

WordNets and Treebanks. A study on the semantic field SEA in Latin and Ancient Greek classical prose.

In Latin and ancient Greek linguistics, onomasiological studies on the lexicalizations of concepts have led to a productive line of research based on in-depth philological analysis of extensive textual evidence (e.g. Moussy 1999; Kaster 2001; Craig 2005; Moussy 2005; 2007; Zamboni 2008; Lecaudé 2010; Kutscher and Werning 2014). The main obstacle to large-scale onomasiological analyses is that they require specialized language resources organized by concepts (Black 2001) rather than by lemmas. The Historical Thesaurus of English (Kay 1984) has enabled studies on diachronic patterns in English lexicalization of various concepts (Wild 2010; Alexander 2018; Allan 2020). For ancient Greek and Latin, we can now use WordNet (Biagetti et al. 2021).

Building on the work done within the MultiWordNet project to automatically generate Latin synsets from dictionaries (Minozzi 2017), the Latin (<https://latinwordnet.exeter.ac.uk>) and Ancient Greek WNs (<https://greekwordnet.chs.harvard.edu>), directed by Dr William Short (University of Exeter), aim to create accurate machine-readable and machine-actionable representations of the semantic structures of Ancient Greek and Latin using the WN framework. Thanks to the same data structures and semantic descriptors (*synsets*, *semfields*) provided by the Princeton WN, they intend to model the idiosyncratic and unique features of the semantics of ancient languages, in terms that are compatible and comparable with models constructed for other languages, as well as integratable into larger computational systems. WNs for the ancient languages try to capture diachronic and generic variations in word meaning, and figurative structures that impact semantic development. Based on conceptual metaphor theory in cognitive linguistics (Lakoff and Johnson 1980), they capture both the metaphorical or metonymical relationships between word senses at the level of word meaning and large-scale figurative relations that may operate supra-lexically. In addition, they contain etymological information.

Linking onomasiological resources such as WNs to corpus data can help us answer quantitative questions on the lexicalization of concepts. In our study, we combine WNs for ancient languages and corpus data, with specific attention to Ancient Greek and Latin. For our analysis, we choose the lexical field of SEA, given its importance in the two cultures. For the Greeks and the Romans, the sea was a source of wealth and trade (e.g. Reed 2003; Wilkinson 2020; Boardman et al. 2021), and a key part of their military strategy (e.g. Harris 2017; Nash 2018), but also a place of mystery and danger (e.g. Berens 1979; Lindenlauf 2004; Nikoloska 2012; Beaulieu 2016).

We focus on Ancient Greek and Latin pairs of nouns (e.g. AGr. *thálassa*, *póntos*, *pélagos*, *háls* : Lat. *mare*, *pontus*, *pelagus*, *aequor* ‘sea’) and verbs (e.g. AGr. *pléō* : Lat. *navigo* ‘sail’) related to SEA. We enrich the Ancient Greek and Latin Dependency Treebanks (Celano 2019; v2.1 http://perseusdl.github.io/treebank_data/) with semantic information from the Ancient Greek and Latin WNs. Our corpus includes the treebanked passages from Herodotus’s *Histories* and Caesar’s *De bello Gallico*. We then compare the quantitative data from the two languages to understand in what way Ancient Greek and Latin lexicalize concepts related to SEA, investigating the best way to add semantic information to the Ancient Greek and Latin Dependency Treebanks. This is a challenging task as WNs often provide a large amount of potentially relevant semantic information, and, even in a specific textual context, words can possess multiple senses, or their sense(s) can be ambiguous. Our study will cast new light on the use of WNs for crosslinguistic comparison for ancient Indo-European languages. It will also represent an attempt to digitally link semantic and syntactic information for classical languages, integrating WNs and treebanks (for modern languages see e.g. Kingsbury et al. 2002; Hajnicz 2014).

