sentence (e.g. ‘nsubj’, ‘obj’)	Only one tag can be chosen.	YES
9.	UD UNIVERSAL POS TAGS (v2)	PART OF SPEECH	List of 17 POS (e.g. ‘NOUN’, ‘VERB’)	Only one tag can be chosen.	YES
10.	VERB CLASSES	VERB CLASS	List of verb classes according to VerbNet (Section 2.2) (e.g. ‘ESCAPE-51.1’)	Only one tag can be chosen.	YES
11.	VERB STEMS	VERB STEM	‘aorist stem’ ‘supine stem’ ‘future stem’ ‘perfect stem’ ‘present stem’	Only one tag can be chosen.	YES
12.	WORDNET SYNSETS	FIGURE SYNSET GROUND SYNSET SYNSET	List of WordNet synsets of different POS (verbs, nouns, adjectives) (e.g. n#00004123 a human being)	More than one tag can be chosen as the annotation considers different levels of granularity (Section 2.5).	NO

Table 5. Annotation layers used for the annotation of preverbed motion verbs.

For the sake of clarity, here I stress that the first inventory of PREVERB MEANING (tagset 5) has been drawn from all the preverb meanings registered in Farina (2021), which are themselves taken from Chantraine (1999 [1968]), Luraghi (2003), and other works on specific prepositions (cf. references in Farina 2021). Other meanings are added following again Chantraine (1999 [1968]), but also Ernout and Meillet (2001 [1985]).

5. Practical recommendations about the annotation

This annotation scheme investigates different aspects of preverbed motion verbs, corresponding to all the linguistic parameters described in Section 2. Some of the annotation layers (DEPENDENCY, EXPRESSED BY, INCLUDES, LEMMA, MORPHOLOGICAL FEATURES, PART OF SPEECH, PLACE, PREVERB, SENTENCE, SPACE, SPATIALITY, VERB STEM) are straightforward, as they mostly rely on morpho-syntactical parameters. All other annotation layers are based on semantics and sentence interpretation, and this can lead to a higher level of disagreement among the annotators (see also Section 2.5.1). Here, I provide relevant information that can guide the annotator to perform the annotation on some of the layers considered above (Section 4.1). I will follow the same order as Table 4.

5.1 ACTIONALITY

Actionality is annotated following the Vendlerian classification (Sections 2.4, 4.1, 4.2). This means that the actional class should be assigned based on Table 3. Note that one lemma can denote more than one actional class in different contexts, depending on actional shifts (Section 2.4.1). Therefore, all the parameters mentioned in Section 2.4.1 must be considered when assigning an actional class to a verb token.

Consider the following examples from Farina (2021: 63) on the Ancient Greek motion verb *plēō* ‘sail’.

- (3) *aútis es héteron ploíon esbàs*
 again to another:ACC.SG boat:ACC.SG enter:PTCP.AOR.NOM.SG
*duōdeka hēméras pleúseai*¹²
 twelve day:ACC.PL sail:FUT.MID.2SG
 ‘Again, after boarding another boat, you will sail for twelve days.’ (Hdt.2.29.6)

- (4) *én te epì tèn khóran hēmôn*
 if PTC against ART.ACC.SG territory:ACC.SG 1PL.GEN
pezêi iōsin, hemeís epì tèn
 on_foot go:SUBJ.PRS.3PL 1PL.NOM against ART.ACC.SG
ekeinōn pleusoúmetha
 DEM.GEN.PL sail:FUT.MID.1PL
 ‘If they march against our territory, we will set sail against theirs.’ (Thuc.1.143.4)

Both *pleúseai* in (3) and *pleusoúmetha* in (4) are inflected in the future, a tense that is not characterized by any specific grammatical aspect (Section 2.4.1) in Ancient Greek. The

¹² In this document I follow the Leipzig glossing rules (<https://www.eva.mpg.de/lingua/resources/glossing-rules.php>). Nonetheless, to keep the glosses shorter I have omitted some categories. I have always omitted gender in nouns, pronouns, and adjectives. I have specified verbal mood only when different from the indicative, and verbal diathesis only when different from active. AGr. *autós* is always glossed as a demonstrative: its function can be retrieved from the translation.

presence of the temporal adverbial *duódeka hēméras* ‘for twelve days’ in (3) suggests that the verb denotes an activity. In (4), there is no linguistic means (Section 2.4.1) that points towards a specific actional class. It is rather the context and the interpretation of the sentence that suggests that the future *pleusoúmetha* denotes an achievement. This is also reflected in the English translation of the passages, as *plēō* means ‘sail’ in (3) and ‘set sail’ in (4).

In some languages, the grammatical aspect is extremely important in actional shifts. For this reason, assigning an actional class to a verb token cannot be based on a mere translation of the original text. Let us consider the two occurrences of *apoplēō* in (5) and (6), from Farina (2021: 81-82).

(5) *héna autôn katalipóntes hoi*
 one:ACC.SG DEM.GEN.PL leave:PTCP.AOR.NOM.PL ART.NOM.PL
loipoi apépleon eis Delphoús
 rest:NOM.PL sail_away:IMPF.3PL to Delphi:ACC.PL
 ‘After leaving there one of them, the rest was sailing to Delphi.’ (Hdt.4.157.1)

(6) *hoútō mèn dē apépleusan apraktoi ek*
 so PTC PTC sail_off:AOR.3PL doing_nothing:NOM.PL from
Milētou hoi Pérsai
 Miletus:GEN.SG ART.NOM.PL Persian:NOM.PL
 ‘Hence, the Persians sailed off from Miletus without success.’ (Arr.An.1.19.11)

In Ancient Greek, the present stem has an imperfective aspect, while the aorist stem has a perfective aspect (Section 2.4.1). Therefore, (5) denotes an activity due to the imperfect *apépleon* (imperfective), but (6) denotes an accomplishment due to the aorist *apépleusan* (perfective).

5.2 EXPRESSED BY

The layer EXPRESSED BY is used to mark the morphological or morpho-syntactic way in which a given spatial relation is expressed. The tagset is not chosen *a priori* (cf. Table 5), so it can be expanded if new cases or prepositions appear in the corpus. Considering only Ancient Greek and Latin, it is evident that there is a limit to the ways in which spatial relations can be expressed, i.e. a limited set of prepositions that can appear in these languages. However, not choosing the tagset *a priori* makes the annotation scheme susceptible to expansions, especially if annotating more languages at once (cf. e.g. Section 6.1).

Apart from grammatical cases, marked with their abbreviation (first three letters) in capital letters (‘ACC’ for accusative, ‘ABL’ for ablative, ‘DAT’ for dative, ‘GEN’ for genitive, ‘LOC’ for locative) and prepositions + grammatical cases (e.g. ‘ab + ABL’, ‘ek + GEN’, ‘in + ACC’), the tag ‘adverb’ is chosen if a spatial relation is expressed by an adverb (e.g. Lat. *huc* ‘hereto’, *illac* ‘through there’, AGr. *ekeíse* ‘thereto’, *ekeí* ‘there’). Grammatical cases are specified even after prepositions that only govern one case (e.g. ‘ek + GEN’, ‘ab + ABL’), so that all the tags are in the form ‘preposition + CASE ABBREVIATION’. The EXPRESSED BY layer is always annotated on the noun (cf. Table 4), even in the case of PPs. For instance, considering (6) again, *ek Milētou* would be annotated with ‘ek + GEN’ on the token *Milētou*.

5.3 FIGURE SYNSET

Given that Figure and Ground are the participants of a motion event¹³ (see Section 2.1.1), the motion verb also carries the layers FIGURE SYNSET and GROUND SYNSET (see Table 4 and Section 5.4). The layer FIGURE SYNSET is always annotated onto a verb token, regardless of the Figure being overtly mentioned in the sentence. For instance, if the Figure is syntactically encoded as the implicit subject of a motion verb, it will still be annotated onto the verb form. This is done in order to collect enough data to analyze e.g. what Figures tend to occur the most with a given verb. Thanks to the layer PARTICIPANTS (see Table 4 and Section 5.7), it will then be possible to isolate those cases in which the Figure is overtly expressed in a sentence.

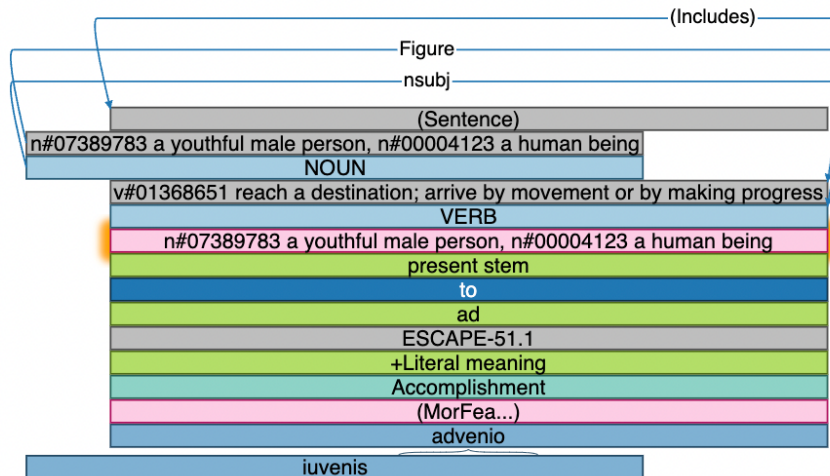
As the Figure is usually a noun, semantic annotation is conducted on different levels of granularity, to allow for different levels of analysis. For this reason, in some cases more than one synset is connected to a nominal token. These cases are analyzed below and include common nouns referring to people as well as proper nouns referring to places.

(7) *Commodum cubuerant ecce quidam longe plures*
just lie_down:PPF.3PL PTC some:NOM.PL far many:NOM.PL
numero iuvenes adveniunt alii
number:ABL.SG young_man:NOM.PL arrive:PRS.3PL other:NOM.PL
‘They had just taken their places when another much larger group of young men arrived.’ (Apul.Met.4.8)

In (7), the Figure is the group of young men. In the English WordNet, the synset connected to ‘young man’ is n#07389783 a youthful male person. However, to assign only this synset for the Figure of this verb occurrence would not allow a less granular analysis distinguishing, for instance, only animate and inanimate entities. In other words, if only this synset was assigned to the Figure, then it would be more time-consuming to group it together with synsets describing other human beings, such as, for instance, n#07391044 an adult male person (as opposed to a woman) (‘man’), n#07258194 a male parent (also used as a term of address to your father) (‘father’), n#07242378 an armed adversary (especially a member of an opposing military force) (‘enemy’), or n#07215549 someone who rules unconstrained by law (‘dictator’), which are all instances of animate entities. WordNet (see Section 2.5) has a hierarchical structure considering hyponyms and hypernyms, which allows to group together different synsets that have, for instance, a shared hypernym synset. However, to facilitate a higher-level analysis and a quicker and more convenient data collection, the hypernym of n#07389783 a youthful male person in the English WordNet, which is the synset for ‘person’ (n#00004123 a human being), is also added to the annotation.

When annotating cases such as (7), the first synset to be selected on INCEpTION must be the synset of the Figure. The hypernym must be selected as the second synset. Figure 9 shows how this double annotation is visualized in the INCEpTION interface.

¹³ In the specific context of the study for which this annotation scheme has been created.



am longe plures numero iuvenes adveniunt alii, quos incunctanter adaeque latrones a
Figure 9. Annotation of the motion verb *adveniunt* in (7). The layer assigning synsets to the Figure is highlighted in orange.

If the Figure is a proper noun denoting a person, the synset of the proper noun is omitted, and only the highest hypernym (n#00004123 a human being) is assigned to the Figure. This occurs for two reasons. First, I am not interested in evaluating how many times a proper noun occurs with a given motion verb. Secondly, and more technically, WordNet does not include proper nouns unless they are well-known, for instance, due to historical reasons. Figure 10 below shows the results for *James* in the English WordNet.

Noun

- [S: \(n\) James, James IV](#) (a Stuart king of Scotland who married a daughter of Henry VII; when England and France went to war in 1513 he invaded England and died in defeat at Flodden (1473–1513))
- [S: \(n\) James, James II](#) (the last Stuart to be king of England and Ireland and Scotland; overthrown in 1688 (1633–1701))
- [S: \(n\) James, James I, King James, King James I](#) (the first Stuart to be king of England and Ireland from 1603 to 1625 and king of Scotland from 1567 to 1625; he was the son of Mary Queen of Scots and he succeeded Elizabeth I; he alienated the British Parliament by claiming the divine right of kings (1566–1625))
- [S: \(n\) James, Jesse James](#) (United States outlaw who fought as a Confederate soldier and later led a band of outlaws that robbed trains and banks in the West until he was murdered by a member of his own gang (1847–1882))
- [S: \(n\) James, William James](#) (United States pragmatic philosopher and psychologist (1842–1910))
- [S: \(n\) James, Henry James](#) (writer who was born in the United States but lived in England (1843–1916))
- [S: \(n\) James, Saint James, St. James, Saint James the Apostle, St. James the Apostle](#) ((New Testament) disciple of Jesus; brother of John; author of the Epistle of James in the New Testament)
- [S: \(n\) James, James River](#) (a river in Virginia that flows east into Chesapeake Bay at Hampton Roads)
- [S: \(n\) James, James River](#) (a river that rises in North Dakota and flows southward across South Dakota to the Missouri)
- [S: \(n\) Epistle of James, James](#) (a New Testament book attributed to Saint James the Apostle)

Figure 10. Results for *James* in the English WordNet.

