

Of nodes and cells. Two perspectives on (and from) *Word Formation Latin*

Marco Budassi, Eleonora Litta, and Marco Passarotti
Contemporary theoretical approaches to word formation in Latin
ICLL 2019
Universidad de Las Palmas de Gran Canaria
19 June 2019

1 Word Formation Latin (WFL)

Word formation based resource for Latin. **Step-by-step morphotactic** model: each WFR is treated individually as the application of one single rule.

cerno > *certo/or* (V-to-V *-(i)t*) > *concerto/or* (V-to-V *con-*) > *concertator* (V-to-N *-(t)or*) > *concertatorius* (N-to-A Conversion).

Web application: directed graphs in the shape of hierarchical trees.

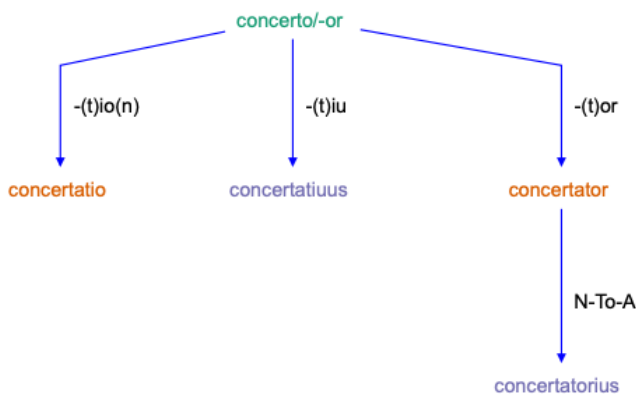


Figure 1: example of directed graph in WFL

Each output lemma can only have one input lemma (except for compounds). **SEARCHABLE:**

- By WFR,
- By Affix,
- By PoS,
- By Lemma.

2 Troubles

Issues raised by fitting all Latin derivation phenomena in the strict morphotactic step-by-step model (Budassi and Litta, 2017). Types of issues:

- Directionality (*amo* > *amicus* A > *amicus* N, or *amo* > *amicus* N > *amicus* A?)
- Double/triple Affixation (*adduco* > *inadducibiliter*)
- Backformation (*iratus* > *irascor*)
- Philological discrepancies (*exhorreo* > *exhorresco*)

- Borrowings (*philosophus* N / *philosophus* A / *philosophia*).

Temporary solutions:

- Trust the dictionary (OLD, TLL), but this does not solve all the problems (i.e. gaps in attestation, backformation, double affixation)
- Creation of **fictional** lemmas (solves only double affixation)

NUMBER OF FICTIONAL LEMMAS IN WFL: 389.
103 fictional lemmas ending in *-bilis*.

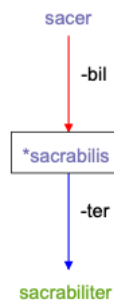


Figure 2: example of fictional lemma

3 Derivational Paradigms

WHAT IS A DERIVATIONAL PARADIGM?

A "derivational paradigm is [...] an arrangement of derivational families. These families are arranged with respect to a set of arrangement relations", that is relations that "connect lexemes formed by a same derivational process" (Hathout and Namer, 2019).

3.1 Advantages of W&P models

- Not limited to base-derivative pairs (Jackendoff, 1975) => allows to conceive the creation of words not necessarily in a linear process.
- No need for directionality (Jackendoff, 1975) => diachronic discrepancies;
- Potentiality (Štekauer, 2014);
- Availability of slots more important than the form filling them (Bauer, 1997);
- Regularity and predictability (Bauer, 1997).

3.2 Construction Morphology (CxM)

“The primary purpose of a good classification is to enable the linguist to make the best generalisations possible about linguistic phenomena.” Geert Booij, 2005

- Sign-based (form and meaning) and Word-based (morphemes are units of analysis in an abstract way).
- 3 notions:
 1. Constructions: conventionalised pairings of form and meaning.
 2. Schemas: they function like rules, similar to WFRs (which are procedural), but they are declarative, they are static generalisations over a set of fully specified items. They are OUTPUT ORIENTED
 3. Constructicon: the collection of all lexical and grammar knowledge of a language.

DERIVED WORDS are represented as items stored in their entirety as simplex, but with internal morphosyntactic structure.

SEMANTIC FEATURES mark fundamental role of semantics in accounting for derivational processes.

EXAMPLE

- immemoratio
- [in[memor](a)(t)io]_N ↔ [something not [mention(ed)]]_N;
- [in[x](t)io]_V ↔ [something not [SEM]]_N.

