

Computer Assisted Annotation of Themes and Motifs in Ancient Greek Epigrams: First Steps

Federico Boschetti

Riccardo Del Gratta

Marion Lamé

ILC-CNR

Via Moruzzi, 1

56124 Pisa - ITALY

{firstname.lastname}@ilc.cnr.it

Abstract

English. This paper aims at illustrating some tools to assist the manual annotation of themes and motifs in literary and epigraphic epigrams for the PRIN 2010/2011 Memorata Poetis Project.

Italiano. *Questo contributo ha lo scopo di illustrare alcuni strumenti per assistere l'annotazione manuale di temi e motivi in epigrammi letterari ed epigrafici, nell'ambito del progetto PRIN 2010/2011 Memorata Poetis.*

1 Overview

The Memorata Poetis Project is a national funded project (PRIN 2010/2011), led by Professor Paolo Mastandrea, “Ca’ Foscari” University of Venice, in continuity with the Musisque Deoque Project (Mastandrea and Spinazzè, 2011). It aims at the study of the intertextuality between epigraphic and literary epigrams in Greek, Latin, Arabic and Italian languages. Some of those epigrams are translated in more languages. Currently the access to the website (<http://memoratapoetis.it>) is restricted to the project workgroups but the access will be public before the end of the project, i.e. February 2016.

To understand the specific goal of this work in progress, a broader presentation of the project is necessary. Epigrams are short poems and follow specific schemes, contents and structures. Those short poems are transmitted both by epigraphs and by manuscripts, with interesting relations between the different traditions: an epigram can have been copied from stone to parchment, losing its original function and contextualization or, on the contrary, a literary epigram can have been adapted to a new epigraphic situation. As inscription, epigrams are

a communication device inserted in a cultural construct. They are part of an information system and this implies, in addition to texts and their linguistics aspects: writings, contexts and iconotextual relationships. This holistic *and systemic* construction creates meanings: in Antiquity and in Middle-Ages, for instance, epigrams, as inscriptions, were often epitaphs.

Intertextuality also takes into account this relation between images of the context and the epigrams. For instance, “fountain” is a redundant motive in epigrams. An epigram that refers to divinities of water could be inscribed on a fountain. Such epigraphic situation participates to the global meaning. It helps to study the original audience and the transmission of epigrams. The reuse of themes and motifs illustrates how authors work and may influence other authors. From epigraphs to modern edition of epigrams, intertextuality draws the movement of languages and concepts across the history of epigrams.

Here is an example of a poetic English translation of a Theocritus’ epigram:

XV. [For a Tripod Erected by Damoteles to Bacchus] The precentor Damoteles, Bacchus, exalts / Your tripod, and, sweetest of deities, you. / He was champion of men, if his boyhood had faults; / And he ever loved honour and seemliness too.

(transl. by Calverly, 1892, <https://archive.org/details/Theocritus/TranslatedIntoEnglishVerseByC.s.Calverley>)

Effectively, European cultures enjoyed epigrams since the Antiquity, copied them, translated them, and epigrams became a genre that philology studies ardently. This intercultural process transforms epigrams and, at the same time, tries to keep their essence identifiable in those themes and

motifs. Naturally, those themes and motifs, such as “braveness”, “pain”, “love” or more concretely “rose”, “shield”, “bee” are reflecting the concepts in use in several different languages. The Memorata Poetis Project tries to capture metrical, lexical and semantic relations among the document of this heterogeneous multilingual corpus.

The study of intertextuality is important to understand the transmission of knowledge from author to author, from epoch to epoch, or from civilization to civilization. Even if the mechanisms of the transmission are not explicit, traces can be found through allusions, or thematic similarities. If the same themes are expressed through the same motif(s), probably there is a relation between the civilizations, which express this concept in a literary form, independently by the language in which it is expressed. For instance, the concept of the shortness of life and the necessity to enjoy this short time is expressed both in Greek and Latin literature:

Anthologia Graeca 11, 56 Πῖνε καὶ εὐφραίνου. τί γὰρ αὔριον ἢ τί τὸ μέλλον, / οὐδεὶς γινώσκει. (transl.: Drink and be happy. Nobody knows how will be tomorrow or the future.)

Catullus, *carmina*, 5 Viuamus, mea Lesbia, atque amemus / ... / Nobis cum semel occidit breuis lux, / Nox est perpetua una dormienda. (transl.: Let us live and love, my Lesbia [...] when our short light has set, we have to sleep a never ending night.)