The same holds for Roman and Greek people. The synsets assigned to *Caesar* and *Alexander* in the English WordNet are n#07703440 conqueror of Gaul and master of Italy (100–44 BC) and n#07694765 king of Macedon; conqueror of

Greece and Egypt and Persia; founder of Alexandria (356–323 BC), respectively. However, most of the proper nouns are not included in WordNet. They will be included in the new versions of the Latin and the Ancient Greek WordNet (see Section 2.5). To avoid any disproportion between proper nouns that possess a synset in the English WordNet (e.g. *Caesar*, *Alexander*), and nouns that do not (e.g. *Eurialus*, *Xerxes*), I omit the synset of the proper noun even when available and annotate the highest hypernym. I also omit in-between layers such as n#07674205 someone engaged in or experienced in warfare (*warrior*, with reference to Eurialus) or n#07354565 a male sovereign; ruler of a kingdom (*king*, with reference to Xerxes), as they do not convey any relevant information for the study for which this annotation scheme has been designed. Of course, if useful, granularity can be added to the annotation.

If the Figure is a proper noun denoting a place, two types of annotations are possible. If the place possesses a synset in PWN, then it is assigned both this synset and the synset of the corresponding common noun. For instance, in PWN the river Rhine is associated to synset n#06829293 a major European river carrying more traffic than any other river in the world; flows into the North Sea. In the annotation scheme described here, the river Rhine would be annotated with both this synset and the synset for river, i.e. n# 06789983 a large natural stream of water (larger than a creek). If the place does not possess a dedicated synset in PWN, which usually occurs with small or less common places, then it will only be annotated with the corresponding common noun. Annotating the common noun also in the first case allows the researcher to filter the annotation and consider both cases jointly.

Personal pronouns – and pronouns in general – are not annotated in the WordNets, but in fact they can act as the Figure of a motion event in context. In these cases, they are given the generic synset n#00004123 a human being, considering the referent of the personal pronoun itself. This is useful, for instance, in order to have a complete overview of how many animate entities perform a motion event with a given lemma.

When a motion verb has a literal meaning, usually its Figure consists in a moving entity. If a verb is used in a non-literal sense, it should not always be possible to identify a proper Figure, as well as a proper Ground.

- (8) *autàr Akhaiois aspasiē trillistos*
but Achaeans:DAT.PL welcome:NOM.SG three_times:NOM.SG
epéluthe nùx erebenné
come_upon:AOR.3SG night:NOM.SG dark:NOM.SG
‘But the dark night, gladly welcome and three times prayed for, came upon the Achaeans.’ (Hom.II.8.488)

In (8), the ‘night’ is represented as a moving entity, even though it is not physically able to perform any motion. This annotation scheme uses the labels Figure and Ground (see Sections 2.1.1 and 5.4) even in metaphorical cases. Therefore, in the case of (8), the Figure synset of the verb form *epéluthe* will be n#10885886 the time after sunset and before sunrise while it is dark outside. This annotation is useful to quantitatively analyze what ‘metaphorical Figures’ are most common with specific verbal base, regardless of their syntactic function in the sentence and without creating another annotation layer to be used only with non-literal meanings.

As the Figure is also annotated in non-literal verbal senses, sometimes it can be a whole sentence.

- (9) *trecentos* *sex* *perisse* *satis* ***convenit***
 three_hundred:ACC.PL six die:INF.PF enough agree:PRS.3SG
 ‘Three hundred and six men died, as it is generally agreed.’ (Liv.2.50)

Here, *convenio* ‘come together’ acquires the metaphorical meaning ‘agree’, and the morphological subject of the sentence, as well as the Figure, is the clause that precedes *convenit*. In such cases, layer FIGURE SYNSET is annotated with **clause**.

Finally, there can be more than one noun serving as the Figure for one verb occurrence.

- (10) *nam* *cum* *omnis* *iuventus*, *omnes* *etiam* *gravioris*
 PTC as all:NOM.SG youth:NOM.SG all:NOM.PL too older:GEN.SG
aetatis [...] *eo* ***convenerant*** [...]
 age:GEN.SG there come_together:PPF.3PL
 ‘For, when all the younger men, yet all the older men [...] had assembled there [...].’
 (Caes.BG.3.16)

There are two Figures in (9), the younger men (*omnis iuventus*) and the older men (*omnes [...] gravioris aetatis*). In this case, the layer FIGURE SYNSET collects both Figures getting the generic synset for ‘person’, i.e. n#00004123 a human being. The meanings of the single Figures will then be distinguished via the layer SYNSET. *Iuventus* will be assigned synset n#05959071 young people collectively, while *omnes*, referring to *omnes [...] gravioris aetatis*, will be assigned synset n#07442529 a man who is old.

Sometimes, the Figure can be a multiword expression.

- (11) *Aulus* *Vitellius* *inferiorem* *Germaniam*
 Aulus:NOM.SG Vitellius:NOM.SG lower:ACC.SG Germany:ACC.SG
ingressus *hiberna* *legionum*
 enter:PTCP.PF.NOM.SG winter_quarters:ACC.PL legion:GEN.PL
cum *cura* ***adierat***
 with care:ABL.SG inspect:PPF.3SG
 ‘[On December 1st of the preceding year,] Aulus Vitellius, entering Lower Germany, had carefully inspected the winter quarters of the legions.’
 (Tac.Hist.1.52)

In (10), the Figure is a multiword expression composed of the praenomen *Aulus* and the nomen *Vitellius*. In this case, the Figure is simply assigned synset n#00004123 a human being. To see where the synset n#00004123 a human being is put in cases such as *Aulus Vitellius*, see the layer SYNSET in Section 5.13.

In some cases, a motion verb can have no Figure.

- (12) *Horatio* *sorte* ***evenit***
 Horatius:DAT.SG destiny:ABL.SG fall_onto:PF.3SG
 ‘[The consuls Valerius and Horatius drew lots to determine who should do it.]
 It fell onto Horatius.’ (Liv.2.8)

In (11), *evenit* is impersonal, and it has no Figure, as it is not followed by any subject clause. Therefore, the layer FIGURE is annotated with NA.

5.4 GROUND SYNSET

This layer is annotated on the verb – I refer to Section 5.3 for this explanation. Unlike the layer FIGURE SYNSET (see Section 5.3), the layer GROUND SYNSET is annotated only if the Ground is overtly mentioned in context, thus not considering anaphora zero, i.e. the omission of an overt reference term, for the Ground. While cases of anaphora zero with the Figure, usually syntactically encoded as the subject of a motion verb, are simple to grasp and are annotated (see Section 5.3), cases of anaphora zero with the Ground may sometimes become extremely ambiguous. Due to the properties of the Ground (Table 1), in some occurrences it may be difficult to evaluate whether the Ground is actually encoded with anaphora zero or whether it is not encoded at all. Moreover, the research for which this annotation scheme has been created does not include a deep investigation on the Ground generically. It rather focuses on those occurrences where the Ground is overtly expressed.

As it occurs for the layer FIGURE SYNSET (see Section 5.3), the Ground is annotated also in cases of metaphorical motion if it is overtly expressed. For instance, recalling (8), the annotation of the layer GROUND SYNSET would be `n#00004123 a human being` (for the use of the generic synset for ‘person’ instead of a more specific synset for ‘Achaean’ see Section 5.3).

5.5 LITERAL MEANING

Literal meaning is annotated based on a Boolean feature whose values are TRUE and FALSE. A bivalent feature allows to keep the annotation simpler and quicker and helps reducing ambiguity, by forcing annotators to make a clear-cut decision. This may lead to more consistent annotation. From a computational point of view, binary features often require less computational resources than gradient features. The end of this section will point out that granularity is not lost with this annotation scheme, as it can be retrieved by other annotation layers.

A literal meaning with a TRUE value is annotated in all the cases in which the verb displays a literal meaning, i.e. when the motion verb describes a real motion as in (3)-(7) and the resulting meaning of the preverbed verb is basically compositional. A literal meaning with a FALSE value is annotated in all the other cases, i.e. when the motion verb has either a metaphorical or – more rarely – a metonymic sense and/or the meaning of the preverbed verb is non-compositional. If a preverb underwent lexicalization (e.g. Meillet and Vendryes 1963; Cuzzolin 1995; López Moreda 1998; McGillivray 2013), the literal meaning of the resulting verb would likely be annotated with FALSE. In a comparative perspective, treating lexicalization cases together with occurrences where the preverb is not lexicalized is extremely useful to quantitatively analyze how different languages behave with respect to the lexicalization of preverbs itself. Other layers such as SEMPREV (see Section 5.10), SYNSET (see Section 5.13), and VERB CLASS (see Section 5.14) will help disambiguation between lexicalized verb or metaphorical use of the verb during the analysis.

(13)	<i>qui</i> REL.NOM.PL	<i>in</i> during	<i>itinere</i> journey:ABL.SG	<i>congressi</i> meet:PTCP.PF.PASS.NOM.PL
	<i>magnopere</i> earnestly	<i>ne</i> that_not	<i>longius</i> further	<i>progredere tur</i> advance:SBJ.IMPF.PASS.3SG
	<i>orabant</i> beseech:IMPF.3PL			

‘They met him during their journey and earnestly besought him not to advance further.’ (Caes.BG.4.11)

- (14) *denique* *hos* *esse* *eosdem* *Germanos*
 finally DEM.ACC.PL be:INF.PRS same:ACC.PL German:ACC.PL
quibus-cum *saepe* *numero* *Helvetii*
 REL.ABL.PL-with often number:ABL.SG Helvetius:NOM.PL
congressi [...] *plerumque* *superarint*
 fight:PTCP.PF.PASS.NOM.PL very_frequently defeat:SBJ.PF.3PL
 ‘Finally, these are the same men with whom the Helvetii had often fought, very frequently defeating them.’ (Caes.BG.1.40)

The verb *congregior* is composed of *cum* ‘together (with)’ and *gradior* ‘walk, go’, therefore it means ‘go, come, meet with’. This is precisely the meaning of *congressi* in (13), annotated with literal meaning TRUE. The meaning ‘fight’ in (14) still results from *cum* + *gradior*, with the addition of a negative connotation. Given that the meaning is not strictly compositional, *congressi* in (14) is annotated with literal meaning FALSE. The difference between (13) and (14) is also reflected in other annotation layers, such as SEMPREV (see Section 5.10; *congressi* is annotated with ‘together’ in (13), and with both ‘together’ and ‘(malefactive)’ in (14)), or VERB CLASS (see Section 5.14; *congressi* is annotated with MEET-36.3 in (13) and BATTLE-36.4-1 in (14)¹⁴).

In the case of metaphorical motion, the Figure is usually an abstract noun (see Section 5.3), as in (8), quoted again, and (15).

- (8) *autàr Akhaioîs* *aspasiē* *trillistos*
 but Achaean:DAT.PL welcome:NOM.SG three_times:NOM.SG
epēluthe *nūx* *erebennē*
 come_upon:AOR.3SG night:NOM.SG dark:NOM.SG
 ‘But the dark night, gladly welcome and three times prayed for, came upon the Achaeans.’ (Hom.II.8.488)

- (15) *diditur* *hic* *subito* *Troiana* *per*
 spread:PRS.3SG.PASS then suddenly Trojan:ACC.PL through
agmina *rumor* *advenisse* *diem* *quo*
 band:ACC.PL rumor:NOM.SG come:INF.PRF day:ACC.SG REL:ABL.SG
debita *moenia* *condant*
 promise:PART.PRF.ACC.PL city_walls:ACC.PL found:SUBJ. PRS.3SG
 ‘Then suddenly a rumor spreads through the Trojan bands, that the day to find their promised city has come.’ (Verg.Aen.7.144-145)

The motion verbs occurring in (8) and (15) acquire a metaphorical meaning due to their Figure. The *night* and the *day* are not concrete entities able to perform any actual motion. For this reason, both *epēluthe* in (8) and *advenisse* in (15) are annotated with literal meaning FALSE.

Metaphors have different levels of depth. This annotation scheme does not distinguish between them, as it marks cases like *congressi* in (14), *evenit* in (12), *epēluthe* in (8), and *advenisse* in (15) with a non-literal meaning. Indeed, there is a difference between the three examples. In (14), the Figure is human, and the metaphorical meaning derives from a negative/hostile connotation acquired by the verb *congregior*. Conversely, while the alternation

¹⁴ Note that also in VerbNet both MEET-36.3 and BATTLE-36.4-1 are subgroups of class 36.

of days and nights for (8) and (15) can be easily perceived as a sort of motion, *evenit* in (12) seems to be used in a “more metaphorical” sense, as no metaphorical motion is actually implied in the sentence. This is because the preverb *ex* with *venio* is highly lexicalized. To distinguish among different degrees of metaphors, we use the layer SYNSET (see Section 5.13). The synset connected to *epéluthe* in (8) and *advenisse* in (15), v#00236668 come to pass; arrive, as in due course, is connected to motion, and this is reflected in WordNet’s hierarchical structure. This does not hold for *evenit* in (11), assigned synset v#01768893 chance to be or do something, without intention or causation, which does not imply any motion.

