

# Homeric Greek WordNet: costruire una risorsa lessico-semanticca fra ricerca e didattica

Antonio Stanzone, Lab. di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa

Giulia Re, Lab. di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa

Gloria Mugelli, Lab. di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa

Andrea Taddei, Lab. di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa

Federico Boschetti, CNR-ILC, federico.boschetti@ilc.cnr.it

Riccardo Del Gratta, CNR-ILC, riccardo.delgratta@ilc.cnr.it

## 1 Introduzione

Questo lavoro illustra i primi risultati dei tirocini curricolari concordati fra il Laboratorio di Filologia Collaborativa e Cooperativa (CoPhiLab) del CNR-ILC di Pisa e il Laboratorio di Antropologia del Mondo Antico (LAMA) dell'Università di Pisa.

Da alcuni anni il CoPhiLab, insieme all'Alpheios Project, cura la risorsa lessico-semanticca AncientGreekWordNet (AGWN), descritta in Bizzoni et al. (2014). Le wordnets multilingui, derivate da Princeton WordNet (Fellbaum 1998), permettono di raggruppare in un insieme (detto synset) parole con lo stesso senso attorno a nodi concettuali descritti da una glossa (ad es. [bat, chiropteran]: nocturnal mouselike mammal...) e consentono di distribuire i diversi sensi di una stessa parola polisemica fra diversi synsets (ad es. bat[1]: nocturnal mouselike mammal..., bat[2]: an implement used for hitting the ball in various games).

Mentre Latin WordNet, realizzata da Stefano Minozzi presso l'Università di Verona in collaborazione con l'FBK è una risorsa matura e validata manualmente, AGWN è ancora ad uno stadio aurorale e necessita di validazione. Per questo motivo si è deciso di iniziare la correzione manuale da un sottoinsieme coerente di AGWN costituito dai synsets correlati al lessico omerico, Homeric Greek WordNet (HGWN).

In effetti, una risorsa digitale sulla semanticca omerica (e in particolare della sinonimia), può avere molteplici impieghi in ambito filologico, ad esempio per la valutazione di varianti e per lo studio dei sistemi e dei tipi formulari (mentre per lo studio delle formule omeriche in senso basta combinare l'analisi metrica con la lemmatizzazione: cf. Pavese e Boschetti 2005). La creazione di HGWN permette inoltre di gettare le basi per una risorsa lessico-semanticca diacronica, dove sia possibile seguire l'evoluzione dei significati dei termini da Omero agli stadi successivi della lingua greca.

## 2 Metodo

Ciascuno studente che partecipa allo stage concorda con il suo docente universitario del LAMA e con i ricercatori del CoPhiLab un campo semantico sufficientemente rappresentato dal lessico omerico, su cui basare la propria attività di 225 ore presso il CNR-ILC. Dopo la prima fase di ricerca bibliografica, gli stagisti procedono alla correzione dei synsets tramite l'interfaccia web di HGWN. Dalla vista sintetica sui sensi pertinenti al testo omerico dei termini da validare (si veda Fig. 1), si accede alla scheda di validazione (Fig. 2). Oltre a dizionari bilingui online, gli studenti hanno a disposizione lessici specifici (Ebeling 1880–1885; Cunliffe 2012; etc.), monografie di carattere generale (ad es. Hernandez 1997) e opere su campi semantici specifici.

