
Exploring *Word Formation Latin*

Eleonora Litta

eleonoramaria.litta@unicatt.it
CIRCSE Research Centre
Università Cattolica del Sacro Cuore, Italy

Marco Passarotti

marco.passarotti@unicatt.it
CIRCSE Research Centre,
Università Cattolica del Sacro Cuore, Italy

Chris Culy

chrisculy@mac.com
Consultant @ CIRCSE Research Center

Paolo Ruffolo

CIRCSE Research Centre
Università Cattolica del Sacro Cuore, Italy

Introduction

Word Formation Latin (WFL) is a derivational morphology resource for Classical Latin, where lemmas are analysed into their formative components, and relationships between them are established on the basis of Word Formation Rules (WFRs). For example *amo* (to love) and *amator* (lover) are connected with a relationship that describes a change from a verb to a noun through the addition of a suffix (-a-tor) that in itself bears semantic information (in this case it characterises agentive and instrumental nouns, i.e. someone or something performing an action).

WFL has received funding from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No 658332-WFL). The resource is still a work-in-progress - having so far covered 5,366 morphological families, 268 WFRs, 22,679 relations - and is due to be completed by October 2017. The lexical basis used for the resource comprises the whole 69,682 lemmas featured in the morphological analyser for [Latin LEMLAT 3.0](#).

The word formation lexicon is built in two steps:

1. Word formation rules (WFRs) are detected using a mixture of previous literature on Latin

derivational morphology (Jenks, 1911; Fruyt, 2011; Oniga, 1988) and semi-automatic procedures (Passarotti and Mambrini 2012).

2. WFRs are applied to lexical data: lemmas and WFRs are paired using a MySQL relational database, and a number of MySQL queries provide the candidate lemmas for each WFR. Input and output pairs are then checked manually, in order to clear out false friends and duplicate results due to homography.

This poster will describe the resource and illustrate the web application that is being developed to easily access the data.

The WFL dataset is both integral part of [Lemlat](#) and used in a [standalone web application](#). The database will be made available for download, so that extensive queries can be run and the data can be used and reused at will. The web application is intuitive and user-friendly. It supports those scholars and students that are not familiar with database querying languages such as SQL, but also Classicists with specific scientific questions.

The lexicon can be browsed either by WFR, affix, input and output Part-of-Speech (PoS) or lemma. Drop-down menus provide the available options for each selection, such as the list of affixes and lemmas. Results are visualised as lists of lemmas and tree graphs, whose nodes are lemmas and edges are WFRs. Trees are interactive. Clicking on a node shows the full derivational tree ("word formation cluster") for the lemma reported in that node. For example, figure 1 shows the word formation cluster for the lemma *computo*, 'to calculate'. Clicking on an edge shows the lemmas built by the WFR described by that edge. Methodological motivations will be given for each browsing option together with suggestions for potential uses of the web to investigate Latin derivational processes. Four browsing choices can help the scholar with an array of linguistic investigations.

1. By WFR - opens research questions on a specific word formation behaviour; for example, it is possible to view and download a list of all verbs that derive from a noun with a conservative derivation process (e.g. *radix* 'root' -> *radicor* 'to grow roots').
2. By Affix - acts similarly as above, but works more specifically on affixal behaviour: for example, it is possible to see all agentive nouns in -tor and verify how many correspond to a female equivalent in -trix.

3. By PoS - useful for studies on macro-categories, such as nominalisation or verbalisation.
4. By Lemma - useful when studying the productivity of one specific morphological family (like the one for *bellum* above) or a group of morphological families.

These explorations lead in many directions through investigations on derivational production and semantics (Can semantic identification of outputs help to show which WFRs are more morphotactically transparent? Which inputs produce a certain kind of outputs? Etc.).

The poster will illustrate a few applications of the resource and a demonstration of case studies. The poster will be accompanied by a live demonstration.

Bibliography

- Forcellini, A.** (1940) "Lexicon totius latinitatis ab Aegidio Forcellini seminarii Patavini alumno lucubratum, deinde a Iosepho Furlanetto eiusdem seminarii alumno emendatum et auctum, nunc vero curantibus Francisco Corradini et Iosepho Perin seminarii Patavini item alumnis emendatius et auctius melioremque in formam redactum." Tom. I AG, Patavii, Typis Seminarii
- Georges, K.E.** (1880) Ausführliches lateinisch-deutsches und deutsch-lateinisches Handwörterbuch... Vol. 2. Hahn'sche Verlags-buchhandlung, 1880.
- Glare, P.G.W.** (1982) Oxford latin dictionary. Clarendon Press. Oxford University Press, 1982.
- Gradenwitz, O.** (1904) Laterculi vocum Latinarum: voces Latinas et a fronte et a tergo ordinandas. S. Hirzel.
- Jenks, P. R.** (1911) A manual of Latin word formation for secondary schools. DC Heath & Company, 1911.
- Fruyt, M.** (2011) "Word-Formation in Classical Latin." A Companion to the Latin Language. 157-175.
- Oniga, R.** (1988) I composti nominali latini: una morfologia generativa. Vol. 29. Pàtron.
- Passarotti, M.C.** (2004). "Development and perspectives of the Latin morphological analyser LEMLAT." *Linguistica computazionale* 20, no. A. 397-414.
- Passarotti, M, and Mambrini, F.** (2012) "First Steps towards the Semi-automatic Development of a Wordformation-based Lexicon of Latin." In Eighth International Conference on Language Resources and Evaluation, (LREC 2012), pp. 852-859. European Language Resources Association (ELRA).