5.6 MORPHOLOGICAL FEATURES

Morphological features of the verb tokens are annotated based on the universal feature inventory¹⁵ of the Universal Dependencies. As occurs within the Universal Dependencies, morphological features are separated with a vertical bar (|) and represented as attribute-value pairs, with an equal symbol (=) separating the attribute from the value¹⁶. Figure 11, drawn from the Universal Dependencies¹⁷, shows an example of annotation for the English sentence *They buy and sell books*.

# text = They buy and sell books.					
1	They	they	PRON	PRP	Case=Nom Number=Plur
2	buy	buy	VERB	VBP	Number=Plur Person=3 Tense=Pres
3	and	and	CCONJ	CC	—
4	sell	sell	VERB	VBP	Number=Plur Person=3 Tense=Pres
5	books	book	NOUN	NNS	Number=Plur
6	.	.	PUNCT	.	—

Figure 11. Annotation of the tokens for the English sentence *They buy and sell books*. Verbs are highlighted in red. Figure taken from <https://universaldependencies.org/format.html>.

Within the layer MORPHOLOGICAL FEATURES, forms such as Lat. *advenit* (perfect) and *adveniens* (masculine) are annotated as in (14) and (15).

(16) Mood=Ind|Number=Sing|Person=3|Tense=Past|VerbForm=Fin|Voice=Act

(17) Case=Nom|Gender=Masc|Number=Sing|Tense=Pres|VerbForm=Part|Voice=Act

Comparing (14) and (15) with the annotation in Figure 11, (14) and (15) are richer in morphological features, as they also include ‘Mood’, ‘VerbForm’, and ‘Voice’.

¹⁵ <https://universaldependencies.org/u/feat/index.html>.

¹⁶ «The FEATS field contains a list of morphological features, with vertical bar (|) as list separator and with underscore to represent the empty list. All features should be represented as attribute-value pairs, with an equals sign (=) separating the attribute from the value. In addition, features should as far as possible be selected from the universal feature inventory and be sorted alphabetically by attribute names. It is possible to declare that a feature has two or more values for a given word: Case=Acc,Dat. In this case, the values are sorted alphabetically» (<https://universaldependencies.org/format.html>).

¹⁷ <https://universaldependencies.org/format.html>.

5.7 PARTICIPANTS

The layer PARTICIPANTS refers to the participants in a (motion) event, i.e. Figure and Ground (see Section 2.1.1).

Usually, the verb describing motion events is inflected in the active voice. However, in some cases it can be inflected in the passive, to reverse the perspective of the sentence, thus swapping the Figure and the Ground (cf. Tất Thắng 2013).

- (18) *ut* [...] *dici* *posset* *eos*
 so_that say:INF.PRS.PASS can:SUBJ.IMPF.3SG 3PL:ACC
ab *se* *per* *fidem* *in*
 by 3SG.REFL:ABL against loyalty:ACC.SG in
conloquio ***circumventos***
 parley:ABL.SG surround:PTCP.PERF.ACC.PL
 ‘[...] so that it could be reported that they had been surrounded by him during a parley after pledge given.’ (Caes.BG.1.46)

In (18), the Figure is *eos* and the Ground is *se*. The entity with the most prominent position in the text are the enemies, i.e. the referent of the pronoun *eos*.

As for the Ground, it is less frequently overtly expressed in context, so it is less frequently annotated both with the layer PARTICIPANTS and with the layer GROUND SYNSET. An instance of Ground expression is given in (19) from Farina (2021: 50).

- (19) *adeôs* *geōrgoûntes* *kai* *tên* *thálattan*
 confidently be_a_farmer:PTCP.NOM.PL and ART.ACC.SG sea:ACC.SG
pléontes *kai* *taîs* *állais* *ergasíais*
 sail:PTCP.NOM.PL and ART.DAT.PL other:DAT.PL work:DAT.PL
epikheiroûntes
 put_the_hand:PTCP.NOM.PL
 ‘Confidently being farmers and sailing the sea and putting our hands to the remaining works.’ (Isoc.8.20)

Here, the Ground is expressed by the phrase *tên thálattan*, and it is morphologically realized as the direct object of the motion verb *pléō*. In (18), the Ground is a person, namely Caesar, expressed by the PP *ab se*.

5.8 PLACE

For the annotation of PLACE, the annotator only selects the name of a place expressing a spatial relation in the motion event. Here, I will comment on how to add a place to the tagset PLACES, in case it is not already present there. Recall (6), quoted again here.

- (7) *hoútō mèn dē* ***apépleusan*** *ápraktoi* *ek*
 so PTC PTC sail_off:AOR.3PL doing_nothing:NOM.PL from
Milētou *hoi* *Pérsai*
 Miletus:GEN.SG ART.NOM.PL Persian:NOM.PL
 ‘Hence, the Persians sailed off from Miletus without success.’ (Arr.An.1.19.11)

If the tagset does not already contain the name of city of Miletus with its WHG identifier, then it must be manually added. To do so, the annotator looks for the city name on the WHG (Figure 12). If the WHG gives more than one result, then the annotator chooses the one referring to ‘Miletus’ in (6). In Figure 13, the first result refers to the city of Miletus mentioned in (6), as the second one is the name of a city in the USA.

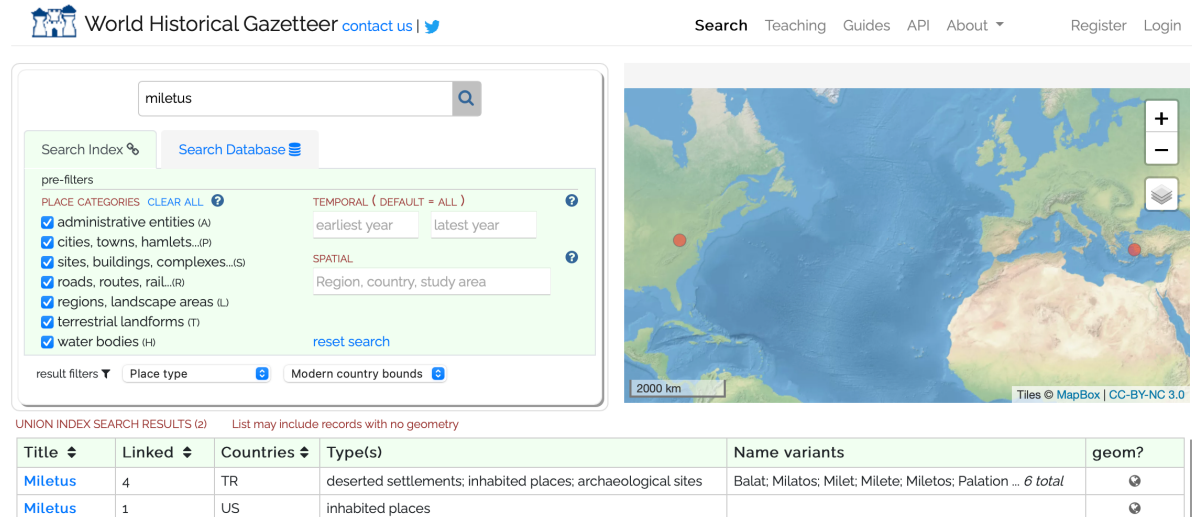


Figure 12. Results for ‘Miletus’ in the WHG.

The annotator might have to further disambiguate within a single WHG entry. Figure 13 shows the results retrieved clicking on the first ‘Miletus’ in Figure 12.

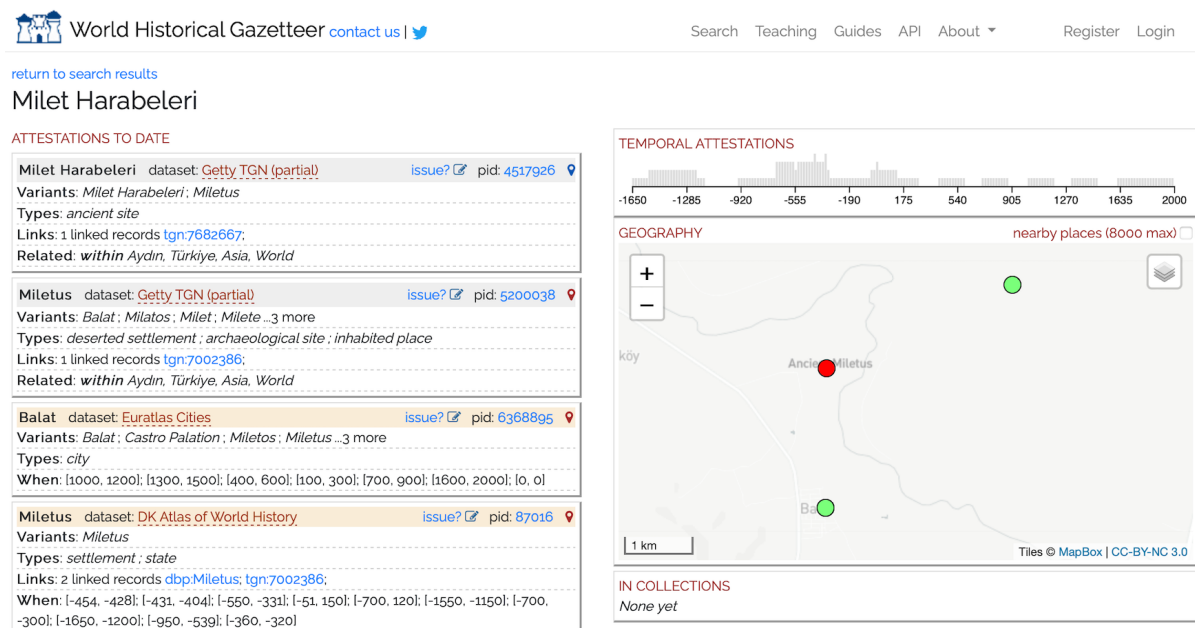


Figure 13. Places connected to ‘Miletus’ (Turkey) in the WHG.

For *Milétou* in (6) the annotator would copy the URL of the page, i.e. <https://whgazetteer.org/places/13745436/portal>, and add it to the tagset PLACES. Then, they will select this tag for the layer PLACE.

5.9 PREVERB

Preverbs are annotated with their corresponding prepositions. Forms such as Lat. *convenio* or *coeo* are assigned *cum* as PREVERB, rather than *con-/co-*. On the one hand, this facilitates lemmatization when a preverb may undergo assimilation (e.g. AGr. *epi-*, *ep-*, *eph-*, all under PREVERB *epi*). On the other hand, it allows for a more direct comparison with prepositions and their meanings.

5.10 SEMPREV

The semantics of the preverb must be annotated depending on the specific meaning that the preverb acquires in context. Consider the following examples from Farina and McGillivray (2022).

- (20) *omnúetō* *dè* *toi* *hórkon* [...] *mé* *pote*
 swear:IMPER.PRS.3SG PTC PTC oath:ACC.SG not ever
tēs *eunēs* ***epibēmenai*** *ēdē* *migēnai*
 ART:GEN.SG bed:GEN.SG go_upon:INF.AOR and mix:INF.AOR
 ‘May he swear [...] that he never went upon her bed, nor he had intercours
 with her.’ (Hom.II.19.176)
- (21) *sè* *d’* *hótan* *plēgē* *Diòs*
 2SG:ACC PTC whenever stroke:NOM.SG Zeus:GEN.SG
epibēi
 go_against:SUBJ.AOR.3SG
 ‘But whenever the stroke of Zeus attacks you.’ (S.Aj.137)
- (22) *hōs* *philosophéōn* *gēn* *pollēn*
 as love_knowledge:PTCP.PRS.NOM.SG earth:GEN.SG much:ACC.SG
theōriēs *heíneken* ***epelēluthas***
 viewing:GEN.SG for come_on:PF.2SG
 ‘As one who loves knowledge you have travelled all around the world for the
 sake of seeing it.’ (Hdt.1.30.2)

Sentences (20)-(22) contain different verbs preverbed with *epi-*. In each of these cases, the preverb acquires a different meaning, depending both on the verbal base and on the context. For (20), (21), and (22), the meanings of *epi-* are ‘upon’, ‘against’, and ‘over’, respectively.

The meaning of the preverb is annotated also when the verb does not have a literal meaning. This is because the annotation layer LITERAL MEANING disambiguates between a literal or non-literal context. Consider (8), quoted again, and (20) above.

- (8) *autàr* *Akhaíois* *apasíē* *tríllistos*
 but Achaeans:DAT.PL welcome:NOM.SG three_times:NOM.SG
epēluthē *nūx* *erebennē*
 come_upon:AOR.3SG night:NOM.SG dark:NOM.SG
 ‘But the dark night, gladly welcomed and three times prayed for, came upon the
 Achaeans.’ (Hom.II.8.488)

Epēluthe in (8) is used metaphorically. The meaning of *epi-* in (8) is still ‘upon’, as in (21) where the form *epibēmenai* has a literal meaning.

Preverb meanings should be considered carefully as some of them are similar (e.g. ‘together’ : ‘together with’, ‘to’ : ‘towards’). Consider the preverb *con-* (< *cum-*) in the following sentences.