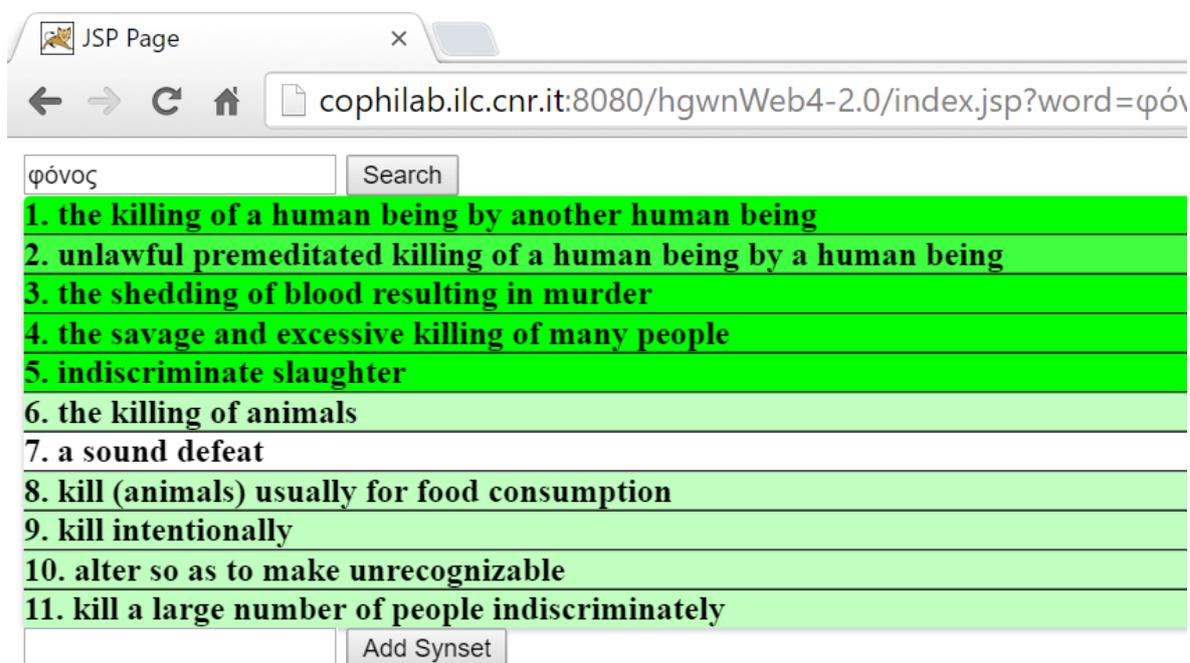


Figura 1: φόνος (phónos: omicidio / strage / spargimento di sangue). Legenda: verde brillante: senso omerico; verde pallido: senso attestati in Greco, ma non in omero; bianco: senso non attestato in Greco

I nostri primi stagisti hanno scelto di lavorare su campi semantici lontani fra di loro: quelli della morte, del sangue e della guerra da un lato e quello della percezione visiva e del colore, dall'altro.

La scelta del campo semantico della morte, del sangue e della guerra (operata seguendo alcuni dei raggruppamenti sinonimici di Paraskevaides 1984 e tenendo presente anche altre monografie, come Di Mauro Battilana 1985 o Mawet 1981) è stata dettata dalla ricchezza (per ovvie ragioni tematiche) di termini omerici di pertinenza. Termini come φόνος (phónos, cfr. Fig. 2), “omicidio” ma anche “sangue versato”, mostrano una caratteristica del lessico omerico, secondo cui il concreto e l'astratto spesso si fondono insieme (e che HGWN separa in synsets diversi).

La scelta del campo semantico ottico parte dall'assunto che “i personaggi di Omero agiscono e reagiscono soprattutto in base alle loro percezioni visive” (Mugler 1964). Di

← → ↻ 🏠

**1. the killing of a human being by another human being**  
**2. unlawful premeditated killing of a human being by a human being**  
**3. the shedding of blood resulting in murder**

synset: 1001100221178 (parent: 1001100220522)

definition (adapted to Ancient Greek)  
 the shedding of blood resulting in murder

note

active  near

**Eng def.:** [the shedding of blood resulting in murder; "he avenged the bloodshed of his kinsmen"]  
**Eng:** bloodshed, gore  
**Ita:** spargimento\_di\_sangue  
**Lat:** cruor

Pertinence	Word	Glimpse Translation	Modification
<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/>	αἷμα	blood	
<input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/> <input type="radio"/>	αἱμαγμός	bloodshed	
<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/>	βρότος	blood that has run from a wound,	
<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/>	λύθρον	defilement from blood, gore	
<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input checked="" type="radio"/>	φόνος	murder, slaughter	

checked

Figura 2: Scheda di validazione. Legenda dei livelli di pertinenza: 1. non pertinente; 2. legame semantico largo; 3. legame semantico stretto; 4. pertinente in Greco post-omerico; 5 pertinente in Omero

grande interesse è l'evoluzione di termini relativi alla vista transitati in ambito filosofico-speculativo, il cui uso è già omerico. Lo studio di tale ambito consente di isolare eventuali sinonimi omerici di termini filosofici successivi, alla ricerca di connessioni semantiche e lessicali (un termine come εἶδος, reso celebre dall'uso platonico come "idea", aveva in Omero un senso molto più concreto, simile a quello di "corpo" / "figura"): vale la pena notare che nei synsets di HGWN sono presenti non solo i termini attestati in Omero, ma anche i termini non omerici ad essi sinonimi e marcati con un livello di pertinenza più basso.

