IJCoL

Italian Journal
of Computational Linguistics

Rivista Italiana
di Linguistica Computazionale

Volume 4, Number 2
december 2018

Further Topics Emerging at the
Fourth Italian Conference on Computational Linguistics

aA
ccademia
university
press

Registrazione presso il Tribunale di Trento n. 14/16 del 6 luglio 2016

Rivista Semestrale dell'Associazione Italiana di Linguistica Computazionale (AILC)

Associazione Italiana di
Linguistica Computazionale

direttore responsabile
Michele Arnese

isbn 978-88-31978-70-5

Accademia University Press
via Carlo Alberto 55
I-10123 Torino
info@aAccademia.it
www.aAccademia.it/IJCoL_4_2

IJCoL

Volume 4, Number 2
december 2018

CONTENTS

E Pluribus Unum. Representing Compounding in a Derivational Lexicon of Latin

M. Silvia Micheli*
Università degli Studi di Pavia
Università degli Studi di Bergamo

Eleonora Litta**
Georg-August-Universität Göttingen

WFL is a word formation based resource for Latin where words are analysed in their formative components and connected to each other on the basis of word formation rules. It represents a wide lexical resource for the study of Latin word formation. This paper describes how compounding is treated in the Word Formation Latin *lexicon: the methodology and workflow employed to insert compound lemmas into the resource are described, as well as the reasons behind some methodological choices that have been taken during the process. Through the analysis of some types of Latin compounds, the theoretical contribution of this resource is highlighted and outlined.*

1. Introduction

Word Formation Latin, (WFL) (Litta, Passarotti, and Culy 2016), is a word formation based resource for Latin where words are analysed in their formative components and connected to each other on the basis of word formation rules (WFRs).[1] The creation of WFL was designed to fill a gap in the list of language resources for Latin, that could answer to the current increased interest in both the theoretical and applied aspects of word formation. It represents a wide lexical resource not only for the study of Latin derivational morphology (i.e. affixal and conversive processes), but also for compounding, which has often been neglected in other most recent resources for other languages.[2] The lexical basis behind WFL is the same as the morphological analyser and lemmatiser for Latin Lemlat. First released at the beginning of the 1990s, and recently made freely available in its version 3.0 (Passarotti et al. 2017), Lemlat can be downloaded in standalone or client version at `https://github.com/CIRCSE/LEMLAT3`. All lemmas in its original lexical basis have been collected from three main Classical Latin dictionaries: (Georges and Georges 1972); (Glare 1982); (Gradenwitz 1904), plus the Onomasticon of Forcellini's (Forcellini 1940) 5th edition of *Lexicon Totius Latinitatis*

* Dipartimento di Studi Umanistici - Corso Strada Nuova 65, 27100 Pavia, Italy.
E-mail: `silvia.micheli@unibg.it`

** Georg-August-Universität Göttingen, Institut für informatik. Goldschmidtstrasse 7, 37077 Göttingen.
E-mail: `e.littamodignani@gmail.com`

[1] The WFL project received funding from EU Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Individual Fellowship grant agreement No 658332-WFL. The database can be downloaded from the project's github repository at `https://github.com/CIRCSE/WFL`.

[2] Among them, notable ones are the lexical network for Czech DeriNet (Ševčíková and Žabokrtský 2014) and (Žabokrtský et al. 2016), the derivational lexicon for German DErivBASE (Zeller, Snajder, and Padó 2013) and that for Italian derIvaTario (Talamo, Celata, and Bertinetto 2016).

(Budassi and Passarotti 2016). All those lemmas that share a common (not derived) ancestor belong to the same "morphological family", represented in the web application (`http://wfl.marginalia.it/`) as a tree-graph.