- (23) *conveniunt* *frequentes* *prima* *luce*
gather:PRS.3PL numerous:NOM.PL first:ABL.SG light:ABL.SG
‘[The Latins] gather at daybreak in large numbers.’ (Liv.1.50)
- (24) [...] *nisi* *tu* *dudum* *hanc* *convenisti* [...]
 unless 2SG:NOM a_while_ago DEM:ACC.SG meet:PF.2SG
‘[I am very surprised, Sosia, that she knows that I received a golden bowl there],
unless you met her a while ago [and told her everything].’ (Plaut.Amph.768)

In (23), *cum-* generically means ‘together’, as the Latins gather altogether in one place. Conversely, *convenisti* in (24) specifies that the moving entity may have met (lit. ‘come together with’) Alcumena (*hanc*). Thus, in (24) *cum-* means ‘together with’.

A similar difference holds between the tags ‘to’ and ‘towards’, usually connected to preverbs inherently expressing the Goal. While ‘to’ indicates that the entity is reached at the end of the motion, ‘towards’ is used when the preverb specifies the direction of the motion, rather than its accomplishment.

When the verb acquires a deeper metaphorical meaning (on this matter, see Section 5.3), the meaning of the preverb in context may be unclear. In these cases, the verb is decomposed between its constituents (the preverb and the verb base) and the preverb is assigned the meaning connected to the metaphorical shift. Recall (9).

- (9) *trecentos* *sex* *perisse* *satis* *convenit*
three_hundred:ACC.PL six die:INF.PF enough agree:PRS.3SG
‘Three hundred and six men died, as it is generally agreed.’ (Liv.2.50)

Here, *convenit* has the metaphorical meaning ‘agree’, decomposable in ‘come (to a conclusion)’ (*venio*) ‘together’ (*cum-*). Therefore, the preverb *cum* is assigned the meaning ‘together’.

- (25) *his* *rebus* *atque* *auditionibus*
DEM.ABL.PL story:ABL.PL and hearsay:ABL.PL
permoti *de* *summis* *saepe*
move:PTCP.PF.PASS.NOM.PL about highest:ABL.PL often
rebus *consilia* *ineunt*
question:ABL.PL decision:ACC.PL go_into:PRS.3PL
‘Moved by these stories and hearsays, they often take decisions about questions
of the highest importance.’ (Caes.BG.4.5)

The phrase *consilia ineunt* in (25) means ‘take decisions’, derived from ‘go [with their minds] into decisions [thus taking them]’ (> ‘make up their mind and take decisions’). Therefore, despite the metaphorical meaning of the occurrence, *in-* is assigned ‘into’ as SEMPREV.

Finally, in some occurrences the preverb may display more than one meaning. For instance, Ernout and Meillet (2001 [1985]: 313) note that the preverbs *inter-* and *per-* convey «une idée de privation, de destruction, de mort», as in *intereo* and *pereo* ‘die, ruin, perish’. However, this

meaning does not exclude the prototypical meaning of the preverbs *inter-* and *per-* ‘across’, which activates the metaphor GO ACROSS SOMEONE’S LIFE is DIE. Let us focus on *per-* considering the two sentences below.

- (26) *reliqui* [...] *vi* *fluminis*
 other:NOM.PL force:ABL.SG river:GEN.SG
oppressi ***perierunt***
 overcome:PTCP.PF.PASS.NOM.PL die:PF.3PL
 ‘The rest [of the Germans] [...] died, overcome by the force of the stream.’
 (Caes.BG.4.15)

- (27) ***periit*** *itaque* *et* *tempus* *magna* *in*
 vanish:PF.3SG and_so also time:NOM.SG great:NOM.SG in
magnis *rebus* *iactura*
 important:ABL.PL operation:ABL.PL loss:ABL.SG
 ‘And so, time was wasted, a great loss in important operations.’ (Sen.Ira.3.21)

There is a difference in the meaning of *pereo* in (26) and (27). In (26), the verb token is attributed SEMPREV ‘across’ and ‘(idea of destruction/death)’, while in (27) it is only attributed ‘across’. Unlike *perierunt* in (26), there is no idea of death in (27). Metaphorically, time is thought to elapse, thus ‘going across’ the temporal axis (cf. It. *tra(n)s-correre* ‘run across’). When two meanings are annotated, the spatial meaning is always annotated before the non-spatial meaning. Moreover, annotating both meanings proves useful for a quantitative analysis of the preverb meaning itself.

5.11 SENTENCE

This layer selects both the verb token and the whole sentence containing it. The general rule is to try and select the whole sentence when possible, and to link it with the verb token using the layers SENTENCE and INCLUDES (see Section 4.1). In this annotation, a sentence includes all the tokens between the following punctuation marks: full stop(s), colon(s), semi-colon(s), exclamation mark(s), question mark(s). Punctuation marks themselves are never included in the annotation, apart from exclamation or question marks. The purpose of the layer SENTENCE is only to register the passage in which a verb token occurs. To exclude punctuation marks means to avoid passages terminating with a colon or a semi-colon – conversely, it may be useful to know whether a sentence has an exclamative or an interrogative sense.

However, some languages can display very long sentences, which can complicate the annotation, also from a technical point of view. Sometimes, INCEpTION does not recognize sentences when they are too long, and this can result in an empty annotation. This may also occur as sometimes INCEpTION interprets the full stops after the Latin *praenomina* (e.g. *A. Irtius*) as the end of the sentence, i.e. as a sentence boundary. To overcome the problem, the layers SENTENCE and INCLUDES allow crossing sentence boundaries by design. In other words, they ignore full stops at the end of the sentences. Nonetheless, technical issues may still occur, and the page may go blank if a SENTENCE annotation crosses sentence boundaries. In these cases, the annotator may decide – or is rather forced – to only select part of the sentence. Suppose that sentence (28) results in a problematic annotation.

- (28) *Hostes ubi et de expugnando oppido et de flumine transeundo spem se fefellisse intellexerunt neque nostros in locum iniquiorem progredi pugnandi causa viderunt,*

*atque ipsos res frumentaria deficere coepit, concilio convocato constituerunt optimum esse domum suam quemque reverti et, quorum in fines primum Romani exercitum introduxissent, ad eos defendendos undique **convenire**, ut potius in suis quam in alienis finibus decertarent et domesticis copiis rei frumentariae uterentur.* (Caes.BG.2.10)

In this case, the preverbed motion verb *convenire* only occurs in the second half of the sentence. If an annotator had to select only part of (28), they could exclude the first part of it, as it does not contain relevant context for *convenire*. The layer SENTENCE in (28) could then be annotated on *convenire* and on *concilio convocato [...] rei frumentariae uterentur*. This is because *convenire* and *reverti* are governed by *esse*, which, in turn, is governed by the verb of the main clause *constituerunt*.

Inverted commas are treated as follows. If the verb token occurs within a direct speech between two of the punctuation marks mentioned above, then the sentence will be treated as if it was not part of a direct speech. Conversely, if the verb token occurs within a direct speech broken up by extra information put between commas, dashes, or parentheses, then the layer SENTENCE will include all the tokens inside the direct speech up to one of the punctuation marks mentioned at the beginning of this section as well as the extra information outside the direct speech.

- (29) “*Ego vero*” *inquam* “*nihil impossibile arbitror, sed utcumque fata decreverint, ita cuncta mortalibus **provenire**: nam et mihi et tibi et cunctis hominibus multa usu venire mira et paene infecta, quae tamen ignaro relata fidem perdant.* [...]” (Apul.Met.1.19)

In (29), the direct speech is broken up by the verbal form *inquam* ‘I say’. As it is not possible to skip tokens within the annotation, the form *inquam* is ignored and the layer SENTENCE will include all the tokens from *Ego vero* to *provenire*. The column after the token *provenire* is the first strong punctuation mark found from the beginning of the direct speech.

5.12 SPATIALITY

The annotation of spatial relations (see Section 2.1.2) is straightforward if a verb possesses a literal meaning. However, when a verb does not possess a literal meaning, spatial relations are not always annotated, as explained below.

- (30) *quid* *te* *mutavit?* *quod* *reginam*
INT.NOM.SG 2SG.ACC change:PF.3SG that queen:ACC.SG
ineo?
have_sex:PRS.1SG
‘What changed you so much? The fact that I have sex with the queen?’
(Svet.Aug.2.69)

- (31) *tum* *sacerdotibus* *creandis* *animum* *adiecit,*
then priest:ABL.PL appoint:OBLG.ABL.PL mind:ACC.SG apply:PF.3SG
quamquam *ipse* *plurima* *sacra* *obibat*
although DET.NOM.SG many:ACC.PL rite:ACC.PL participate:IMPF.3SG
‘Then, he turned his attention to the appointment of priests, although he himself took part in many rites.’ (Liv.1.20)

- (32) *inde alios ineunt cursus*
 then other:ACC.PL perform:PRS.3PL movement:ACC.PL
alios-que recursus
 other:ACC.PL-and counter_movement:ACC.PL
 ‘Then they perform other movements and counter movements.’
 (Verg.Aen.5.583)

In the sentences above, *ineo* and *obeo* are used non-literally. However, a difference can be noticed between (30) and (31) on the one hand, and (32) on the other hand. Despite their non-literal sense (see Section 5.5), *ineo* in (30) and *obibat* in (31) still denote motion ‘into/towards’ (*in-*, *ob-*) someone (*reginam*) or something (*sacra*), expressed in the accusative case. This means that *reginam* in (30) and *sacra* in (31) constitute the figurative Goal (see Section 2.1.2) of *ineo* and *obibat*, respectively. As such, the layer SPATIALITY is annotated for both verbs. Conversely, the motion meaning of *ineunt* in (32) is more opaque, and the subject of the sentence is not said to go ‘into’ other movements. For this reason, the layer SPATIALITY is not annotated for a phrase such as *ineunt cursus aliosque recursus*.

5.13 SYNSET

In Section 5.3, I discussed how to annotate synsets of nouns. The annotation of verb synsets is simpler, because one verb is annotated with only one synset.

The annotation is conducted at a less deep level of granularity, with reference to the English WordNet. Consider the two examples below, respectively from Farina (2021: 153) and Farina (2021: 91).

- (33) *kaí pote tèn dēmarkhían apolipòn*
 and once ART.ACC.SG tribunate:ACC.SG leave:PTCP.AOR.NOM.SG
áphnō pròs Pompēion exépleusen eis
 suddenly toward Pompey:ACC.SG sail_off:AOR.3SG to
Surían, eîť ekeîthen epanêlthen
 Syria:ACC.SG then fom_there come_back:AOR.3SG
alogōteron
 more_irrational:ACC.SG
 ‘Once, suddenly leaving his tribunate, he sailed off to reach Pompey in Syria.
 Then he came back from there with even less reason.’ (Plut.Cic.26.7)
- (34) *epléomen dē eis tèn Aínon [...]*
 sail:IMPF.1PL PTC to ART.ACC.SG Aenus:ACC.SG
sunéplei dē tà te andrápoda
 sail_together:IMPF.3SG PTC ART.NOM.PL PTC slave:NOM.PL
hà édei autòn apolúsai
 REL.ACC.PL need:IMPF.3SG DEM.ACC.SG release:INF.AOR
 ‘We were sailing towards Aenus [...]. The slaves whom he was to release were
 also sailing.’ (Antiph.5.20)

In (33), the meaning of *ek-* ‘off from’ is reflected in the synset assigned to this verb, v#01318250 | steer away from shore, of ships. However, the English WordNet does not provide any synset for *sunéplei* ‘sail together’ in (34). In this case, the form *sunéplei* is assigned synset v#01260993 | travel by boat, which seems to hide *sun-*

. Despite this, other layers of annotation such as PREVERB and SEMPREV will make clear that the verb form is preverbed and that the meaning of *sun-* is ‘together with’.

A verb denoting the same idea may be assigned two different synsets, depending on the sentence, as occurs for *obeo* below.

- (35) *decedens* *Macedonia* [...] *mortem*
 depart:PTCP.PRS.NOM.SG *Macedonia*:ABL.SG death:ACC.SG
obiit *repentinam*
 meet:PF.3SG sudden:ACC.SG
 ‘While departing from Macedonia, [...] he died suddenly.’ (Svet.Aug.4.1)
- (36) *Iuliam* *primum* *Marcello* [...] *deinde ut*
Julia:ACC.SG first *Marcellus*:DAT.SG then as_soon_as
is *obit* *M.* *Agrippae*
 3SG.NOM die:PF.3SG *Marcus*:DAT.SG *Agrippa*:DAT.SG
nuptum *dedit*
 marry:SUP give:PF.3SG
 ‘He first gave Julia in marriage to Marcello, [...] then, after he died, to Marcus Agrippa.’ (Svet.Aug.63.1)

Although the resulting meaning of *obiit* in (35) and *obit* in (36) is connected to the semantic field of DEATH, there is a difference between the two sentences. In (35), *obiit* is constructed with the accusative *mortem*, but this does not happen in (36). In order to assign a synset to the occurrences above, this difference should be taken into account. Therefore, *obiit* in (35) is given synset v#01444459 undergo or suffer, as the figurative Ground *mortem* will specify what is undergone or suffered. On the contrary, the occurrence in (36) is assigned v#00250254 pass from physical life and lose all bodily attributes and functions necessary to sustain life. Of course, despite the difference between the two synsets, both verbs are connected to the meaning ‘die’. This is pointed out by means of the layer VERB CLASS (see Section 5.14).

The layer SYNSET is also annotated on nominal tokens representing the Figure/Ground or spatial relations (see Table 4). In the case of a multiword expression (see also Section 5.2), the layer SYNSET is annotated on the syntactic head, if present. The parts of the multiword expression are annotated with dependency relations following the section MWE (Multi-Word-Expression) of the Universal Dependency Relations¹⁸ and its related tags.

