

The making of Ancient Greek WordNet

Yuri Bizzoni*, Federico Boschetti[◇], Riccardo Del Gratta[◇],
Harry Diakoff[‡], Monica Monachini[◇], Gregory Crane*

[◇]CNR-ILC “A. Zampolli”, Pisa - Italy, Via Moruzzi 1
{firstname.lastname}@ilc.cnr.it

*Università degli Studi di Pisa, Pisa - Italy, Via Santa Maria, 53
yuri.bizzoni@gmail.com

[‡]Alpheios Project
<http://alpheios.net>, harry.diakoff@gmail.com

*Perseus Digital Library Project, Department of Classics - Eaton Hall 134C
Tufts University - Medford MA, 02155 USA
<http://www.perseus.tufts.edu>, gregory.crane@tufts.edu

Abstract

This paper describes the process of creation and review of a new lexico-semantic resource for the classical studies: AncientGreekWordNet. The candidate sets of synonyms (synsets) are extracted from Greek-English dictionaries, on the assumption that Greek words translated by the same English word or phrase have a high probability of being synonyms or at least semantically closely related. The process of validation and the web interface developed to edit and query the resource are described in detail. The lexical coverage of Ancient Greek WordNet is illustrated and the accuracy is evaluated. Finally, scenarios for exploiting the resource are discussed.

Keywords: Ancient Greek, Multilingualism, Classical Philology

1. Overview

This paper describes a work in progress in its early stage for the creation of the Ancient Greek WordNet (AGWN) and its linkage to other WordNets (WNs).

Rich and deep lexical and grammatical tradition, coupled with the changes in meaning due to modern developments makes the creation of lexical resources for classical languages quite a complicated task. The literature has plenty of examples of attempts at endowing ancient languages with Wordnets. We only mention here (Kulkarni et al., 2010), which describes the construction of a Sanskrit WordNet, built using the expansion approach.

The need for a WordNet of ancient Greek, in particular after the creation of Minozzi’s WordNet for Latin (Minozzi, 2009; McGillivray, 2010), has become increasingly evident as other digital resources for the Classics have appeared.

In the field of linguistic and literary analysis, text processing techniques offer the possibility to investigate the vocabulary of a large amount of classical texts, as explained in (Bamman and Crane, 2008; Bamman and Crane, 2011).¹ Most of philological digital instruments, such as concordance tools, would profit from the availability of a WordNet to allow for the extension of keyword-based searches to semantically related lemmas. But the main motivation for having computational resources for classic languages is the possibility to perform automatic analysis. An Ancient

Greek WordNet provides a lexical resource that can be used as a tool for applying many computational linguistic techniques such as methods for Word Sense Disambiguation, Word Similarity, etc. and enhancing the performance of Information retrieval.

In the long-standing field of computational lexicography, the development of computational lexical resources has gone on for the last thirty years at least, producing large scale resources that are now commonly used in tandem with tools for the automatic extraction of lexical items and relations to foster e.g. the production of thesauri.

Natural Language Processing tools and Lexical Extraction tools are an aid both for enhancing access to electronic texts and for supporting the analysis of texts. The annotation of texts at different levels of linguistic analysis allows for more refined search instruments to be offered to the reader and for a more refined set of features to be used with algorithms of computational text analysis and classification.

The integration of WordNets with Treebanks is nowadays recognized as one of the most compelling needs for both research and educational purposes.

If we open the lens, making language resources and technologies available and easily usable to scholars of digital humanities, we will help to overcome the present fragmentation within the discipline and will lead to new research frontiers thus promoting a methodological change.

This paper presents the first results of efforts by an international collaboration among the Institute of Computational Linguistics “Antonio Zampolli” in Pisa, the Perseus Project in Boston, the Open Philology Project in Leipzig and the Alpheios Project in New York, to address this need, which

¹The Dynamic Lexicon is “an NEH-funded project to automatically create bilingual dictionaries (Greek/English and Latin/English) using parallel texts [...] along with the syntactic data encoded in treebanks.”: <http://nlp.perseus.tufts.edu/lexicon>

the late Emanuele Pianta (Bruno Kessler Foundation), in collaboration with the University of Pavia (Sausa, 2012), had also been planning to address before his tragic death.

2. Methodology

2.1. Creation

The initial automatic construction of the AGWN was achieved using Greek-English digitized lexicons provided by the Perseus Project: the LSJ (Liddell et al., 1940), the Middle-Liddell (Liddell and Scott, 1889) and Autenrieth's *Homeric Lexicon* (Autenrieth, 1891), to extract Greek-English word pairs.

The Middle-Liddell proved to be more consistently structured than the other two and thus provided the most reliable parsing of English synonyms with the least "noise."