Il campo della percezione visiva è inoltre legato alla sfera dei sentimenti, spesso descritta attraverso metafore o associazioni luminose: in Omero ἀγλή connota il bagliore di una luce riflessa, ma subisce un ulteriore sviluppo semantico figurato ad esempio in Pindaro, in cui indica la gloria acquisita, la magnificenza dell'eroe o dell'atleta.

Le glosse dei synsets sono ereditate da Princeton WordNet, ma gli stagisti sono invitati a modificare la glossa (e stabilire la relazione di near equivalence con Princeton WordNet)

nei casi in cui i termini facciano riferimento a paradigmi epistemici differenti: si pensi alla tassonomia degli animali, alla definizione dei corpi celesti o delle malattie, etc.

### 3 Risultati ottenuti finora e risultati attesi

La dimensione del lessico omerico è di circa 9000 lemmi, di cui oltre il 50% è contenuto nella AGWN attuale. I circa 50 termini, presi in considerazione dai primi due stagisti nei tre mesi di operatività del progetto, hanno già portato alla validazione di un totale di circa 1000 synsets, (data una media di 20 synsets per parola). Mano a mano che il lavoro progredisce, sarà possibile validare sempre più termini, grazie al fatto di ritrovarli in synsets già validati contenenti sinonimi dei termini ancora da validare.

### 4 Conclusioni e sviluppi futuri

Considerate le esperienze svolte e i risultati raggiunti, ad altri studenti che intendono lavorare nel gruppo del LAMA sarà proposto di sfruttare la possibilità di uno stage presso il CoPhiLab. L'esperienza maturata può condurre a programmare sempre meglio (ottimizzando così, tempi e risultati) gli stage che saranno svolti. Riteniamo opportuno, ove possibile, continuare a orientare il lavoro di stage intorno agli interessi specifici dei singoli studenti, in modo da ampliare le loro conoscenze e competenze nello studio della lingua e della letteratura greca antica.

Abbiamo intenzione di creare wordnet di altri autori greci per osservare diacronicamente gli sviluppi lessico-semantiche in prospettiva multidisciplinare: il primo passo già pianificato è la risemantizzazione in tragedia di termini omerici. Date le conseguenze attese sull'apprendimento linguistico (dinamico vs statico, autonomo vs dipendente da vocabolario, ragionato vs mnemonico), riteniamo che lo strumento possa avere impieghi non solo di ricerca, ma anche didatticamente rilevanti.

## Bibliografia

- Bizzoni, Y., et al. 2014. «The making of Ancient Greek WordNet». In *Proceedings of the 9th Annual Conference of LREC*.
- Cunliffe, R. J. 2012. *A lexicon of the Homeric dialect*. Norman, OK: University of Oklahoma Press.
- Di Mauro Battilana, G. 1985. «*Moira*» e «*Aisa*» in *Omero: una ricerca semantica e socioculturale*. Roma: Edizioni dell'Ateneo.
- Ebeling, H., cur. 1880–1885. *Lexicon Homericum*. Lipsiae: Teubner.
- Fellbaum, C. 1998. *Wordnet, an Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Hernandez, M. M. 1997. *Semàntica del Griego Antiguo*. A cura di Clàsticas. Madrid: Ediciones Clàsticas.
- Mawet, F. 1981. «Evolution d'une structure sémantique: le vocabulaire de la douleur: Apollonios de Rhodes et Homère». *L'Antiquité Classique* 50 (1-2): 499–516.
- Mugler, C. 1964. *Dictionnaire Historique de la Terminologie Optique des Grecs*. Paris: C. Klincksieck.
- Paraskevaides, H. A. 1984. *The use of synonymous in the Homeric formulaic diction*. Amsterdam: A. M. Hakkert.
- Pavese, C., e F. Boschetti. 2005. *A Complete Formular Analysis of the Homeric Poems*. Amsterdam: Hakkert.