- (11) *Aulus* *Vitellius* *inferiorem* *Germaniam*
Aulus:NOM.SG *Vitellius*:NOM.SG lower:ACC.SG *Germany*:ACC.SG
ingressus *hiberna* *legionum*
 enter:PTCP.PF.NOM.SG winter_quarters:ACC.PL legion:GEN.PL
cum cura *adierat*
 with care:ABL.SG inspect:PPF.3SG
 ‘[On December 1st of the preceding year,] Aulus Vitellius, entering Lower Germany, had carefully inspected the winter quarters of the legions.’ (Tac.Hist.1.52)

¹⁸ <https://universaldependencies.org/u/dep/>.

In (11), for instance, the layer SYNSET, whose tag is n#00004123 a human being, is annotated on the token *Aulus*. *Aulus* and *Vitellius* are then related with the tag ‘flat’¹⁹ with the layer UD UNIVERSAL DEPENDENCIES (v. 2).

Annotation of pronouns, not included in the English WordNet, has already been discussed in Section 5.2. Adverbs are present in the English WordNet, so they are normally annotated.

5.14 VERB CLASS

Verb class is chosen depending on the meaning of the verb in context. In Section 2.2, I mentioned that classes of motion verbs in VerbNet can possess subclasses and sub-subclasses (e.g. ESCAPE-51.1-1, ESCAPE-51.1-1-1, ESCAPE-51.1-1-2, ESCAPE-51.1-1-3). These have not been considered in this annotation scheme for three reasons. First, in Section 2.2 I have already stressed that subclasses and sub-subclasses are ambiguous, and overlaps can be noticed. Second, VerbNet is modelled on English verbs, so these further categorizations may be specific to English and not fully reflected in other languages. Third, in the annotation, the layer SYNSET contains the specific meaning of the verb in context, so it is more useful to set the layer VERB CLASS as more generic to allow for more generic groupings during the analysis. Subclasses are mostly excluded also when the verb possesses a non-literal sense, i.e. the verb class assigned to the verb is not part of class 51 (see Section 2.2). For instance, Lat. *convenio* is assigned ESCAPE-51.1 when it means ‘come together’ (literal meaning), but it gets CORRESPOND-36.1 when it means ‘agree’ (non-literal meaning), as this is the verb class including English verbs such as *agree* and *decide* in VerbNet (see Section 2.2). In rare cases, it has become necessary to include subclasses. For instance, Lat. *congregior* can mean ‘fight’, as in (14). The English *fight* is associated with class BATTLE-36.4-1 (e.g. *battle*, *combat*, *duel*, *fight*), which is different from the simple BATTLE-36.4 (e.g. *argue*, *clash*, *collide*, *compete*, *dispute*, *quarrel*) as BATTLE-36.4-1 specifically describes physical violence (Figure 14). If *congregior* ‘fight’ was annotated with BATTLE-36.4, it would then be grouped together with verbs of quarrelling, which may lead to problematic results.

Figure 14. Verbs included in VerbNet under class BATTLE-36.4 (left) and BATTLE-36.4-1 (right).

Meanings can slightly differ even when a motion verb occurs in its literal sense (e.g. Lat. *abeo* ‘go away’ or ‘leave’). It is true that in VerbNet a verb can rarely appear in two classes, as for

¹⁹ <https://universaldependencies.org/u/dep/flat.html>.

leave, occurring both in ESCAPE-51.1 and in LEAVE-51.2. Nonetheless, a closer look at the classes reveals that *leave* is assigned a verb class depending on the constructions it occurs in, which is typical of the English language. When the place away from which the Figure moves is specified, *leave* gets class ESCAPE-51.1 (Figure 15). Otherwise, when *leave* occurs without any specification of the Source, it gets class LEAVE-51.2 (Figure 16).

VerbNet FrameNet PropBank OntoNotes

9 result(s) matching query (click to view):

- become-109.1-1-1
- escape-51.1-1-1
- fulfilling-13.4.1
- future_having-13.3
- keep-15.2
- leave-51.2-1
- render-29.90
- resign-10.11-2
- sustain-55.6-1

2 result(s) by Class ID:

ESCAPE-51.1-1-1

Parent Classes:

ESCAPE-51.1

This is not the parent class. For full class view click below

Full Class View

Member Verb Lemmas:

DEPART DISEMBARK ESCAPE EXIT FLEE

LEAVE VACATE

ROLES:

NP V NP

EXAMPLE:

The convict escaped the prison.

SHOW DEPENDENCY PARSE TREE

SYNTAX:

Theme VERB Initial_Location

Figure 15. Syntax of *leave*, with an example, when it is assigned class ESCAPE-51.1.

VerbNet FrameNet PropBank OntoNotes

9 result(s) matching query (click to view):

- become-109.1-1-1
- escape-51.1-1-1
- fulfilling-13.4.1
- future_having-13.3
- keep-15.2
- leave-51.2-1
- render-29.90
- resign-10.11-2
- sustain-55.6-1

leave-51.2

leave-51.2-1

Parent Classes:

LEAVE-51.2

This is not the parent class. For full class view click below

Full Class View

Member Verb Lemmas:

DESERT LEAVE

ROLES:

NP V

EXAMPLE:

He deserted.

SHOW DEPENDENCY PARSE TREE

SYNTAX:

Theme VERB

Figure 16. Syntax of *leave*, with an example, when it is assigned class LEAVE-51.2.

Similar occurrences in other languages must therefore be considered on a case-by-case basis also for this annotation.

Normally, to identify a verb class in VerbNet, the English translation of the Ancient Greek or Latin occurrence is searched in the resource. If more than one result match the query, each verb class is analyzed in detail before selecting the right one. Figure 17 shows the results for the English *die* in VerbNet.

Figure 17. Results for the query *die* in VerbNet (left). Verb class DIE-42.4-1 (right).

A careful look at all the seven verb classes on the left in Figure 17 reveals that when *die* means v#00250254 pass from physical life and lose all bodily attributes and functions necessary to sustain life, as for *obit* in (34), then it is part of verb class DIE-42.4-1. For instance, verb class DISFUNCTION-105.2.2 includes verbs such as *die*, *go down*, *go off*, which refer to objects that stop working (e.g. *My phone died*), so this class would not match the meaning of *obit* in (36).

Not all English verbs are registered in VerbNet.

- (37) *quam superba fuerit crudelitas*
how insolent:NOM.SG be:SUBJ.PF.3SG cruelty:NOM.SG
eius ad rem pertinent scire
3SG.GEN to topic:ACC.SG be_relevant:PRS.3SG know:INF.PRS
quamquam aberrare alicui possimus
although stray:INF.PRS someone:DAT.SG can:SUBJ.PRS.1PL
videri et in devium exire
seem:INF.PRS.PASS and to off_the_road:ACC.SG exit:INF.PRS
‘It is relevant to the topic to know how insolent his cruelty was, although it may seem that I am straying from the subject and making a digression.’ (Sen.Ira.3.19)

In (37), *exire* is annotated with synset v#00530582 turn aside esp. from the main subject of attention or course of argument in writing or speaking, connected to the English lemmas *digress*, *stray*, *divagate*, and *wander*. Some of these verbs, *digress* and *divagate*, are not included in VerbNet. *Stray* is included in class RUN-51.3.2, and *wander* is included both in RUN-51.3.2 and MEANDER-47.7 (see Section 2.2). However, a closer look at RUN-51.3.2 and MEANDER-47.7 reveals that none of these classes fits the meaning of *exire* in (37). Even though it is not assigned synset v#00530582 turn aside esp. from the main subject of attention or course of argument in writing or speaking in the English WordNet, *deviate* is surely another synonym of the four verbs mentioned above. In VerbNet, it is included class DIFFER-23.4. Therefore, the layer VERB CLASS for *exire* in (37) is annotated with verb class DIFFER-23.4.

Two verbs annotated with different synsets may be annotated with the same verb class. Recall (35) and (36), quoted again below.

- (35) *decedens* *Macedonia* [...] *mortem*
depart:PTCP.PRS.NOM.SG *Macedonia*:ABL.SG death:ACC.SG
obiit *repentinam*
meet:PF.3SG sudden:ACC.SG
‘While departing from Macedonia, [...] he died suddenly.’ (Svet.Aug.4.1)
- (36) *Iuliam* *primum* *Marcello* [...] *deinde ut*
Julia:ACC.SG first *Marcellus*:DAT.SG then as_soon_as
is *obit* *M.* *Agrippae*
3SG.NOM die:PF.3SG *Marcus*:DAT.SG *Agrippa*:DAT.SG
nuptum *dedit*
marry:SUP give:PF.3SG
‘He first gave Julia in marriage to Marcellus, [...] then, after he died, to Marcus Agrippa.’ (Svet.Aug.63.1)

Obiit mortem in (33) can be considered a multiword expression associated with the meaning ‘die’. Similarly, *obit* in (34) implies an accusative *mortem*. Despite the different synsets assigned to these occurrences (see Section 5.13), *obiit mortem* and *obit (mortem)* share the same verb class, which, in this case, is DIE-42.4-1. Attributing different synsets but the same verb class for these occurrences is useful to analyze the constructions that a verb may possess.

6. Other possible applications of this annotation scheme

The annotation scheme described in this work has been specifically created to analyze preverbed motion verbs in Ancient Greek and Latin (see Section 1). Nevertheless, it can be personalized depending on other scholars’ purposes. Layers can be removed if some of the parameters considered for this annotation (e.g. actionality, spatial relations, etc.) are not relevant for a specific study. In the sections below, I provide three examples of other applications of this annotation scheme, namely: (i) annotation of preverbed verbs of motion in (other) ancient or modern languages; (ii) annotation of preverbed verbs not expressing motion in ancient or modern languages; (iii) annotation of non-preverbed verbs (of motion or not) in ancient or modern languages.

From the technical point of view, the annotation scheme designed here has been used with the annotation tool INCEpTION. However, other textual annotation tools such as CATMA²⁰ (Petris 2010; Schüch 2010) are available and can replace INCEpTION.

6.1 Preverbed verbs of motion in (other) ancient or modern languages

Ancient Greek and Latin, the languages for which this annotation scheme has been designed, are not the only languages possessing preverbs. For this reason, a similar analysis may be replicated for other ancient or modern languages. Consider the examples below, from one ancient language (Vedic Sanskrit) and seven modern languages. The Sanskrit sentences come from the Rigveda, while the modern sentences are drawn from the TenTen Corpus Family²¹, available on SketchEngine²². I provide two sentences for each language, for a total of 16 sentences. In the first sentence, the verb possesses a literal meaning and the meaning of the

²⁰ <https://catma.de>.

²¹ <https://www.sketchengine.eu/documentation/tenten-corpora/>.

²² <https://www.sketchengine.eu>.

preverbed verb is basically compositional, while in the second sentence the meaning of the preverbed verb is non-literal.

Sanskrit, *pari-gam-* ‘go around, surround’ and *ni-gam-* ‘go down’:

- (38) *sá* *pravo!hṛ́n* *parigátyā* *dabhīter*
 3SG.NOM one_who_carries_off:ACC.PL surround:ABS Dabhīti:GEN.SG
viśvam *adhāg* *āyudham* *iddhé*
 all:ACC.SG burn:AOR.3SG weapon:ACC.SG kindle:PTCP.PF.PASS.LOC.SG
agnau
 fire:LOC.SG
 ‘Having surrounded the raiders against Dabhīti, he burned their every weapon in the kindled fire.’ (RV.2.15.4, trad. Jamison and Brereton 2014)
- (39) *kím* *bhrátāsad* *yád* *anāthám*
 INT.NOM.SG brother:NOM.SG-be:SBJ.PRS.3SG when helpless:NOM.SG
bhāvāti *kím* *u* *svāsā* *yán*
 be:SBJ.PRS.2SG INT.NOM.SG conversely sister:NOM.SG when
nīrtir *nigāchāt*
 dissolution:NOM.SG come_down:SBJ.PRS.3SG
 ‘What will “brother” (mean) when there will be no refuge. And what will “sister,” if Dissolution will come down?’ (RV.10.10.11, trad. Jamison and Brereton 2014)

English, *circum-navigate* and *over-come*:

- (40) *Saildrone is first to **circumnavigate** Antarctica, in search for carbon dioxide.*
 (enTenTen21)
- (41) *Latham is dedicated to working with clients to help them achieve their business goals and **overcome** legal challenges anywhere in the world.* (enTenTen21)

German, *aus-gehen* ‘go out’ (see also Figure 18 and Figure 19):

- (42) *wir können relaxen, Ausflüge machen,*
 we can:PRS.1PL relax:INF.PRS excursion:ACC.PL do:INF.PRS
*Sport treiben, abends **ausgehen** und*
 sport:ACC.SG do:INF.PRS in_the_evening go_out:INF.PRS and
gut essen
 well eat:INF.PRS
 ‘We can relax, go on excursions, do sports, go out in the evening, and eat good food.’ (deTenTen20)
- (43) *die große Wirkung, die von diesen*
 ART.NOM.SG big:NOM.SG effect:NOM.SG which from DEM.DAT.PL
*Bildern **ausgeht**, rührt von*
 painting:DAT.PL derive:PRS.3SG move:PRS.3SG from
dem Bildthema her
 ART.DAT.SG image_topic:DAT.SG here