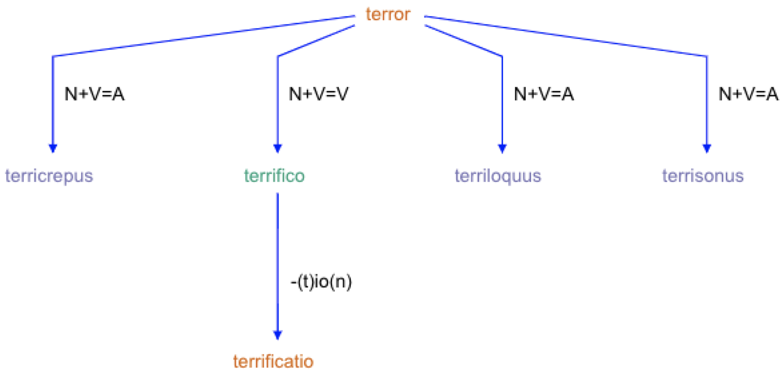

Figure 1
Derivation graph of *terror*

The aim of this paper is twofold: on the one hand, it describes how compounding is represented into the WFL derivational lexicon; on the other hand, it aims at highlighting the contribution of this resource to theoretical linguistics issues through the analysis of some aspects of the Latin compounds collected in it. By providing scholars with all Latin compounds included in the lexical basis mentioned above, WFL can give a more precise portrait of what kind of WFRs (i.e. combination of constituents) are involved in compounding, which are the most used input and output PoS and which are the most common first and second constituents.

2. Latin compounding

Compared to other Indo-European languages (e.g. Sanskrit or Ancient Greek), compounding in Latin is generally considered to be not very productive. According to (Grenier 1912) and (Puccioni 1944), most of Latin compounds are *hapax legomena* and mainly occur in poetic, religious and legal texts. Furthermore, they seem to be strongly influenced by Ancient Greek models.[3]
In the last decades, Latin compounding (henceforth LC) has received more attention (see, for example, works by (Oniga 1992); (Oniga 1988); (Benedetti 1988); (Fruyt 2002); (Brucale 2012). However, most of the available studies are qualitative descriptions of the compounding mechanism, which are based on a small amount of data, usually extracted from dictionaries, and cited as examples of the main types of compounds. These studies have mainly focussed on formal features of LC. Essentially stem-based, Latin compounds are almost always made up of bound units (i.e. roots, stems) connected by a linking element (LE) *-i-* (or sometimes *-u-* when the second constituent starts with a bilabial or a labio-dental sound, e.g. *cornupeta* 'that pushes with the horns,

[3] The influence of Ancient Greek on LC can be observed in three modalities: a) in compounds made up of a Latin and a Greek constituent (e.g. *diversiclinia* 'words irregularly inflected', *diversus* 'different' + κλίνω); b) in calques (e.g. *multiclinatum*, from πολύπτωτον); c) in adapted borrows (e.g. *filosofia* 'philosophy' from φιλοσοφία).

offensive'), as in (1):[4]

(1) *pisciceps*A
pisc-i-ceps
piscis+LE+*capio*+INFL
N+V+INFL=N

As shown in (1), the compound *pisciceps* 'fisherman' is made up of two stems, i.e. *-pisc-* and *-ceps-*, which display formal differences compared to their corresponding lemmas, i.e. *piscis* 'fish' and *capio* 'to take'. In particular, within a compound, *capio* can occur as *-capio* (e.g. *pignoriscapio* 'to take a pledge'), *-ceps* (e.g. *vesticeps* lit. 'who takes the virile toga', 'adolescent'), *-cip-* (e.g. *muscipulum* 'mousetrap').[5] In a number of cases, Latin compounds display a suffix, i.e. *-ium*, which determines the lexical category as well as the syntactic-semantic features of the compound, as shown in (2):

(2) *naufragium*N
nau-frag-ium
navis+frango+SUFF
N+V+SUFF=N

The compound *naufragium* 'shipwreck' results from a process of parasynthesis (cf. (Brucale 2012), (Grandi and Pompei 2010), (Bisetto and Melloni 2008)) since neither **nau-frag* nor **fragium* are attested as full words. In these cases, compounding and suffixation take place simultaneously: this mechanism highlights the strong interaction between compounding and derivation in Latin.
The nature of compound constituents represents the main difference between LC and Romance compounding (henceforth RC), where compounds are mostly made up of two autonomous words. In Italian, for instance, compounds consist of two words, each of which displays its stem and its inflectional morpheme, as exemplified in (3):

(3) *caporedattore*N
cap-o-redattor-e
capo+INFL+*redattore*+INFL
N+INFL+N+INFL=N

In *caporedattore* 'editor-in-chief', both constituents, i.e. *capo* 'chief' and *redattore* 'editor', maintain their inflectional morphemes and their status of full words. [6]
The nature of compound constituents and of the linking element *-i-*,[7] the relationship between compounding and derivation in Latin, and the classification of Latin compounds are the main theoretical topics on which attention is traditionally focussed.

[4] Throughout the paper, we use N to indicate nouns, A for adjectives, V for verbs, INFL for inflectional morphemes, SUFF for suffixes.

[5] In a number of cases, this formal mismatch makes it harder to recognize the lemmas corresponding to the constituents. This point has implications from a computational point of view, since it makes the task of morphological analysis tools harder.