The Greek word of the extracted bilingual pair was linked to every synset in the Princeton WordNet (PWN) (Fellbaum, 1998) in which the English member of the pair appeared. This has been a common approach in the creation of a number of modern WNs (Sagot and Fišer, 2011), because of the great richness and detail of the PWN, although it raises both problems common to all uses of English as a pivot language and issues arising from the attempt to map concepts across cultures that are so remote from one another (Vossen, 1996).

2.2. Validation of a Sample

In this early phase of the work, we had two main goals: the identification of the principal sources of error in the automatic extraction and the evaluation of a relevant sample of synsets, comparable with traditional studies of synonymy.

"For this reason, the sample to be manually corrected and validated was composed by the largest synsets (often due to spurious synonymy related to very generic terms) and by synsets including at least one word from J.H. Schmidt's *Synonymik der griechischen Sprache* (Schmidt, 1876).

Before the manual correction, misaligned polysemy had been reduced by filtering out English meanings expressly identified as colloquial in the PWN as well as anachronistic MultiWordNet (MWN) domains, such as those related to modern science and technology and any other domains consisting primarily of recent neologisms (e.g. aviation, telecommunication, football, etc.). The MWN Domains² resource was used for this purpose (Magnini et al., 2001).

For example, the English word "plane" assumes in different domains specific meanings, a geographic entity (*rhykánē*, ῥυκάνη), a kind of tree (*plátanos*, πλάτανος), but also "aircraft". The latter meaning can be filtered out by the identification of the anachronistic domain "aviation".

However, misaligned polysemy, although reduced and highly filtered, remains the main source of error, nor did anachronisms completely disappear.

The manual review was performed by an Italian native language speaker, a graduate student in Digital Humanities, with a BA in Classics and an intermediate level in English. He participated in a pilot-project aimed at evaluating the localization in the Italian language of part of the protocol

established in the Perseus Project for the creation, correction and validation of resources for the study of Classics.

During the manual validation process, the student had to rank each word of the Ancient Greek synsets with a score between 0 (semantically not related) and 2 (synonym with the other words in the synset). The possibility to assign an intermediate score 1 to the translation was made necessary by linguistic and cultural problems.

When a word was considered inadequate for inclusion in a synset but still semantically related to it, the type of the specific semantic relation (e.g. hyponymy, meronymy, etc.) was marked by the student, in order to insert, in a second stage of the work, the inspected word into the correct synset (whether existent or newly created).

For example, the synset that contains the English word "bird", glossed by "warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings", attracts a high number of Greek lemmas. Many of these, however, do not refer to the general concept "bird" (*órnis*, ὄρνις, in Greek), but to some species of birds: *kérkēris* (κέρκηρις) is an aquatic bird, *drepanís* (δρεπανίς) is defined by LSJ "a bird, so called from the shape of its wings, probably the Alpine swift, *Cypselus melba*", etc. All these terms are marked with the suitable semantic relation, in order to be placed in hyponymic synsets in a second stage of the work.

If the list of Greek words in a specific synset is ill-formed, either because the gloss for the concept is inadequate to express the correct meaning or because all the related English words are inadequate to translate the Ancient Greek term, the student logically isolated³ the synset.

The deactivation of a synset is a challenging choice, especially when the synset expresses a modern concept that is an evolution of the ancient concept. In this case the student marks a near-equivalent-like relation⁴ between the modern and the ancient concept that must be inserted and glossed. The first stage of the validation process was mainly focused on deletion of inadequate words from synsets and isolation of inadequate synsets from the semantic net, but the second stage will take into account also the necessity to add new words to the synsets or new synsets to the semantic net.

2.3. Linkage to other WNs

English / Greek bilingual resources are available under free licenses and frequently used not only by native speakers but also by the entire community of scholars and students of Classics. But we believe that scholars and students belonging to the international community of classicists and who are involved in crowdsourcing efforts to extend available resources for the study of classical languages are strongly assisted in their work if they have at their disposition bilingual resources in various languages, and especially in their native languages. These resources are aimed at understanding the nuances of meaning expressed by terms belonging to different synsets.

³The synset is temporary removed (deactivated) from the net because the links are considered inadequate at the moment of the analysis, but they can be activated again after further investigation.

⁴The correct statement is "near to the concept expressed by a definition that needs adjustments"

²The version 3.2 is available at <http://wndomains.fbk.eu>

Accordingly, the manual review during the pilot-project was facilitated by consultation with several thesauri of Classical Greek and bilingual dictionaries (in particular Greek / English and Greek / Italian, even if this latter unfortunately not available under free licence) and by alignment of the AGWN not only with PWN but also with Italian WordNet (IWN), developed at the Institute of Computational Linguistic in Pisa (Roventini et al., 2003), the Italian section of the MWN, developed at Bruno Kessler Foundation and a Latin WordNet automatically produced by the Alpheios Project linked to Minozzi’s Latin WordNet.