‘The great impact that emanates from these images is due to the image's subject matter.’ (deTenTen20)

Italian, *ac-correre* ‘come to’ and *in-correre* ‘incur’:

- (44) *Frate Elia, a quella notizia, accorse in*
 brother Elia to DEM news come_to:PST.3SG in
fretta da lontano
 hurry from afar
 ‘Brother Elia, receiving that piece of news, came in a hurry from afar.’
 (itTenTen20)

- (45) *anche Elia incorse dunque nella scomunica*
 also Elia incur:PST.3SG therefore into_the ban
di Gregorio IX
 of Gregory 9th
 ‘Therefore, Elia too incurred in the ban of Gregory 9th.’ (itTenTen20)

French, *ac-courir* ‘run to’ and *in-courir* ‘incur’:

- (46) *un vieil émigré à cheval accourt et*
 a old emigrant on horse run_to:PRS.3SG and
se fait expliquer la cause du tumulte
 REFL make:PRS.3SG explain:INF.PRS the cause of_the tumult
 ‘An old emigrant on horseback runs up and has the cause of the tumult explained to him.’ (frTenTen20)

- (47) *il sera tenu livrer un autre*
 he be:FUT.3SG required deliver:INF.PRS a other
personnage honneste en son lieu, sur peine d’
 person honest in his place under penalty of
incourir l’amende à appliquer comme
 incur:INF.PRS the fine to appliquer as
dessus
 above
 ‘He will be required to deliver another honest person in his place, under the penalty of incurring the fine to be applied as above.’ (frTenTen20)

Spanish, *circun-navegar* ‘sail around’ (see also Figure 20) and *sobre-volar* ‘overfly’ (see also Figure 21):

- (48) *estos "nómadas del mar", que incluso llegaron a*
 these nomads of_the sea REL even reach:PST.3PL to
circunnavegar el continente africano [...], también
 sail_around:INF the continent African also
ocultaron [...] sus rutas marítimas y sus descubrimientos
 hide:PST.3PL their routes maritime and their discoveries
geográficos
 geographical

‘These “nomads of the sea”, who could even sail around the African continent [...], also hid [...] their maritime routs and their geographical discoveries.’ (esTenTen18)

- (49) *hay una pregunta ética que sobrevuela*
 exist:PRS.3SG a question ethical REL hover_over:PRS.3SG
a todas estas anteriores cuestiones y que es
 to all these previous issues and REL be:PRS.3SG
con lo que comenzábamos este texto
 with the REL begin:IMPF.1PL this text
 ‘There is an ethical question that hovers over all these previous issues, and that is what we began this text with.’ (esTenTen18)

Portuguese, *circum-navegar* ‘sail around’ and *sobre-voar* ‘overfly’:

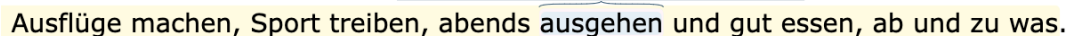
- (50) *ele realmente desejava circumnavegar o globo*
 he really want:IMPF.3SG sail_around:INF the
globo ou queria simplesmente chegar às
 world or want:IMPF.3SG simply reach:INF to_the
ilhas Molucas e regressar?
 islands Maluku and come_back:INF
 ‘Did he really want to sail around the world, or did he just wanted to reach the Maluku islands and then come back?’ (ptTenTen20)
- (51) *por estas linhas sobrevoamos afetos,*
 through these lines soar_over:PST.1PL affections
inquietações, memórias, traições, destinos de pessoas
 concerns memories betrayals destinies of people
banais expostas a circunstâncias únicas
 ordinary exposed to circumstances unique
 ‘Through these lines, we soar over affections, concerns, memories, betrayals, destinies of ordinary people exposed to unique circumstances.’ (ptTenTen20)

(Modern) Greek, *ane-vénō* ‘climb, go up’:

- (52) *strívis dexiá kai anevénis tà skaliá*
 turn:PRS.2SG right and go_up:PRS.2SG ART.ACC.PL stair:ACC.PL
 ‘You turn right and go up the stairs.’ (elTenTen19)
- (53) *to Fevrouário tou 2001 anévike*
 ART.ACC.SG February:ACC.SG ART.GEN.SG 2001 come_out:AOR.3SG
sto théatro Emprós i parástasi
 in_the theatre:ACC.SG Embros ART.NOM.SG play:NOM.SG
To imerológio tis ámmou
 ART.NOM.SG diary:NOM.SG ART.GEN.SG sand:GEN.SG
 ‘In February 2001, the play The Diary of the Sand was staged at the Embros Theater.’ (elTenTen19)

The annotation scheme produced for this research and applied to Ancient Greek and Latin, could also be adopted for sentences (38)-(53) without any substantial modification. Figure 18,

Figure 18. Example of annotation for Ger. *ausgehen* in (42).



The diagram illustrates the syntactic structure of the sentence: "an artistic composition made by applying paints to a surface -Literal meaning n#00102821 influencing strongly von + DAT Bild von Wirkung ausgehen".

Top Level Structure:

- acl** (adpositional clause) is the main modifier of the noun phrase.
- nsubj** (nominal subject) is the main subject of the sentence.
- det** (determiner) is the main determiner of the noun phrase.

Noun Phrase (NP) Structure:

- NOUN** (Includes) is the main noun.
- (Space)** is the space between the noun and the modifier.
- (Sentence)** is the sentence structure.
- (Sentence)** is the sentence structure.

Modifier Structure:

- State** (present stem)
- APPEAR-48.1.1** (out)
- aus** (out)
- DET** (an artistic composition made by applying paints to a surface)
- ADP** (von)
- von#00381912** (proceed or issue forth, as from a source)
- Literal meaning** (n#00102821 influencing strongly)
- von + DAT** (von)
- (MorFea...)** (Bild)

Verb Phrase (VP) Structure:

- VERB** (ausgehen)
- von** (von)
- Wirkung** (Wirkung)
- der** (der)

Diagram Details:

- The diagram uses color-coded boxes to represent different syntactic categories: yellow for NOUN, green for DET, blue for ADP, and pink for VERB.
- Arrows indicate the grammatical relations between the words and the phrases they form.
- The diagram shows the hierarchical structure of the sentence, from the main subject and predicate down to the individual words and their grammatical functions.

The diagram illustrates a semantic network for the sentence: "one of the large landmasses of the earth, the second largest continent, located south of Europe and bordered to the west by the South Atlantic and to the east by the India Ocean travel around, either by plane or ship".

Top Level (Sentence):

- (Sentence) - amod - (Sentence)
- (Sentence) - (Space) - (Sentence)

Second Level (Verb Phrase):

- v#01304328 travel around, either by plane or ship - VERB - det -
- DET - NOUN - ADJ -

Third Level (Accomplishment):

- Accomplishment
- present stem
- ESCAPE-51.1
- around
- circum
- +Literal meaning

Fourth Level (MorFea):

- n#06835944 one of the large landmasses of the earth, n#06840536 the second largest continent, located south of Europe and bordered to the west by the South Atlantic and to the east by the India Ocean
- n#00004123 a human being
- (MorFea...)
- circunnavegar

Bottom Level (URL and Continente):

- https://whgazetteer.org/places/14043174/portal
- continente

46

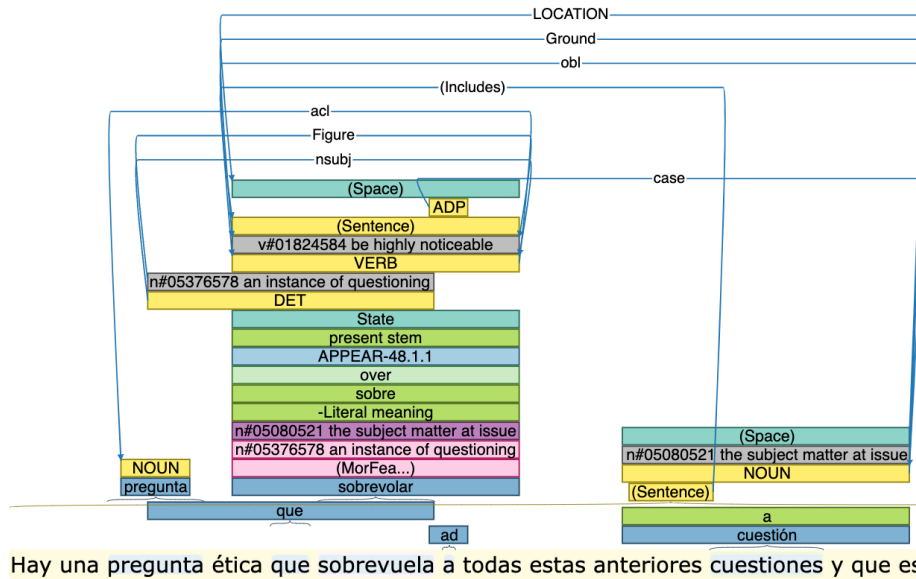


Figure 21. Example of annotation for Sp. *sobrevolar* in (49).

No substantial differences can be noticed between the annotations above on the one hand, and the annotations in Figure 7 and Figure 9 on the other hand²³. All the layers mentioned in Section 3 still hold for modern languages. Therefore, the annotation guidelines described in this document (cf. especially Section 4 and Section 5) may also be used in these cases.

Tagsets (Section 4.2) for language-specific layers (EXPRESSED BY, PREVERB) change depending on the language. For instance, the tagset SR EXPRESSION, used to annotate the layer EXPRESSED BY (see Section 5.2) includes new prepositions and cases for languages with cases such as German, but it contains only prepositions for languages without cases such as Spanish. In (43), Ger. *ausgehen* is constructed with a complement introduced by the preposition *von*, only governing the dative. For this reason, the noun *Bildern* in Figure 19 is annotated with ‘*von* + DAT’ under the layer EXPRESSED BY. Conversely, Sp. *sobrevuela* in (49) is constructed with the complement *a [todas estas anteriores] cuestiones*. The noun *cuestiones* in Figure 21 is annotated with ‘*a*’ under the layer EXPRESSED BY as Spanish is a fusional language without cases. The layer EXPRESSED BY and its tagset SR EXPRESSION could even be removed from the annotation for languages without cases as prepositions would be easily retrieved through syntactic annotation, thus using layer DEPENDENCY. The same holds for languages in which prepositions only govern one case (e.g. Modern Greek, in which prepositions only govern the accusative).

6.2 Preverbed verbs not expressing motion in ancient or modern languages

To better grasp the meaning of preverbs and the role they have in changing the semantics of the verb base, other verb classes can be considered. Compared to Section 6.1, in this case not all annotation layers can still be relevant.

The annotation scheme presented here is designed for a case study on motion verbs, so parameters such as Figure/Ground (see Section 2.1.1), spatial relations (see Section 2.1.2), and geographical places (see Section 2.6) are annotated. Preverbed stative verbs are perhaps the

²³ The colors of the layers have changed as these sentences are part of a different “project” (see INCEption guidelines for this terminology, <https://inception-project.github.io/releases/30.2/docs/user-guide.html>). The tool randomly assigns colors to labels in different projects.

most similar to motion verbs. The three parameters mentioned above, reflected in layers FIGURE SYNSET, GROUND SYNSET, PARTICIPANTS, PLACE, SPACE, and SPATIALITY, may still be considered for an analysis of this verb class, and the difference lies in that no motion is implied in the event expressed by a stative verb. The layers SPACE and SPATIALITY, however, would need to be omitted. Consider, for instance, Lat. *absum* and its literal and non-literal uses, as in (54) and (55), respectively.