[6] This implies that it is not always easy to draw a distinction between compounds and phrasal lemmas since Italian compounds (unlike e.g. Russian compounds, cf. (Benigni and Masini 2009)) do not show any formal marker for compounding. Therefore, Italian compounding shows a stronger interaction with syntax rather than with derivation.

[7] A survey of the literature on this linking element is outlined in (Brucale 2012).

However, there are still many questions that could not be answered exhaustively due to the scarcity of data collected so far. Which were the most productive compound types and constituents in Latin? Which WFRs formed compounds? What Part-of-Speech (henceforth PoS) did Latin compounds were made of, most frequently? What kinds of meaning were expressed by compounding in Latin?

Furthermore, very little attention has been paid to the fate of LC in Romance Languages (henceforth RLs). Notably, unlike other areas of morphology where one can perceive a significant continuity between RL and Latin (e.g. derivation), there is a strong discontinuity as fa as compounding is concerned.

For instance, at least until the XIX century, Italian compounding is a rather rare, but not totally unproductive, word formation mechanism which shows many formal differences compared to LC (e.g. the unbound nature of Italian compound constituents, the constituent order, etc.). The rise of a new Italian compounding represents an issue which has been poorly studied so far. In this respect, the availability of data on the Latin word formation system is crucial in order to investigate the gap between LC and RC. WFL makes answering to these questions easier, by providing a large account of quantitative data which can help to better understand the mechanisms of LC and its fate.

3. Compounding in Word Formation Latin

The methodology behind Word Formation Latin is consistent with the Item-and-Arrangement (I&A) model outlined in (Hockett 1954), which considers morphemes, not words, the basic units for the study of utterances. Because the I&A model is based on the assumption that morphemes contain both form and meaning, it provides the best theoretical basis for the outcomes of a computational language resource of this kind, which fits into the current need for semantic processing of linguistic data. As a matter of fact, word formation and semantics are strongly connected: words that share the same formative morphemes often share also basic semantic features.

WFL relies on a fairly strict morphotactic approach. Morphotactics is the way morphemes combine with each other. In WFL, to the basic component of the uninflected word, the so-called les ("LExical Segment", a list of which was initially extracted from Lemlat), one derivational morpheme (prefix/suffix) or phenomenon (conversive PoS change) is considered for each single derivational passage. As a result, the output of a WFR is always a lemma richer than the input one, and the output lemma always contains one additional morpheme, or a change of PoS, in case of conversions. For example: *cerno* 'to sift, distinguish, decide' > *certo/or* 'to contend for superiority' (V-to-V *-(i)t*) > *concerto/or* 'to engage in a contest, fight' (V-to-V *con-)* > *concertator* 'rival' (V-to-N *-(t)or*)> *concertatorius* 'controversial' (N-to-A Conversion).

The WFL data was collected and structured in a MySQL relational database in four main phases. First, a list of WFRs was obtained both manually and automatically (Passarotti and Mambrini 2012); the WFRs were then identified and formalised into a table according to their type (prefixal, suffixal, compound and conversion) and to the category of transformation undergone by the lexical element in input (N-to-N, N-to-V, N-to-A, etc.). Then, a series of SQL queries was applied to the lexical data in order to pair input (origin of the derivation) with output (derived) lemmas, one WFR at a time. Subsequently, the resulting list of candidate pairs is thoroughly checked manually for coherence and amended where needed. The final necessary step consisted in the manual insertion of lemmas not picked up by the SQL queries due to phonetic change or non-concatenative formations (i.e. the formation of lemmas not necessarily created by linear morphotactic rules, like backformation or analogy (Budassi and Litta 2017)).

However, the insertion of compounds into the database was necessarily different: as seen in Section 1, with the example of *-capio/-ceps/-cip*, compounds display a combination of morphemes that not necessarily correspond to anything that is recorded into Lemlat's list of lexical segments. This means that the insertion of compounds into WFL not always involved the same workflow based on the pairing of input and output candidates from lists of lexical segments and lemmas, but were often added to the database manually. An initial list of possible compounds was drawn by taking into account all possible combinations of V (verb), N (nouns), A (adjectives), PR (pronouns), and I (invariables - e.g. adverbs, some numerals). Some categories were filled semi-automatically with the help of SQL queries. This happened however after considering popular categories of widely used compounds. These usually matched a string that combines a certain lexical element + *-i-* + a customised string. This method was possible for example for those verbs including *-fico* (from verb *facio* 'to make', e.g. *clarifico* 'to make illustrious'), or those adjectives or nouns featuring noun *pes* 'foot' as a second constituent (e.g. celer-i-pes, lit. 'fast foot'). However, morphotactically obscure compounds like *fidicina* 'lyre player' (*fides* 'lyre' + *cano* 'to sing'), needed to be inserted completely manually.
The WFL lexicon is accessible in two different ways. First, online through a visualisation query system currently at http://wfl.marginalia.it.[8] Online, the data can be browsed according to four different perspectives implemented as four different screens, which can be accessed via a top-level menu. These have been modelled on four kinds of research questions and results that a user might be interested in (by WFR, by affix, by PoS, by lemma).