2.4. Comparison to Latin WN

The comparison with Latin WN is interesting, because it is an available resource manually checked by a classicist, linked to PWN, which strengthens the evaluation of a synset: if a Greek term is associated with an apparently inadequate synset but manually checked Latin terms are associated with the same synset, additional attention is needed before rejecting the relation.

However, Latin WordNet seems to be less restrictive about anachronisms that we decided to reject, managing them by an “extended polysemy” policy: existing Latin words that acquired modern senses are extended also to those senses, in agreement with the *modus operandi* of the *Lexicon Recentis Latinitatis* (Egger, 2004). Accordingly, for instance, “cliens” can have the meaning of “any computer that is hooked up to a computer network” and, thus, find a place as a hyponym of “machina”, which, in turn, can also signify a “4-wheeled motor vehicle; usually propelled by an internal combustion engine”, hyponym of “vehiculum”. In Latin WordNet (LWN) the term “accitus” means ‘an order to appear in person at a given place and time’, ‘a writ issued by authority of law’ and also ‘a telephone connection’.

Latin WordNet seems in other words to address the problem of polysemy from a modernist perspective: the persistence of at least one common sense between two words justifies the inclusion of the ancient lemma in its modern counterpart’s synset.

3. The Ancient Greek WordNet GUI

In this section we describe the graphical user interface to query and edit Ancient Greek WN, which has been developed according to the needs of manual checking and correction and validation. The interface is available at the URL:

http://www.languagelibrary.eu/new_ewnui

Figure 1 illustrates the search page, which is divided in two sections: the top one is devoted to the personal profile of the user, including his/her activities, while the bottom is devoted to the search panel with options for the source and target languages.

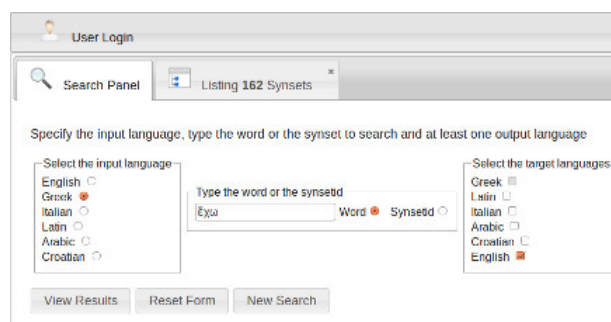


Figure 1: The search page of the GUI

3.1. Structure of the model

The data model behind the GUI has been designed to manage WN-like data structures. So far it deals with the following WNs:

- Princeton WordNet;
- Italian WordNet;
- Croatian WordNet;
- Arabic WordNet;
- Latin WordNet;
- Ancient Greek WordNet.

but it has been planned to have pluggable components, so a new WN can be inserted into the model and added to the search panel.

The main feature of the model is the possibility to have a set of mapped concepts in different languages. According to section 2.3., thus starting from English-Greek concepts and words, each WN is mapped onto PWN so that the English is *the pivot language* and each concept is mapped to the corresponding English concept. For example, the Croatian “107543288 (n) snažan osjećaj naklonosti, strastvene privrženosti; duhovna i/ili spolna privlačnost jednog bića prema drugome [*ljubav*]”, is mapped onto the English: “107543288 (n) a strong positive emotion of regard and affection [*love*]”, which correspond to the Ancient Greek synset that contains the terms to be accepted or rejected *agápē* (ἀγάπη), *philótēs* (φιλότις), *érōs* (ἔρως), etc.

3.2. Search Panel

The search panel in Figure 1 contains two main zones: an input language area where users can select the source language and a list of output languages from which users can select the desired target language(s).

Once the users have selected the input language, they start typing the word to search in the textbox, where an auto-complete mechanism suggests the list of words (which are contained in the input WN, selected according to the input language) that starts with the characters they have been typed.

Figure 2 describes how the English pivot language interface between input (Greek) and target (Latin and Italian) to show the target concepts that are mapped to the input synset(s).⁵

⁵As explained in section 3.3., the input synsets which are presented to the users contain the input word.

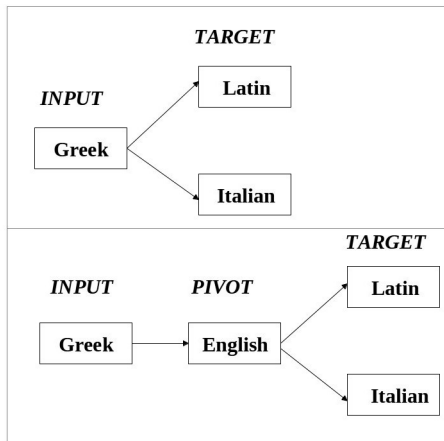


Figure 2: English as pivot between input and target languages

3.3. From words to synsets and related operations

Once the “View Results” button is pressed, the list of synsets that contain the typed word is presented to the users. The entire list is presented to the users, so that they have to click on the appropriate synset and navigate to the corresponding target concept.