- (54) *itaque* *exiguum* *spatii* *vallum* *a*
 and_so short:NOM.SG distance:GEN.SG rampart:NOM.SG from
vallo ***aberat***
 rampart:ABL.SG be_away:IMPF.3SG
 ‘Therefore, there was only a short distance between one rampart and the other.’
 (Liv.22.24)
- (55) *neque* *ulla* *re* *longius* ***absumus***
 and_not any:ABL.SG thing:ABL.SG further differ:PRS.1PL
a *natura* *ferarum*
 from nature:ABL.SG beast:GEN.PL
 ‘And in no respect are we farther removed from the nature of wild beasts.’
 (Cic.Off.1.16.50)

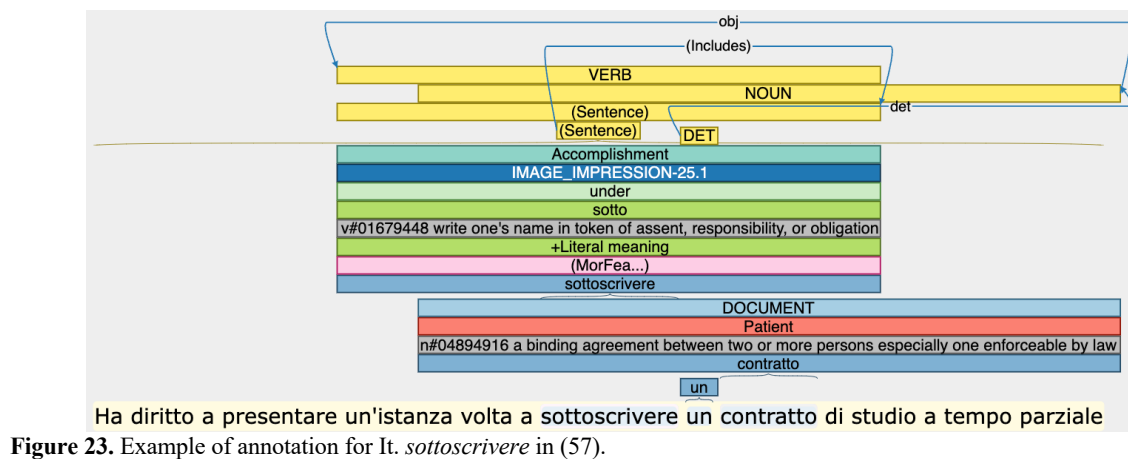
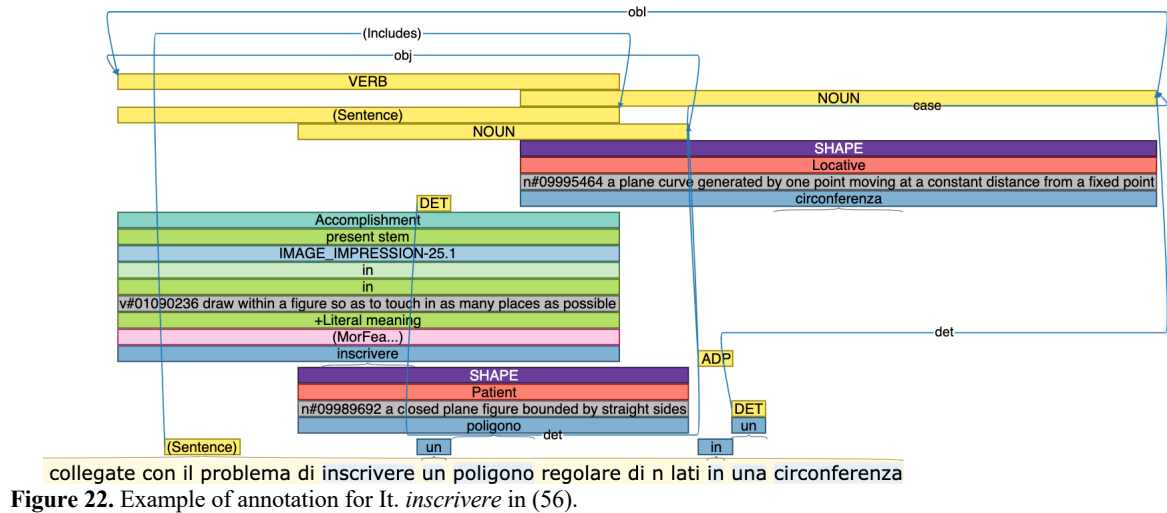
As the layers FIGURE SYNSET, GROUND SYNSET, PARTICIPANTS, and PLACE can be utilized for both motion and stative verbs, they should probably be omitted for other types of verbs. Suppose we want to annotate the meanings of the preverbed forms of It. *scrivere* ‘write’.

- (56) *ha* *diritto a* *presentare* *un’* *istanza* *volta*
 have:PRS.3SG right to present:INF a request turned
a ***sottoscrivere*** *un* *contratto* *di* *studio a* *tempo*
 to sign:INF a contract of study to time
parziale
 part
 ‘He has the right to present a request to sign a part-time study contract.’
 (itTenTen20)
- (57) [...] *il* *problema* *di* ***inscrivere*** *un* *poligono* *in*
 the problem of inscribe:INF a polygon in
una *circonferenza*
 a circle
 ‘The problem of inscribing a polygon in a circle.’ (itTenTen20)

To maintain something similar to the layers FIGURE SYNSET, GROUND SYNSET, and PARTICIPANTS, semantic roles of the verb arguments (Grimshaw 1990, but also Tesnière 1959 and Lazard 1994), as well as their semantic type (Jezek 2016; 2018; Farina 2020 for a non-computational case study on three preverbed verbs) could be annotated in such cases. Semantic roles of English verbs have already been annotated within the Berkeley FrameNet²⁴ project (Baker et al. 1998; Gildea and Jurafsky 2002; Ruppenhofer et al. 2010), then expanded to other languages (Chinese, Danish, German, Japanese, Korean, Polish, Brazilian Portuguese, Spanish, Swedish). To include semantic roles and types in the study of preverbed verbs would allow a

²⁴ <http://framenet.icsi.berkeley.edu>.

better grasp not only of the semantics of the preverb, but also of the semantics of the resulting preverbed verb and of the semantic roles and types themselves. Figure 22 and Figure 23 provide an example of annotation for the Italian *sottoscrivere* and *inscrivere* based on (56) and (57), respectively.



Compared to Figure 7 and Figure 9, for the annotation of (56) and (57) the following layers have been excluded: EXPRESSED BY, FIGURE SYNSET, GROUND SYNSET, PARTICIPANTS, PLACE, SPACE, SPATIALITY, and VERB STEM. This means that the annotated layers on the verbal forms *inscrivere* and *sottoscrivere* are (from the bottom to the top): LEMMA, MORPHOLOGICAL FEATURES, LITERAL MEANING, SYNSET, PREVERB, SEMPREV, VERB CLASS, ACTIONALITY, SENTENCE, PART OF SPEECH. Nominal tokens (*poligono* and *circonferenza* in Figure 22, and *contratto* in Figure 23) have been annotated with the following layers (from the bottom to the top): LEMMA, SYNSET, SEMANTIC ROLE (marked in red), SEMANTIC TYPE (marked in purple in Figure 22 and in light blue in Figure 23), PART OF SPEECH. Layers SEMANTIC ROLE and SEMANTIC TYPE have been designed for this annotation and are not present in Table 4.

The annotations provided above are just examples of the way in which the annotation scheme described in this document can be used and adapted. Other parameters can be added or removed, depending on the researchers' research questions.

6.3 Non-preverbed verbs (of motion or not) in ancient or modern languages

Verb semantics is extremely interesting also when considering the difference in meaning between a verb base and its preverbed forms (Farina 2020; Farina et al. 2023a). Moreover, preverbed forms are also useful in casting new light on the meanings of the verbal bases themselves (Farina et al. 2023a). In the case of motion verbs, other parameters such as deixis may be added within this annotation scheme, if the analysis focuses on deictic oppositions (Nuti 2016; Farina et al. 2023a). Studies on verb semantics also benefit from analyses on possible similarities/differences between (non-)preverbed synonymic pairs of verbs. As already suggested in Section 6.2, these types of research questions may need to consider other linguistic parameters such as verb arguments and their semantic and/or thematic roles, ultimately reaching more sophisticated and complex semantic theories such as the Qualia structure (Pustejovsky and Jezek 2012).

7. Conclusions and future work

This document has provided guidelines for the linguistic annotation of preverbed verbs of motion, also including non-linguistic parameters such as the annotation of places (see Sections 2.6 and 5.8). However, I have also shown that this annotation scheme is extremely versatile, as it can be easily adapted to other languages, either ancient or modern. Furthermore, layers can be added or removed on a case-by-case basis, depending on research questions. Sections 6.1–6.3 show a possible complete semantic analysis of verbs – and their preverbs (Sections 6.1 and 6.2) –, encompassing different linguistic parameters that are usually not considered jointly.

Performing a complete semantic analysis of preverbed motion verbs following these guidelines would have several benefits. First, researchers across different languages, projects, and studies could use a common approach to identify and study the same linguistic feature. Standardization and consistency are crucial to enhance reproducibility for comprehensive cross-linguistics analyses around preverbs and their relationship with motion verbs, which have not been done yet, especially on a large scale. This can lead to insights into universal patterns or language-specific phenomena that we are not yet aware of. Second, in a computational perspective, a detailed annotation scheme is essential for training and evaluating machine learning algorithms, and these guidelines can be instrumental in developing and refining NLP tools and models, starting from manual annotation. Third, this document could also serve as a valuable resource for educational purposes. These guidelines can be used in linguistics courses focused on corpus linguistics, syntax, and semantics to teach students about the complexities of annotating and analyzing linguistic data, with a specific focus on preverbed motion verbs.

8. References

- AlMousa, Mohannad, Rachid Benlamri, and Richard Khoury. 2022. A novel word sense disambiguation approach using WordNet knowledge graph. In *Computer Speech & Language*, 74, 101337.
- Andreevskaia, Alina, and Sabine Bergler. 2006. Semantic Tag Extraction from WordNet Glosses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, 413–416, Genoa, Italy. European Language Resources Association (ELRA).
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational linguistics*, 86–90.

- Barker, Elton, Leif Isaksen, Nick Rabinowitz, Stefan Bouzarovski, and Chris Pelling. 2013. In *Leeds International Classical Studies*, 9(1), 1–24.
- Barker, Elton, Rainer Simon, Leif Isaksen, and Pau de Soto Cañamares. 2016a. Peripleo: A Tool for Exploring Heterogeneous Data through the Dimensions of Space and Time. In *Code4Lib*, 31.
- Barker, Elton, Rainer Simon, Leif Isaksen, and Pau de Soto Cañamares. 2016b. The Pleiades Gazetteer and the Pelagios Project. In Merrick L. Berman, Ruth Mostern, and Humphrey Southall (eds.), *Placing Names: Enriching and Integrating Gazetteers*, 97–109. Bloomington: Indiana University Press.
- Basu, Chumki, Laura Dietz, and Christiane Fellbaum. 2018. WordNetContext: Information Retrieval-friendly Access to WordNet Senses. In *Joint Proceedings of the First International Workshop on Professional Search (ProfS2018); the Second Workshop on Knowledge Graphs and Semantics for Text Retrieval, Analysis, and Understanding (KG4IR); and the International Workshop on Data Search (DATA:SEARCH'18)*, Ann Arbor, Michigan (USA), 63–64.
- Bertinetto, Pier Marco. 1986. *Tempo, Aspetto e Azione Nel Verbo Italiano*. Firenze: Accademia della Crusca.
- Biagetti, Erica, Chiara Zanchi, and William Michael Short. 2021. Toward the Creation of WordNets for Ancient Indo- European Languages. In *Proceedings of the 11th Global Wordnet Conference, University of South Africa (UNISA)* Global WordNet Association: 258–266.
- Bizzoni, Yuri, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, and Gregory Crane. 2014. The Making of Ancient Greek WordNet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1140–1147, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Bizzoni, Yuri, Riccardo Del Gratta, Federico Boschetti, Marianne Reboul. 2015. Enhancing the Accuracy of Ancient Greek WordNet by Multilingual. Distributional Semantics. In *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it*, 47–50.
- Black, Jeremy. 1999. *DK Atlas of World History*. London: Dorling Kindersley Limited.
- Booij, Geert, and Ans Van Kemenade. 2003. “Preverbs: An Introduction.” In *Yearbook of Morphology 2003*, G. Booij and J. van Marle (eds.), 1–12. Dordrecht: Kluwer.
- Brinton, Laurel J. 1985. “Verb Particles in English: Aspect or Aktionsart?” *Studia Linguistica* 39 (2): 157–68.
- Chakravarthi, Bharathi Raja, Michael Arcan, and John P. McCrae. 2019. WordNet Gloss Translation for Under-resourced Languages using Multilingual Neural Machine Translation. In *Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation*, 1–7, Dublin, Ireland. European Association for Machine Translation.
- Chantraine, Pierre. 1999 [1968]. *Dictionnaire étymologique de la langue grecque. Histoire des mots*. Paris: Klincksieck.
- Ciravegna, Fabio, Bernardo Magnini, Emanuele Pianta, and Carlo Strapparava. 1994. Multilingual Lexical Knowledge Bases: Applied WordNet Prospects. Talk presented at the *International Workshop on "The Future of the Dictionary"*, Grenoble (France), October 1994.
- Cole, Michael J., and Jacek Gwizdka. 2008. Tagging semantics: investigations with wordnet. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries (JCDL '08)*, 446. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/1378889.1378991>