Note: Constructions and schemas can contain further specifications regarding the morphological features involved, e.g. details on PoS, of the base, thematic vowels, theme etc.

4 LiLa

Knowledge base of linguistic resources for Latin. Makes resources (textual resources, lexical resources, NLP tools) work with each other by connecting them using Semantic Web and Linked Data standards and practices. LiLa is based on an ontology made of:

- **Individuals:** instances of objects (one specific token, lemma etc.)
- **Classes:** types of objects/concepts (token, lemma, PoS etc.)
- **Data properties:** attributes that objects can/must have (morphological features for lemmas/tokens)
- **Object properties:** ways in which classes and individuals can be related to one another: RDF triples. Labels from a restricted vocabulary of knowledge description: `hasLemma`, `hasPoS`

Each component of the ontology is uniquely identified through a Uniform Resource Identifier (URI).

5 WFL in LiLa

In CxM words are described in their formative elements. These are organised into connected classes of objects into the LiLa ontology. Three classes of objects:

1. **Lemma**
2. **Prefix/Suffix**
3. **Base:** automatically generated nodes that act as connectors between Lemmas of the same WF family.

Connected by three possible relationships:

1. `hasPrefix`
2. `hasSuffix`
3. `hasBase`

LEMMAS ARE NEVER RELATED TO EACH OTHER

6 Index of Resources

- LiLa: <https://lila-erc.eu/> - This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme – Grant Agreement No. 769994
- LiLa prototype triplestore: <https://lila-erc.eu/data/>
- WFL: <http://wfl.marginalia.it/> - This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 658332-WFL
- Lemlat: <https://github.com/CIRCSE/LEMLAT3>.

7 Abbreviations

V = Verb
N = Noun
A = Adjective
WFL = Word Formation Latin.
WFR = Word Formation Rule.
W&P = Word and Paradigm.
CxM = Construction Morphology
PoS = Part of Speech
OLD = Oxford Latin Dictionary
TLL = Thesaurus Linguae Latinae

Selected References

- Bauer, L. 1997. Derivational paradigms, *Yearbook of Morphology 1996*, ed. Booij, G. and van Marle, J.. Springer Netherlands, 243–256.
- Booij, G. 2005. Compounding and Derivation: Evidence for Construction Morphology, *Current Issues in Linguistic Theory*, 264, eds. Dressler, W. U. and Kastovsky, D. and Pfeiffer, O. E. and Rainer, F.. John Benjamins Publishing Company: 109–132.
- Booij, G.. 2010. *Construction morphology*. Oxford University Press.
- Booij, G. 2018. *The Construction of Words: Advances in Construction Morphology*. Springer.
- Budassi, M. and Litta, E. 2017. In Trouble with the Rules. Theoretical Issues Raised by the Insertion of -sc- Verbs into Word Formation Latin, *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*. EDUCatt, 15-26.
- Cuzzolin, P. 2014 L'espressione della totalità in latino, *QUADERNI DI AION*, 2, 53–70.
- Hathout, N. and Namer, F. 2019. Paradigms in word formation: what are we up to?, *Morphology*, 2, 29, 153–165.
- Karl Ernst Georges and Heinrich Georges. 1972. *Ausführliches Lateinisch-Deutsches Handwörterbuch*, Hahn, Hannover.
- Peter G.W. Glare. 1982. *Oxford Latin Dictionary*. Oxford University Press, Oxford.
- Otto Gradenwitz. 1904. *Laterali Vocum Latinarum*. Hirzel, Leipzig.
- Charles F. Hockett. 1954. Two Models of Grammatical Description. *Words*, 10: 210-231.
- Jackendoff, R. 1975. Morphological and semantic regularities in the lexicon. *Language*, 639–671.
- Litta, E. 2018. Morphology Beyond Inflection. Building a Word Formation Based Lexicon for Latin, in *Formal Representation and the Digital Humanities*, eds. Cotticelli-Kurras, P. and Giusfredi, F. Cambridge Scholars Publishing.
- Litta, E. and Budassi, M. 2018. What we talk about when we talk about paradigms. Paper presented at Word-formation theories III, Typology and universals in word-formation IV, Košice, 27-30 June 2018.
- Marco Passarotti, Marco Budassi, Eleonora Litta, and Paolo Ruffolo. 2017. The Lemlat 3.0 Package for Morphological Analysis of Latin. *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, 133, 24-31.
- Štekauer, P. 2014 Derivational Paradigms, in *The Oxford Handbook of Derivational Morphology*, eds. Lieber, R. and Štekauer, P. Oxford University Press.