- 2001100141806 (V)
the act of inspecting or verifying [ἐπισχεθεῖν, κολάζω, ἀντεφοράω, ἰσχανάω, ἴσχω, ...]
- 2001100167446 (V) (chess)
a direct attack on an opponent’s king [ἐπισχεθεῖν, κολάζω, ἀντεφοράω, ἰσχανάω, ἴσχω, ...]
- 2001100318735 (V)
the act of carrying something [ἐμβαστάζω, ἐποχετεύω, πορίζω, μυριαγωγέω, κυέω, ...]
-

3.4. Edit the data

While everyone can search and browse the data, the editing features of the GUI are available only to logged users. Users can log on the system through their profile tab in the search page, see Figure 1.

For example, logged users can edit the specific meanings of *échō* (ἔχω): “the act of carrying something”.

Once clicked on the identifier of the synset, 2001100318735, logged users can do the following activities:

Browse This tab of the GUI allows logged users to edit the part of speech as well as the gloss of the concept (for example a typical activity consists in modifying or translating the gloss in the target language) see Figure 3.

Managing words in the synsets As anticipated in section 2.2., logged users can validate words within the

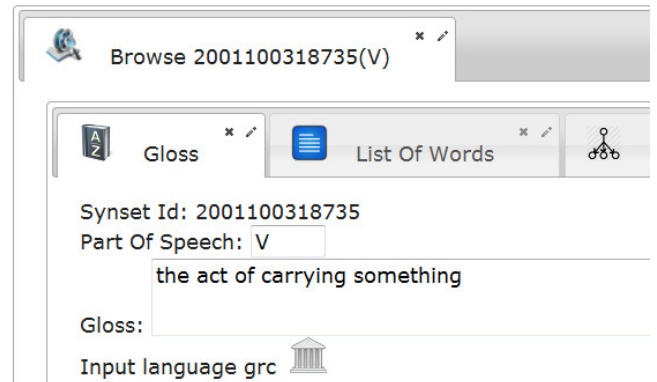


Figure 3: POS and gloss of the selected concept

synset: they can choose 0 to exclude the word from the synset, 2 to say the word is fully pertinent to the synset, 1 to imply that further investigations are needed. In addition they can add up to 5 words to the synset: see Figure 4.

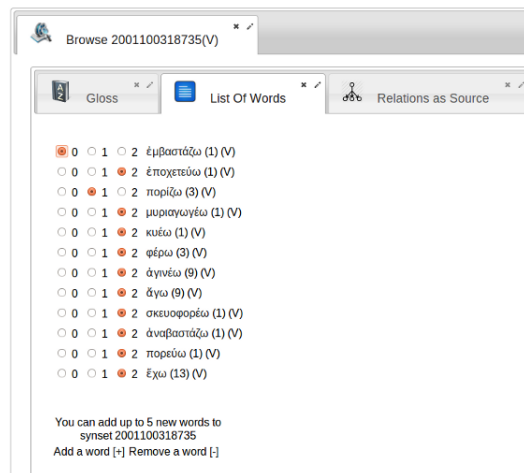


Figure 4: Validate and add words to the synset

Managing Relations The GUI displays the relations that involve the selected synset and the synsets to which it is connected. For example, the synset “100318735, the act of carrying something” has the *hypernym* “100315986, the act of moving something from one location to another”. Logged users have the possibility of validating/adding the relations, in the same way to the word validation/addition, see figure 5. This functionality allows the users to modify the conceptual network, changing the original graph, which is inherited from the PWN structure.