- Collins, Allan M., and M. Ross Quillian. 1969. Retrieval time from semantic memory. In *Journal of Verbal Learning and Verbal Behavior*, 8, 240–247.
- Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge Univ. Press.
- Cuzzolin, Pierluigi. 1995. A proposito di *sub vos placo* e della grammaticalizzazione delle adposizioni. In *Archivio Glottologico Italiano*, 80: 122–143.
- Dowty, David R. 1986. “The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics?” *Linguistics and Philosophy* 9 (1): 37–61.
- Edmonds, Philip, and Adam Kilgarriff. 2002. Introduction to the special issue on evaluating word sense disambiguation systems. In *Natural Language Engineering*, 8(4), 279–291.
- Elliott, Tom. 2021. The Pleiadic Gaze: Looking at Archaeology from the Perspective of a Digital Gazetteer. In *Classical Archaeology in the Digital Age – The AIAC Presidential Panel*. <https://doi.org/10.11588/PROPYLAEUM.708.C10612>.
- Ernout, Alfred, and Alfred Meillet. 2001 [1985]. *Dictionnaire étymologique de la langue latine. Histoire des mots*. Paris: Klincksieck.
- Farina, Andrea. 2020. Struttura argomentale e significato: *ridere, navigare, scrivere* e i prefissati *deridere, circumnavigare* e *descrivere*. University of Pavia: student essay.
- Farina, Andrea. 2021. *Aquamotion Verbs in Ancient Greek. A Study on Plēō and Its Compounds*. University of Pavia: MA Thesis.
- Farina, Andrea, and Barbara McGillivray. 2022. “The Ancient Greek preverb *epi-*. A case study on three motion verbs”. Talk presented at *Zadar Linguistic Forum 2022. Predicate: from polysemy to arguments*. April 6–8, 2022. Zadar (Croatia).
- Farina, Andrea, Barbara McGillivray, and Stephen Colvin. 2023a. Motion and prevervation in Latin. The relation between the preverbs *ab-* and *ad-* and the verbal bases *eo* and *venio*. Talk presented at *International Colloquium on Latin Linguistics*. June 19–23, 2023, Prague (Czechia).
- Farina, Andrea, Barbara McGillivray, and William Michael Short. 2023b. WordNets and corpora. A case study on the semantic field SEA in Latin and Ancient Greek. Talk presented at *International Colloquium of Historical Linguistics*. September 4–8, 2023, Heidelberg (Germany).
- Fedriani, Chiara, Irene De Felice, and William Michael Short. 2020. The Digital *Lexicon Translativum Latinum*: Theoretical and Methodological Issues. In *Atti del IX Convegno Annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD)*, 106–112.
- Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press.
- Filip, Hana. 1999. *Aspect, Eventuality Types, and Nominal Reference*. New York: Garland.
- Franzini, Greta, Andrea Peverelli, Paolo Ruffolo, Marco Carlo Passarotti, Helena Sanna, Edoardo Signoroni, Viviana Ventura, and Federica Zampedri. Nunc Est Aestimandum: Towards an Evaluation of the Latin WordNet. In *Proceedings of the Sixth Italian Conference on Computational Linguistics*, 1–8, Bari (Italy).
- Gildea, Daniel, and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. In *Computational Linguistics*, 28(3): 245–288.
- Grimshaw, Jane. 1990. *Argument Structure*. Cambridge (MA): MIT Press.
- Haynes, Sherwood. 2001. Semantic Tagging Using WordNet Examples. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, 79–82, Toulouse (France). Association for Computational Linguistics.
- Isaksen, Leif, Elton Barker, Eric C. Kansa, and Kate Byrne. 2012. *GAP: A NeoGeo Approach to Classical Resources*. In *Leonardo*, 45(1), 1–2.

- Jamison, Stephanie W., and Joel P. Brereton. 2014. *The Rigveda. The Earliest Religious Poetry of India. Volume 1*. Oxford: Oxford University Press.
- Jezek, Elisabetta. 2016. *The Lexicon: An Introduction*. Oxford: Oxford University Press.
- Jezek, Elisabetta. 2018. Partecipanti impliciti nella struttura argomentale dei verbi. In Sara Dallabrida, and Patrizia Cordin (Eds.), *La grammatica delle valenze*, 55–71. Firenze: Franco Cesati Editore.
- Johanson, Lars. 2000. “Viewpoint Operators in European Languages.” In *Tense and Aspect in the Languages of Europe*, Ö. Dahl (ed.), 27–187. Berlin: De Gruyter.
- Kahn, Rebecca, Leif Isaksen, Elton Barker, Rainer Simon, Pau de Soto, and Valeria Vitale. 2021. Pelagios – Connecting Histories of Place. Part II: From Community to Association. In *International Journal of Humanities and Arts Computing*, 15(1-2), 85–100.
- Kenny, Anthony. 2003. *Action, Emotion, and Will*. London /New York: Routledge.
- Kilgarriff, Adam. 2002. English lexical sample task description. In *Proceedings SENSEVAL-2: The Second International Workshop on Evaluating Word Sense Disambiguation Systems*, 17–20. Toulouse, France.
- Lakoff, George. 1987. *Women, Fire and Dangerous Things. What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.
- Lazard, Gilbert. 1994. *L’actance*. Paris: Presses Universitaires de France.
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- López Moreda, Santiago. 1998. Interferencias semántico-sintácticas entre preverbios y preposiciones. In *Estudios de Lingüística latina. Actas del IX Coloquio Internacional de Lingüística Latina, Universidad Autónoma de Madrid*, 953–969.
- Luraghi, Silvia. 2003. *On the Meaning of Prepositions and Cases*. Amsterdam/Philadelphia: John Benjamins.
- Manning, Patrick. 2015. World-Historical Gazetteer Research Report. In *Journal of World-Historical Information*, 2-3(1).
- Manning, Patrick, and Ruth Mostern. 2015. World-Historical Gazetteer. White paper report at the University of Pittsburgh. <https://doi.org/10.17613/M6GQ00Q>.
- Marongiu, Paola, Fahad Khan, and Barbara McGillivray. 2023. *Tools and resources for diachronic lexical semantic analyses: state of the art*. Zenodo. <https://doi.org/10.5281/zenodo.8103978>
- McCarthy, Diana. 2006. Relating WordNet Senses for Word Sense Disambiguation. In *Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*, 17–24.
- McGillivray, Barbara. 2013. Latin preverbs and argument structure. New insights from new methods. In Elly van Gelderen, Jóhanna Barðdal, and Michela Cennamo, *Argument Structure in Flux: The Naples-Capri Papers*, 119–134. Oxford: Oxford University Press.
- Meillet, Antoine, and Joseph Vendryes. 1963. *Traité de grammaire comparée des langues classiques*. Paris: Librairie Ancienne Honoré Champion.
- Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to wordnet: an on-line lexical database. In *International Journal of Lexicography*, 3(4):235–244.
- Miller, George A., and Christiane Fellbaum. 2007. WordNet then and now. In *Language Resources and Evaluation*, 41:209–214.
- Minozzi, Stefano. 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, 707–716.

- Minozzi, Stefano. 2017. Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'Information Retrieval. In Paolo Mastrandrea (ed.), *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, 123–133. Venezia: Edizioni Ca' Foscari.
- Moens, Marc, and Mark Steedman. 1988. "Temporal Ontology and Temporal Reference." *Computational Linguistics* 14 (2): 15–28.
- Mostern, Ruth. 2017. World-Historical Gazetteer Research Report. In *Journal of World-Historical Information*, 4(1).
- Navigli, Roberto. 2006. Reducing the Granularity of a Computational Lexicon via an Automatic Mapping to a Coarse-Grained Sense Inventory. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, 841–844. European Language Resources Association (ELRA).
- Ng, Hwee T., Chung Y. Lim, and Shou K. Foo. 1999. A case study on the inter-annotator agreement for word sense disambiguation. In *Proceedings of ACL Workshop: Standardizing Lexical Resources*. College Park, Maryland.
- Ng, Hwee T., Bin Wang, and Yee S. Chan. 2003. Exploiting Parallel Texts for Word Sense Disambiguation: An Empirical Study. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 455–462. Association for Computational Linguistics, Sapporo (Japan).
- Ngo, Vuong M., Tru H. Chao, and Tuan M. V. Le. 2018. WordNet-Based Information Retrieval Using Common Hypernyms and Combined Features. In *Proceedings of the 5th International Conference on Intelligent Computing and Information Systems (ICICIS-2011)*, Cairo (Egypt).
- Nuti, Andrea. 2016. A Matter of Perspective: Aspect, Deixis, and Textual Exploitation in the Prototype Semantics of *Eo* and *Venio*. In W. M. Short (ed.), *Studies in Language Companion Series*, 15–56.. Amsterdam: John Benjamins.
- Padave, Shailesh. 2014. *Incorporating WordNet in an Information Retrieval System*. San José State University: MA thesis.
- Palmer, Martha, Olga Babko-Malaya, and Hoa Trang Dang. 2004. Different Sense Granularities for Different Applications. In *Proceedings of the 2nd International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004) at HTL-NAACL 2004*, 49–56.
- Passarotti, Marco, and Mambrini, Francesco. 2021. Linking Latin: Interoperable Lexical Resources in the LiLa Project. In Erica Biagetti, Chiara Zanchi, and Silvia Luraghi (eds.), *Building New Resources for Historical Linguistics*, 103–124. Pavia: Pavia University Press.
- Petris, Marco. 2010. *CATMA*. Software-Wiki. <https://software.fandom.com/de/wiki/CATMA>.
- Pianta, Emanuele, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: Developing an aligned multilingual database. In *Proceedings of the 1st International Global WordNet Conference*, 293–302. Mysore (India).
- Purves, Alex C. 2002. *Telling Space: Topography, Time and Narrative from Homer to Xenophon*. University of Pennsylvania: PhD Thesis.
- Purves, Alex C. 2010. *Space and Time in Ancient Greek Narrative*. Cambridge: Cambridge University Press.
- Pustejovsky, James. 1991. "The Syntax of Event Structure." *Cognition* 41 (1–3): 47–81.
- Pustejovsky, James and Elisabetta Jezek. 2012. Introducing Qualia Structure. In *Generative Lexicon Theory: a Guide*. Oxford: Oxford University Press.
- Romm, James S. 1994. *The Edges of the Earth in Ancient Thought: Geography, Exploration, and Fiction*. Princeton: Princeton University Press.

- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2010. *FrameNet II: Extended Theory and Practice* (revised ed.). Berkeley, CA: International Computer Science Institute.
- Snyder, Benjamin, and Martha Palmer. 2004. The English all-words task. In *Proceedings of ACL 2004 SENSEVAL-3 Workshop*. Barcelona, Spain.
- Salm, Khan M., Mumit Khan, and Testuro Nishino. 2009. Example based English-Bengali machine translation using wordnet. In *Conference Papers (Centre for Research on Bangla Language Processing)*.
- Sasse, Hans-Jürgen. 2002. “Recent Activity in the Theory of Aspect: Accomplishments, Achievements, or Just Non-Progressive State?” *Linguistic Typology*, 199–271.
- Sausa, Eleonora. 2012. Toward an Ancient Greek WordNet. In *Workshop on WordNet and SketchEngine*, Pavia (Italy).
- Schüch, Lena. 2010. *CATMA User Manual for CATMA 3.0*. http://catma.de/webarchive/catma-4.0/webfm_send/7.
- Siemiński, Andrzej. 2011. WordNet Based Word Sense Disambiguation. In Piotr Jędrzejowicz, Ngoc Thanh Nguyen, and Kiem Hoang (eds.), *Computational Collective Intelligence. Technologies and Applications. ICCCI 2011. Lecture Notes in Computer Science*, vol. 6923. Springer, Berlin, Heidelberg, 405–414. https://doi.org/10.1007/978-3-642-23938-0_41.
- Simon, Rainer, Elton Barker, and Leif Isaksen. 2012. Exploring Pelagios: A Visual Browser for Geo-Tagged Datasets. In *International Workshop on Supporting Users’ Exploration of Digital Libraries*, 1–6, Paphos (Cyprus).
- Simon, Rainer, Leif Isaksen, Elton Barker, and Pau de Soto Cañamares. 2016. The Pleiades Gazetteer and the Pelagios Project. In Merrick Lex Berman, Ruth Mostern, and Humphrey Southall (Eds.), *Placing Names: Enriching and Integrating Gazetteers*, 97–109. Bloomington: Indiana University Press. <https://doi.org/10.2307/j.ctt2005zq7>.
- Snow, Rion, Sushant Prakash, Daniel Jurafsky, and Andrew Y. Ng. 2007. Learning to Merge Word Senses. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 1005–1014, Prague (Czechia).
- Swart, Henriëtte de. 1998. “Aspect Shift and Coercion.” *Natural Language & Linguistic Theory* 16 (2): 347–285.
- Talmy, Leonard. 1975. “Figure and Ground in Complex Sentences.” In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 1:419–30.
- Talmy, Leonard. 1983. “How Language Structures Space.” In *Spatial Orientation. Theory, Research, and Application*, Herbert L. Pick and Linda P. Acredolo (eds.), 225–82. New York: Plenum Press.
- Talmy, Leonard. 1985. “Lexicalization Patterns: Semantic Structure in Lexical Forms.” In *Language Typology and Syntactic Description, III: Grammatical Categories and the Lexicon*, Timothy Shopen (ed.), 57–149. Cambridge: CUP.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics. Vol. 1: Concept Structuring Systems*. Cambridge (MA): Mit Press.
- Tất Thắng, Nguyễn. 2013. Figure and Ground in passive sentences. *Dalat University Journal of Science*, 7, 111-123.
- Tatevosov, Sergej. 2002. “The Parameter of Actionality.” *Linguistic Typology* 6 (3): 317–401.
- Tesnière, Lucien. 1959. *Éléments de Syntaxe Structurale*. Paris: Klincksieck.
- Vendler, Zeno. 1957. “Verbs and Times.” *The Philosophical Review* 66 (2): 143.
- Vitale, Valeria, Pau de Soto, Rainer Simon, Elton Barker, Leif Isaksen, and Rebecca Kahn. 2021. Pelagios – Connecting Histories of Place. Part I: Methods and Tools. In *International Journal of Humanities and Arts Computing*, 15(1-2), 5–32.

- Zanchi, Chiara. 2019. "Multiple Preverbs in Ancient Indo-European Languages: A Comparative Study on Vedic, Homeric Greek, Old Church Slavic, and Old Irish." *Sprachvergleich*, Band 2. Tübingen: Narr Francke Attempto.
- Zucchi, Sandro. 1998. Aspect Shift. In *Events and Grammar*, S. Rothstein (ed.), 349–70. Dordrecht: Springer.