For what compounds are concerned, the WFL web application allows browsing specifically in three ways:

1. By WFR - this option opens research questions on a specific word formation behaviour; for example, it is possible to view and download a list of all adjectives formed by a A+V=A rule.
2. By PoS - this is useful for studies on macro-categories, such as nouns, adjectives, verbs and invariables, and because it allows for deeper refinement of constituent PoS (e.g. Nouns can be selected to be of one specific declension or gender).
3. By Lemma - allows for quick search of a specific lemma.

The data is visualised as a list of lemmas matching a query, or as tree-shaped graphs representing the derivational cluster for a specific lemma. In each graph, nodes are lemmas, and edges are relations showing the kind of WFR involved. The tree includes all the lemmas derived from the lemma selected, as well as all those words the lemma is derived from. For each compound, a derivational tree-graph is provided (as in Figure 1). It is possible to change the perspective on the view of Figure 1, by clicking on any of the lemmas to inspect their relative word formation families. Special provisions are made in order to collapse and hide compounding relations according to the user's choice. Because a compound is the result of the combination of two (or more) lemmas belonging to different families, this option is useful when very productive constituents are displayed in massive multi-tree graphs. Each visualisation can be downloaded as an

[8] Accessed 17th April 2018.

image.

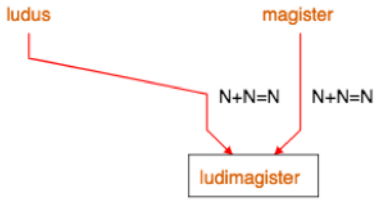

Figure 2
Derivation graph of *ludimagister*

At the time of writing, a few improvements on morphological families visualisations are being planned. In particular for compounds we will implement the opportunity to only see compounding relations for each lemma (e.g. when clicking on *ludus* from *ludimagister*, it will be possible to display also its combination with *facio* resulting in verb *ludificor*).

3.1 Some caveats

The main bedrock of WFL methodology lies in its strict relation to the morphological analyser Lemlat and on the PoS categorisation dictated by its lexical basis. As a consequence, the way compound constituents are classified in their lexical category can sometimes be unconventional. Participles that act as adjectives, even if habitually listed in the dictionaries, are not included in the Lemlat lexical basis, because they are seen as part of the verbal paradigm, this means that certain compounds that would be expected to have a A as one of their constituents have a V instead; e.g *altivolans* (*altus* 'tall, high', past participle of verb *alo* 'to feed, grow' + *volo* 'to fly') 'high flying' can be found among V+V=A compounds rather than among A+V=A. In a similar fashion, certain type of adverbs ending in *-e* are considered in Lemlat ablative cases of the adjectival declension, so *dulciloquus* (*dulce* 'sweetly', from *dulcis* 'sweet' + *loquor* 'to speak') 'sweet talking' is to be found among A+V=A, rather than I+V=A.

Because WFL's lexical basis is drawn from (Glare 1982), (Georges and Georges 1972) and (Gradenwitz 1904), *Oxford Latin Dictionary* and *Georges&Georges* act like manuals for solving a number of theoretical conundrums. However, not all decisions regarding the nature of affixes have respected this rule of thumb. For instance, unlike some traditional studies on Latin word-formation (i.e. (Benedetti 1988), (Fruyt 2002) and (Fruyt 2011)), prepositions (e.g. *cum* 'with' or *in* 'in') are not considered constituents, because one of the main characteristics of the I&A model consists in ascribing the distinction between derivation and compounding to the presence of bound morphemes. As a matter of fact, in (Glare 1982), prepositions that can attach to another lemma to form a new word, can be found under two different entries, one of which is followed by a dash character, indicating a bound lexeme. Hence, those formations that are more traditionally considered preverbal compounds are however to be found among prefixed verbs.