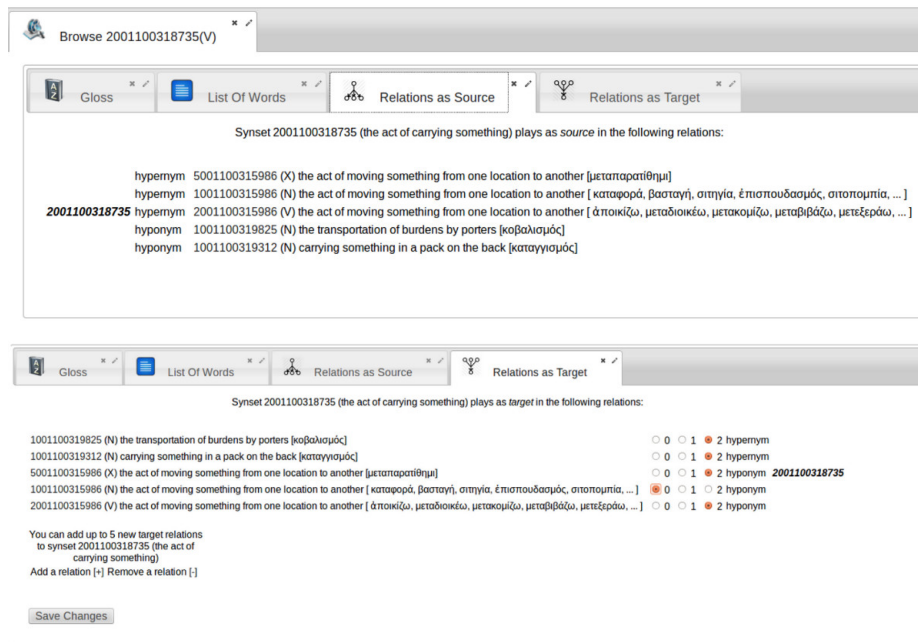


Figure 5: Validate and add relations

4. Results

4.1. Evaluation of the corrections applied to the Sample

1013 out of 33910 synsets have been checked, in order to evaluate the performance of the system and in order to start correcting errors.

84 out of 1013 synsets (8.3%) have been deactivated, because of an erroneous association to modern concepts alien to antiquity, such as “a series of linked atoms (generally in an organic molecule)” automatically associated to *hórmos* (ὄρμος), *hàlysis* (ἄλυσις), *sýsphigma* (σύσφιγμα), *psállion* (ψάλλιον), *hormathós* (ὄρμαθός), due to the polysemy of the English translation “chain”.

14 out of 1013 synsets (1.4%) have been marked as “near to the concept expressed by a definition that needs adjustments”. These cases are interesting because they clearly demonstrate the gap between *Sinn* (sense) and *Bedeutung* (denotation), to use Frege’s categories.

For instance, the concept associated to *gê* (γῆ) and *gâia* (γαῖα), is defined as “the third planet from the Sun; the planet we live on; [...]”

The denotation of *γαῖα* is clearly our planet, but the sense that defines the concept is related to the scientific paradigm (Ptolemaic or Copernican).

The 1013 checked synsets contain 6457 senses, i.e. possibly repeated words with a specific different meaning.

4.2. Comparison with Schmidt’s *Synonymik*

A comparison with Schmidt’s is not straightforward, because Schmidt’s groupments are more similar to semantic fields than synsets. However, a couple of lists of terms are worthy of note:

- a the main Greek synonyms to indicate the sea are present both in the AGWN and in Schmidt (which adds also the co-hyponym *ōkeanós*, ὠκεανός):

thálassa (θάλασσα), *háls* (ἅλς), *pélagos* (πέλαγος), *róntos* (ρόντος)

- b the Greek synonyms to express the concept “moving quickly and lightly” (in English: agile, nimble, quick and spry) can be divided in a subset shared by both Schmidt and the Ancient Greek WordNet and two complimentary subsets. Common synonyms are: *aiólos* (αἰόλος), *aipsērós* (αἰψηρός), *thoós* (θόος), *krairpnós* (κραιπνός), *laipsērós* (λαιψηρός), *tachýs* (ταχύς), *ōkýs* (ὠκύς).

Terms present only in Schmidt with this meaning are: *argós* (ἀργός), *baliós* (βαλιός), *elaphrós* (ἐλαφρός), *karpalímos* (καρπαλίμος), *oksyís* (ὄξύς), *panáiolos* (παναίολος), *sobarós* (σοβαρός), *trochalós* (τροχαλός), plus three terms related to the quickness of the foot: *argírous* (ἀργίπους), *podōkēs* (ποδώκης), *ōkýrous* (ὠκύπους).

Finally, it is worthy to note that there are nine relevant terms that are present only in the Ancient Greek WN: *euagēs* (εὐαγής), *eukínētos* (εὐκίνητος), *dierós* (διερός), *íksalos* (ἰξάλος), *kōuphos* (κοῦφος), *ksouthós* (ξουθός), *otrērós* (ὄτρηρός), *polýskarthmos* (πολύσκαρθμος), *spoudaios* (σπουδαῖος).

4.3. Coverage

The total Greek lexicon counts up to 120k different lemmas, while the number of distinct lemmas contained in the AGWN consists of 35k lemmas, with a coverage of 28%. This is mainly due to the fact that translations constituted by single words or phrases present in the PWN are used to link the WNs, whereas translations with mismatching phrases currently are unparsed. For example, *tráchouros* (τράχουρος) is associated to the correct PWN synset, which correspond to “horse mackerel”, but *óchanon* (ὄχανον) is discarded, because the PWN does not contain

the phrase “bar across of the shield”. These cases will be managed in a further stage of the work.

