Computational Approaches to Approximation in Latin

Eleonora Litta, Federica Iurescia

Università Cattolica del Sacro Cuore Milano

This paper investigates lexical strategies of approximation in Latin from a morphological perspective, combining theoretical categorisation with computational methods. Morphological approximation is defined as the encoding of imprecision, borderline category membership, or non-prototypical instances of a concept (Grandi to appear). It allows speakers to refer to entities that are not fully instantiations of a category but are sufficiently close to be contextually recognised as such. Approximation is thus understood as a grammatical and morphological encoding of semantic proximity without categorical completeness. Building on recent categorisations of approximation in morphology (Masini et al. 2023), we examine whether modern theoretical distinctions adequately capture Latin data. Latin employs various morphological means to express approximation. Attenuation appears through the prefix *sub-* (e.g., *subrufus* "reddish", *subabsurdus* "somewhat absurd") and diminutive suffixes like *-ulus* and *-culus* (e.g., *albulus* "whitish", *frigidulus* "chilly"). Disproximation manifests through *pseudo-* (*pseudopropheta* "a false prophet") and the suffix *-aster* (*philosophaster* "a bad philosopher"), indicating items falling short of standards. Faking and imitation can be expressed through compounds with *-formis* (*serpentiformis* "in the form of a serpent").

In our work, the functional domain of approximation is mapped through a systematic extraction of lemmas from the LiLa:Linking Latin Knowledge Base, an open-ended Linked Open Data Knowledge Base relying on a lexically based architecture connecting multiple lexical and textual resources for Latin (Passarotti et al. 2020; Passarotti et al. 2024; Passarotti, Litta 2026). Our case study focuses on adjectives and nouns with *sub-*, addressing the computational challenge of distinguishing spatial meanings ("under") from approximative ones ("somewhat"). SPARQL queries were used to extract prefixed lemmas across a wide chronological span (3rd century BCE to 21st century CE), integrating morphological information from Word Formation Latin, a derivational morphology lexicon (Litta, Passarotti 2019) and semantic data from bilingual dictionaries. We employed dual annotation methods: manual categorisation by domain experts and automatic classification using Large Language Models (LLMs).

The final part of our investigation consists in the evaluation of the viability of leveraging LLMs for the semantic disambiguation of complex morphological structures in historical languages. We assess the specific types of linguistic ambiguity where LLMs align with or diverge from expert manual annotation, particularly in distinguishing literal spatial senses from abstract approximative values. Furthermore, we reflect on the impact of prompt engineering and context window size on the models' performance when dealing with diachronic lexical data. Ultimately, this evaluation aims to define a balanced, hybrid workflow where automated LLM annotation can effectively scale up the semantic enrichment of Linked Open Data linguistic

resources without compromising the philological accuracy required for classical and medieval texts.

References

Grandi, N. (to appear). Evaluative morphology: forty years and not feeling it. *Lingue e linguaggio*.

Litta, E., Passarotti, M. (2019). (When) inflection needs derivation: a word formation lexicon for Latin. In N. Holmes, M. Ottink, J. Schrickx & M. Selig (eds.), *Lemmata Linguistica Latina. Volume I Words and Sounds*. Berlin: De Gruyter, pp. 224-239.

Masini, F., Norde, M., Van Goethem, K. (2023). Approximation in Morphology. Special issue of *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1): 1–26.

Passarotti, M. C., Mambrini, F., Franzini, G., Cecchini, F. M., Litta, E., Moretti, G., Ruffolo, P., Sprugnoli, R. (2020). Interlinking through lemmas. the lexical collection of the LiLa Knowledge Base of linguistic resources for Latin. *Studi e Saggi Linguistici (SSL)* 58(1): 177–212.

Passarotti, M. C., Mambrini F., Moretti G. (2024). The Services of the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin. In Ch. Chiarcos, K. Gkirtzou, M. Ionov, F. Khan, J. P. McCrae, E. Montiel Ponsoda, P. Martín Chozas, eds., *Proceedings of the 9th Workshop on Linked Data in Linguistics @LREC-COLING 2024, Turin*. ELRA and ICCL, pp. 75–83.

Passarotti M., Litta E. (2026). The LiLa Knowledge Base. Querying Interoperable Latin Resources, *Umanistica Digitale* 23: 151–169.