REFERENCES

- Alexander, M. 2018. Lexicalization Pressure. In *International Conference of English Historical Linguistics XX*, August 2018, Edinburgh, plenary paper.
- Allan, K. 2022. Degrees of lexicalization in the history of English: antonymy and asymmetry. In *International Conference on Historical Linguistics*, Oxford, 1-5 August 2022. Book of Abstracts.
- Biagetti, E., Zanchi, C. & Short, W. M. 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.
- Beaulieu, M-C. 2016. *The Sea in the Greek Imagination*. Philadelphia: University of Pennsylvania Press.
- Berens, E. M. 1979. *The Myths and Legends of Ancient Greece and Rome*. Boston: Longwood Press.
- Blank, A. 2001. Words and concepts in time: Towards diachronic cognitive onomasiology. In *Words in Time: Diachronic Semantics from Different Points of View*, 37–65.
- Boardman, J., G. R. Tsatsikhladze, A. Avram, and J. Hargrave. 2021. *The Greeks and Romans in the Black Sea and the Importance of the Pontic Region for the Graeco-Roman World (7th Century BC-5th Century AD): 20 Years on (1997-2017). Proceedings of the Sixth International Congress on Black Sea Antiquities (Constanta - 18-22 September 2017)*. Oxford: Archaeopress.
- Celano G. G. 2019. The Dependency Treebank for Ancient Greek and Latin. In *Digital Classical Philology*, 279–298.
- Craig, W. 2015. The language of gender: lexical semantics and the Latin vocabulary of unmanly men. In M. Masterson, N. S. Rabinowitz, and J. Robson, (eds.), *Sex in antiquity: exploring gender and sexuality in the ancient world*, 461-481. London: Routledge.
- Hajnicz, E. 2014. Lexico-Semantic Annotation of Składnica Treebank by Means of PLWN Lexical Units. In *Proceedings of the Seventh Global Wordnet Conference*, 23–31.
- Harris, W. V. 2017. Rome at Sea: The Beginnings of Roman Naval Power. In *Greece and Rome* 64 (1), 14–26.
- Kaster, R. A. 2001. The Dynamics of ‘Fastidium’ and the Ideology of Disgust. In *Transactions of the American Philological Association*, 131, 143–89.
- Kay, C. 1984. The Historical Thesaurus of English. In R. R. K. Hartmann (ed.), *Lexeter '83 Proceedings*, 87-90. Tübingen: Niemeyer.
- Kingsbury, P., M. Palmer, and M. Marcus. 2002. Adding Semantic Annotation to the Penn Treebank. In *HLT '02: Proceedings of the second international conference on Human Language Technology Research*, 262–256.
- Kutscher, S. and D. A. Werning. 2014. *On ancient grammars of space: linguistic research on the expression of spatial relations and motion in ancient languages*. Berlin/Boston (Mass.): De Gruyter.
- Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. London: The University of Chicago Press.
- Lecaudé, P. 2010. *La notion de puissance: les équivalents latins du grec δύναμις*. Thèse dactylographiée soutenue le 11 décembre 2010 à l'Université de Paris-Sorbonne.
- Lindenlauf, A. 2004. The Sea as a Place of No Return in Ancient Greece. In *World Archaeology* 35 (3), 416–433.

- Minozzi, Stefano. 2017. Latin WordNet, una rete di conoscenza semantica per il latino e alcune ipotesi di utilizzo nel campo dell'information retrieval. In P. Mastrandrea (ed.), *Strumenti digitali e collaborativi per le scienze dell'antichità*. Venezia: Edizioni Ca' Foscari.
- Moussy, C. 1999. Les vocables latins servant à désigner le sens et la signification. In M. Baratin and C. Moussy (eds.), *Conceptions latines du sens et de la signification*, 13- 27. Paris: PUPS.
- Moussy, C. 2005. Champ lexical et polysémie: l'expression du pardon en latin. In *Journal of Latin Linguistics*, 9(1), 359–370.
- Moussy, C. 2007. Ambiguus, ambiguitas, anceps, utroqueuersus dans le vocabulaire de l'ambiguïté. In C. Moussy and A. Orlandini (eds.), *L'ambiguïté en Grèce et à Rome. Approche linguistique*, 57–64. Paris: PUPS.
- Nash, J. M. 2018. *Rulers of the Sea – Maritime Strategy and Sea Power in Ancient Greece 550-321 BC*. The Australian National University: PhD Thesis.
- Nikoloska, A. 2012. The Sea Voyage of Magna Mater to Rome. In *Histria Antiqua* 21, 365–371.
- Reed, C. M. 2003. *Maritime Traders in the Ancient Greek World*. Cambridge/New York: Cambridge University Press.
- Wild, K. 2010. Angelets, trudgeons and bratlings: the lexicalization of childhood in the Historical Thesaurus of the Oxford English Dictionary. In M. Adams (ed.), *Cunning passages, contrived corridors: Unexpected essays in the history of lexicography*, 289–308. Monza: Polimetrica.
- Wilkinson, Troy. 2020. *The Indian Ocean Trade and the Roman State*. University of Wales Trinity Saint David: MA Thesis.
- Zamboni, A. 2008. Sul lessico latino, tardo e neolatino dell'alienazione mentale. In R. Wright (ed.), *Latin vulgaire–latin tardif VIII: actes du VIIIe Colloque international sur le latin vulgaire et tardif, Oxford, 6-9 septembre 2006*. Hildesheim/New York: Olms-Weidmann.