The coverage of the AGWN on the Homeric lexicon is 69% cf. table 1, due also to the fact that Autentrieth’ *Homeric Dictionary* has been used for the construction of the resource.

Part of Speech	% of lexicon	AGWN coverage
Nouns (N)	32%	76%
Adjectives (A)	27%	59%
Verbs (V)	33%	72%
Adverbs (R)	80%	61%
N+A+V+R	100%	69%

Table 1: AGWN coverage of the Homeric lexicon

4.4. Propagated Polysemy

We have compared the Ancient Greek WordNet with the Princeton WordNet in order to verify how polysemy in the two resources is propagated, considering also that PWN covers 148k different lemmas and AGWN only 35k.

Lemma	# of senses
break	59
make	49
give	44
take	42
cut	41

Table 2: Top five English polysemous verbs

Lemma	# of senses
<i>échō</i> (ἔχω)	162
<i>kóptō</i> (κόπτω)	125
<i>téuchō</i> (τεύχω)	105
<i>tektáinomai</i> (τεκταίνομαι)	104
<i>ágō</i> (ἄγω)	91

Table 3: Significant polysemous verbs extracted from the top ten polysemous verbs

The 5 most polysemous Greek words spread into many corresponding English words: for instance ἔχω spreads into 171 different English words among which some of the 5 words in table 2 are contained. This holds also for the other most polysemous Greek words, due to the inheritance of the senses from English: see table 4.

A similar analysis can be carried out also for other parts of speech.

4.5. Semantic relations

Currently semantic relations are inherited by the PWN, although it is possible to modify them through the graphical interface. The study of semantic relations can have fruitful didactic applications, especially if focused on the lexicon of specific authors. For instance, the AGWN terms can be filtered by the Homeric lexicon, in order to identify the parts of the ship (*náus*, ναῦς) in Homer. *histíon* (ἱστίον)

Greek Verb	English Verb
<i>échō</i> (ἔχω)	carry (40) hold (36) have (19) take (6) make (3) give (1) break (1)
<i>kóptō</i> (κόπτω)	cut (41) strike (21) take (1)
<i>téuchō</i> (τεύχω)	make (49) work (27) give (1) take (1)
<i>tektáinomai</i> (τεκταίνομαι)	make (49) work (27) give (1) take (1)
<i>ágō</i> (ἄγω)	carry (40) lead (15) bring (11) take (1)

Table 4: Propagated polysemy

and *spêiron* (σπεῖρον): “sail”, *prýmna* (πρύμνα): “stern”, *oiéion* (οἰήιον) and *póus* (πόυς): “steering-paddle”, *ántlos* (ἄντλος): “hold of a ship” are correctly retrieved, even if the precision and the recall need improvements.

5. Discussion

5.1. Missing synsets in AGWN

Among the limitations of the methods used in the present approach should be noted the obvious inability to identify concepts present in ancient Greek that have no counterpart in the Princeton WordNet, which was initiated in the mid 1980’s with American English. But even within Greek, the use of general lexicons without specifying the authors and time periods represented by the entries must necessarily create many associations that were in fact valid only for specific time periods, or even authors. Ideally a WordNet should reflect the semantic relationships of a specific text or collection of texts, linking the lemmas in each synset with the lexemes in the texts where they have that particular synset’s meaning. Some of this procedure can be partially automated using collocations and the synonymous relations identified within the WordNet itself, but obviously there will be a need for manual curation if

this level of precision is to be attempted at present.

The exercise of creating the AGWN also provided many interesting opportunities to compare distinctive characteristics of the two languages. English is often surprisingly polysemic in a quite different way from ancient Greek. Provided the right context, English makes it easy to detect the part of speech of a word without morphological clues: exactly what ancient Greek doesn't allow, vividly illustrating a major difference between ancient Greek and modern English: the highly polysynthetic nature of the former and the relatively isolating character of the latter. A difference that clearly contributed to our difficulties with spurious polysemy among the Greek equivalents of the same English word.

5.2. Study of multilingual intertextuality

The AGWN is aimed at supporting the study of multilingual intertextuality inside the *Memorata Poetis* Project (Boschetti et al., 2014), an Italian PRIN 2010/2011 funded project focused on literary and epigraphic poetic texts in Greek, Latin, Italian and Arabic, in order to evaluate the transmission of themes and motives across different civilizations.