Whereas other units are working on Greek, Latin, and Italian texts, the ILC-CNR unit of the project currently has in charge the semantic annotation of a small part of the Greek and of all the Arabic texts and it is developing computational tools to assist the manual annotation, in order to suggest the most suitable tags that identify themes and motifs. The semantic annotation of literary and historical texts in collaborative environments is a relevant topic in the age of the Semantic Web. At least two approaches are possible: a top-down approach, in which an ontology or a predefined taxonomy is used for the annotation, and a bottom-up approach, in which the text can be annotated with unstructured tags that will be organized in a second stage of the work. By combining these approaches, it is possible to collect more evidence to establish agreement on the annotated texts.

2 Manual Annotation

The distinction between theme and motif is a challenging theoretical question that this large scale intertextual work aims at studying in depth when the critical mass of annotations will be reached. Up to now, among heterogeneous and opposite discussions (see Lefèvre, 2006), the position shared by the Memorata Poetis working groups is that a theme is composed of motifs, even if the taxonomy adopted for the annotation does not reflect a neat distinction between these complex and inter-related concepts. The taxonomy of the themes and motifs has been established by an expert of ancient Greek literature, Gian Carlo Scarpa (“Ca’ Foscari” University of Venice), and an expert of Latin literature, Paola Paolucci (University of Perugia). The items of the taxonomy are close to one hundred. The number varies due to the periodic revisions, coordinated by the central unit, according to the proposals of the other operative units.

Despite the large number of items for the classification, the taxonomy has only three levels of depth, for instance:

Res > Alimenta et potiones > Vinum
(Things > Food and drinks > Wine)

The repertory of themes and motifs is based on the study of the indices of “notabilia” in authoritative editions of Greek and Latin poetic collections of the last five centuries. Thus, the taxonomy adopted is grounded in a long tradition of studies, which organizes the themes in spheres of pertinence, such as the semantic spheres of plants, animals, human beings and gods. However, its hierarchical structure prevent transversal relationships, such as the relation between the body parts of human beings and the body parts of animals, which obviously do not share the same kind of body parts. An ontology-driven organisation of such themes and motifs should enrich the expressivity of the description.

3 Granularity Issues

Manual annotations are performed at the level of the entire epigram, when the annotator is aware that the theme (or the motif) interests the entire document, or at the level of a single verse, if the annotator identifies the line interested by a specific theme or motif. It is not possible to annotate a single word, because the annotation of single words is slower and the citation practices in the domain of

classical philology, related to the identification of themes and motifs, usually require the indication of the verse (or verse sequences). The automated tools that we are developing requires a finer granularity, at word level. Even if the manual annotation is performed at the granularity of verse, the individuation of the highest correlations between peculiar words and themes or motifs that they contribute to express, is a useful exploratory strategy. For this reason it is necessary to lemmatize the texts and to calculate the correlation between a specific word and a specific theme or motif. Relevant associations are used to rank the epigrams not yet manually annotated, but candidate to contain the pertinent theme or motif. In addition, those levels of granularity illustrate the complementarity between top-down and bottom-up approaches. A granularity based on words refers to a semantics defined by the strict content of the text and not on the prototype that the annotator has in mind. For instance, in one case the concept of flower is extensionally defined by the occurrences in the actual texts, through the names of specific flowers, such as “rose” and “violet”. In the second case, the annotator could associate to a sequence of words that never contain the term “flower” a projection of his or her prototypical idea of flower.

4 Lemmatization

The lemmatization of the epigrams has been performed using *Morpheus*, the morphological analyzer for Ancient Greek developed at the Perseus Project (Crane, 1991). Multiple suggestions related to lemma and pos are scored according to the probabilities calculated on the Ancient Greek Treebank, <http://nlp.perseus.tufts.edu/syntax/treebank/greek.html>.