5.3. Peculiarities of the user interface

Software to edit WordNets, such as DEBVisDic (Horák et al., 2006a; Horák et al., 2006b) or WordNet Atlas (Abrate et al., 2012; Abrate and Bacciu, 2012) and Wikyoto Knowledge Editor (Ronzano et al., 2011), have been evaluated before the creation of a new user interface to query and edit the WordNet and we decided that some peculiarities of the targeted language need to be managed accurately.

For example, in ancient Greek and in some other languages, such as Arabic, present participles can be systematically used as adjectives and nouns, whereas in other languages, such Italian, only a few infinitives, past participles and present participles (e.g. “cantante”, that means “singer”), lexicalized in the dictionaries, are synonyms of the corresponding *nomina actionis* (“canzone, *nomina rei actae* or *nomina agentis* (e.g. “cantore”, also with the meaning of “singer”). Currently, the automated procedure that extracts the synsets from bilingual dictionaries, due to the semantic relations among different parts of speech, in many cases puts in the same set both nouns and verbs. Through the user interface, the reviewer is able to generate from the verbal lemma the correct inflected form (e.g. the participle, or in other cases the infinitive, etc.) that is synonym to some nouns in a nominal synset, to lexicalize it, preserving the morphological information and the lexical relation with the original lemma, and eventually to validate it.

6. Future Work

In order to reach different groups of users, modules with the same functionality and similar design must be developed for different platforms. In particular, we are planning to develop a module for pedagogical use with the Moodle (<https://moodle.org>) platform and one for more advanced use with the Perseids platform for scholarly annotation of Classical texts. Furthermore, a variety of data should be

linked to AGWN, such as etymological relations with other WNs, through the crowdsourced Etymological WordNet (<http://www1.icsi.berkeley.edu/demelo/etymwn>).

The GUI needs to reflect these new ideas: while the customization for specific communities is *in nuce* within the specific lexicons related to single authors, other features must be added to make the GUI as complete as possible.

The following features are planned:

Bilingual Search It is essential to be able to perform a bilingual search, crossing words in two different languages, in order to have a clear idea of missing couples in the created resource. In this case, experts may see a direct snapshot of the vulnerabilities of the created resource and may add the missing words using the features developed and described in section 3.

Semantic tagging of specific texts The GUI needs to help users to select the correct sense of a specific word in a given text. This is essential for experts to have a plain idea of what authors mean in a specific contexts. First experiments have been performed on Homer.

Access and Identity Management This is already available, but we need to create work-groups for specific authors. Even if the data model is designed to perform these features, an investigation is needed to create correct groups of users who share the same knowledge of specific Greek authors.

Validation by a superuser The user profile panel contains the list of activities performed by a user. This list is editable by both the specific owner and by the superuser. Added and/or removed words and relations must be validated by the superuser in order to be effectively part of the WordNets. As long as the single activity is not validated the WN remains unaltered.

6.1. Distribution

The data contained in the Ancient Greek WordNet will be released as (Linguistic) Linked Open Data ((L)LOD) following the way PWN is released into RDF.⁶

Indeed Italian WordNet (IWN) has been already released as (L)LOD, (Del Gratta et al., 2013; Bartolini et al., 2013) and the other WNs shortly will be released.

(L)LOD represents a new trend in the publication of linguistic resources: a survey on the formats and frameworks used in the last 20 years to exchange linguistic resources, (Lezcano et al., 2013) found “an increase in recent years in approaches adopting the Linked Data initiative”.

(L)LOD are still quantitatively a minority within the linked data cloud (Chiaros et al., 2011; Lezcano et al., 2013) but they are growing and becoming a central modality for linguistic data and especially for lexical data publication. Even if not big in number of triples,⁷ they are significant in specific weight - especially the resources manually developed/checked as the ones contained in Ancient Greek WordNet.

⁶For example, <http://www.w3.org/2006/03/wn/wn20/instances/wordbank.rdf>

⁷<http://linguistics.okfn.org/resources/lld/>.

Acknowledgments

We acknowledge Eleonora Sausa (University of Pavia), for her contribution to the initial design of the AGWN, Antonio De Prisco (University of Verona) for the interconnection to Latin WordNet, Neven Jovanović (University of Zagreb) for his contribution to the connection with Croatian WordNet.

This research has been partially co-funded by the NEH and the Italian CNR.