5 Identification of words highly associated to specific themes and motifs

Currently more than 10,000 manual annotations have been performed by the collaborators to the project. The annotated verses have been tokenized by the ILC-CNR unit and tokens have been lemmatized. By evaluating the correlation between lemma and theme or motif, the system that we are developing is able to suggest the most suitable tags. A couple of examples can clarify:

Bákchos is the name of the god of wine, and he is highly correlated to the following themes and motifs:

Bacchus (Bacchus),
 Crapula (pleasure),
 Vinum (wine, as a drink),
 Vinum curis remedium
 (wine, as a solution to cares)
sakós, which means *shield*, is highly associated to the motif *Instrumenta belli* (war instruments)

The association between synsets and themes is derived from the association between words and themes. Each verse has been lemmatized and the correlation of each lemma with the leading theme associated to the verse is evaluated calculating the log likelihood ratio, which involves as parameters the frequency of the pairs lemma - theme under observation; the frequency of the pairs lemma - different theme; the frequency of the pairs different lemma - theme under observation and, finally, the frequency of the pairs other lemma - other theme. From the pairs with the highest scores, lemmas are extracted and searched on Ancient Greek WordNet, in order to identify relevant synonyms.

6 The Ancient Greek WordNet

Even if a word has never been previously annotated by hand, it can be associated to a specific theme or motif, recurring to the growing Ancient Greek WordNet (Bizzoni et al., 2014). As shown in Fig. 1, the Ancient Greek WordNet is connected to Latin, Italian, Arabic, English and Croatian WordNets.

Powered by [Riccardo Del Gratta](#) [View License](#) [Acknowledgment](#)

Figure 1: Ancient Greek WordNet GUI.

Searching for the aforementioned word *sákos*, it provides the following synset, composed both by frequent (such as *aspís*) and rare words (such as *prothorákion*).

LSJ Entry	LSJ Translation
sákos	shield
aspideîon	shield
aspís	shield
proballós	shield
párme	buckler
prothorákion	shield

Table 1: Greek-English pairs.

7 Improving the Coverage of Ancient Greek WordNet

The coverage of Ancient Greek WordNet (27% of the lexicon) is still poor, because the associations between Greek and English, the pivot language, has been performed extracting pairs from the LSJ bilingual dictionary and only precise matchings with an English word (or phrase) in Princeton WordNet create an association.

In order to improve the coverage, multiword English definitions have been parsed with the Stanford parser, and the identified head is assumed to be a word belonging to a hypernymic synset. An example:

word: boágrion

English translation: a shield of wild bull's hide

head: shield

part of speech: NN

number of words: 6

syntactic structure:

```
(ROOT (NP (NP (DT) (NN) )
  (PP (IN)
    (NP (NP (JJ)
      (NN) (POS) ) (NN) ) ) ) ) )
```

8 Conclusion

In conclusion, we have presented a work in progress related to the lexico-semantic instruments under development at the ILC-CNR to assist the annotators that collaborate to the Memorata Poetis Project.

Acknowledgments

We acknowledge Luigi Tessarolo, “Ca’ Foscari” University of Venice, for the data related to the current status of annotation of the Memorata Poetis database and Marina Caputo, University of

Perugia, for information related to the procedures adopted for manual annotation.

References

- Luisa Bentivogli and Emanuele Pianta. 2004. *Extending WordNet with Syntagmatic Information*. In Proceedings of the Second Global WordNet Conference, Brno, Czech Republic, January 20-23: 47-53.
- Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini and Gregory Crane. 2014. *The Making of Ancient Greek WordNet*. LREC 2014: 1140-1147
- Gregory Crane. 1991. Generating and Parsing Classical Greek. *Literary and Linguistic Computing*, 6(4): 243-245.
- Matteo Lefèvre. 2006. *Per un profilo storico della critica tematica*. In C. Spila, ed., *Temi e letture*, Roma, 11-29.
- Paolo Mastandrea and Linda Spinazzè. 2011. Nuovi archivi e mezzi d’analisi per i testi poetici. I lavori del progetto Musisque Deoque. Amsterdam.
- Pavese Carlo Odo. 1997. I temi e i motivi della lirica corale ellenica. Pisa.
- Gunawan Pranata and Erick Pranata. 2010. *Acquisition of Hypernymy-Hyponymy Relation between Nouns for WordNet Building*. In Proceedings of the 2010 International Conference on Asian Language Processing (IALP ’10). IEEE Computer Society, Washington, DC, USA: 114-117.
- Di Donato Francesca and Morbidoni Christian and Fonda Simone and Piccioli Alessio and Grassi Marco and Nucci Michele. 2013. *Semantic Annotation with Pundit: A Case Study and a Practical Demonstration*. In Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities, DH-CASE ’13, 16:1–16:4