7. References

- Abrate, M. and Bacciu, C. (2012). Visualizing word senses in wordnet atlas. In *LREC*, pages 2648–2652.
- Abrate, M., Bacciu, C., Marchetti, A., and Tesconi, M. (2012). Wordnet atlas: a web application for visualizing wordnet as a zoomable map. In *GWC 2012 6th International Global Wordnet Conference*, page 23.
- Autenrieth, G. (1891). *A Homeric Dictionary for Schools and Colleges*. Harper and Brothers, New York.
- Bamman, D. and Crane, G. (2008). Building a Dynamic Lexicon from a Digital Library. In *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2008)*, Pittsburgh, PA, USA.
- Bamman, D. and Crane, G. (2011). Measuring Historical Word Sense Variation. In *Proceedings of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2011)*, Ottawa.
- Bartolini, R., Del Gratta, R., and Frontini, F. (2013). Towards the establishment of a linguistic linked data network for Italian. In *Proceedings of the 2nd Workshop On Linked Data in Linguistic, Collocated with the 6th International Conference on Generative Approaches to the Lexicon*, Pisa, Italy, September.
- Boschetti, F., Del Grosso, A. M., Khan, A. F., Lamé, M., and Nahli, O. (2014). A top-down approach to the design of components for the philological domain. In *DH 2014 (accepted)*.
- Chiarcos, C., Hellmann, S., and Nordhoff, S. (2011). Towards a linguistic linked open data cloud: The open linguistics working group. *TAL*, 52(3):245–275.
- Del Gratta, R., Frontini, F., Khan, F., and Monachini, M. (2013). Converting the PAROLE SIMPLE CLIPS Lexicon into RDF with lemon. *Semantic Web Journal (submitted)*.
- Egger, C. (2004). *Lexicon Recentis Latinitatis*. Officina Libraria Editoria Vaticana (LEV), Città del Vaticano.
- Fellbaum, C., editor. (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, Cambridge, MA, USA.
- Horák, A., Pala, K., Rambousek, A., and Povolný, M. (2006a). DEBVisDic - First Version of New Client-Server Wordnet Browsing and Editing Tool. In *Proceedings of the Third International WordNet Conference - GWC 2006*, pages 325–328, Brno, Czech Republic. Masaryk University.
- Horák, A., Pala, K., Rambousek, A., and Rychlý, P. (2006b). New clients for dictionary writing on the DEB platform. In *DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writings Systems*, pages 17–23, Torino, Italy. Lexical Computing Ltd., U.K.
- Kulkarni, M., Dangarikar, C., Kulkarni, I., Nanda, A., and Bhattacharya, P. (2010). Introducing sanskrit wordnet. In *The 5th International Conference of the Global Word-Net Association (GWC-2010)*, 31st Jan-4th Feb.
- Lezcano, L., Sanchez, S., and Roa-Valverde, A. J. (2013). A survey on the exchange of linguistic resources: Publishing linguistic linked open data on the web. *Program: electronic library and information systems*, 47(3):3.
- Liddell, H. G. and Scott, R. (1889). *An Intermediate Greek-English Lexicon*. Clarendon Press, Oxford.
- Liddell, H. G., Scott, R., Jones, H. S., and McKenzie, R. (1940). *A Greek-English lexicon / compiled by Henry George Liddell and Robert Scott*. Clarendon Press, Oxford, new edition edition.
- Magnini, B., Strapparava, C., Pezzulo, G., and Gliozzo, A. (2001). Using domain information for word sense disambiguation. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems, SENSEVAL '01*, pages 111–114. ACL.
- McGillivray, B. (2010). Automatic selectional preference acquisition for latin verbs. In *Proceedings of the ACL 2010 Student Research Workshop, ACLstudent '10*, pages 73–78. ACL.
- Minozzi, S. (2009). The Latin WordNet Project. In Anreiter, P. and Kienpointner, M., editors, *Latin Linguistics Today. Akten des 15. Internationalem Kolloquiums zur Lateinischen Linguistik*, volume 137 of *Innsbrucker Beiträge zur Sprachwissenschaft*, pages 707–716.
- Ronzano, F., Marchetti, A., and Tesconi, M. (2011). Editing Knowledge Resources: The Wiki Way. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*, pages 2533–2536, New York, NY, USA. ACM.
- Roventini, A., Alonge, A., Bertagna, F., Calzolari, N., Girardi, C., Magnini, B., Marinelli, R., and Zampolli, A. (2003). Italwordnet: building a large semantic database for the automatic treatment of italian. *Computational Linguistics in Pisa, Special Issue, XVIII-XIX, Pisa-Roma, IEPI*, 2:745–791.
- Sagot, B. and Fišer, D. (2011). Extending wordnets by learning from multiple resources. In *LTC'11 : 5th Language and Technology Conference*, Poznań, Poland, November.
- Sausa, E. (2012). Toward an ancient greek wordnet. <http://goo.gl/y3H3qu>.
- Schmidt, J. H. H. (1876). *Synonymik der griechischen Sprache*. B.G. Teubner, Leipzig.
- Vossen, P. (1996). Right or wrong: Combining lexical resources in the eurowordnet project. In *Euralex*, volume 96, pages 715–728. Citeseer.