On the other hand, even if this has led to inconsistencies, the same choices have not been made around lexicalised suffixes such as *-ficus* and *-fex* from *facio*, *-ceps* from *capio*, *-ger* from *gero* etc. Even if these would make more sense as suffixes, due to the unique way they have changed over time to only resemble etymologically their origin lemma, we have decided to consider them as verbs in order to more precisely depict the full extent of compounding in the Latin language.
Special consideration should be given to numerals. As we have seen above, there is a clear distinction, in the *Oxford Latin Dictionary*, between affixes and isolated words where the lemma's formative elements are specified. For what numerals are concerned, words preceded by *bi-*, *tri-* and *quadri-* are, as a consequence, included among prefixed lemmas, while lemmas including number from five upwards are to be found among compounds. For instance, *quadriennium* 'period of four years' is inserted in WFL as *quadri-* 'consisting of four of the things following' + *annus* 'year' and not *quatuor* 'four' + *annus*). Additionally, most numerals are categorised under the generic tag N, indicating indeclined nouns (rather than adjectives), which means that any numeral bigger than 4 will appear as a N in a compound. For example, *sexennium* 'period of six years' has been inserted as a N+N=N compound, because *Oxford Latin Dictionary* does not list *sex* as a prefix but only as an indeclinable adjective. Likewise, there is no *uni-* prefix in the dictionary, hence *unus, -a, -um* 'one' always appears as and adjectival first constituent, as in *unoculus* 'that has one eye', a A+N=A compound from *unus* and *oculus* 'eye'.

4. Case studies

4.1 Word Formation Rules

The compounds extracted from the WFL lexical basis are 2003. The fact that all compounds collected from the three dictionaries mentioned above are for the first time categorised and labelled into a language resource allows for a more in-depth overview and for a quantitative analysis on many aspects of LC (e.g. productivity, WFRs, lexical categories involved in compounding). Compound words collected in WFL are created through 59 WFRs. In table 1, the first twenty most productive WFRs are shown.
The most productive pattern in LC is Noun+Verb: this rule creates adjectives, nouns and verbs, e.g. *soporifer* 'soporific' (*sopor+fero*), *artifex* 'artisan' (*ars+facio*) and *aedifico* 'to build' (*aedes+facio)*. This word formation process is no longer productive in RLs, where the reverse order (i.e. the Verb+Noun pattern, e.g. Italian *portafoglio* 'wallet' or French *porte-parole* 'spokesman') is the most frequent for the exception of Romanian, (Grossmann 2012). This change in constituent order can be related to the more general syntactic shift from Object-Verb (OV) to Verb-Object (VO) order which occurred in the transition from Latin to RLs. This correspondence between the constituent order in compounds and in phrases would support the hypothesis sustained by (Gaeta 2008) according to which morphology is not autonomous from syntax for what this specific property is concerned.
Following the classification proposed by (Bisetto and Scalise 2005), three kinds of relation between constituents can be identified, i.e. subordinative, coordinative and attributive relations. In particular, subordination involves asymmetry between the two constituents (one of which is considered as the head, the other as the modifier), while coordination refers to same-level ordering; attributive relation occurs within compounds made up of a noun and an adjective. In nominal compounds made up of

Table 1
Compounding WFRs in WFL

	WFRs	Examples	Compounds
1	N+V=A	*amorifer*	470
2	N+V=N	*agricola*	253
3	A+N=A	*aequanimis*	160
4	A+V=A	*amoenifer*	158
5	N+N=N	*arcuballista*	135
6	N+N=A	*flammipes*	128
7	N+V=V	*bellifico*	88
8	V+V=V	*calefacio*	87
9	A+V=V	*certifico*	76
10	A+N=N	*angustiportum*	41
11	V+V=A	*candificus*	40
12	A+A=A	*dulcacidus*	39
13	V+N=A	*versiformis*	35
14	A+V=N	*infanticida*	33
15	I+I=I	*etiam-tum*	27
16	N+A=A	*animaequus*	21
17	I+N=N	*paeninsula*	16
18	PR+PR=PR	*alteruter*	15
19	N+A=N	*pedeplana*	13
20	PR+V=PR	*qualislibet*	13

two nouns (i.e. compounds obtained through the N+N=N WFR),[9] subordination is the most frequent relation (78 forms, 90%) between the two constituents (e.g. *ludimagister* lit. 'school+teacher', 'schoolteacher'). However, we can also observe a number of cases of coordinative compounds (7 forms, 10%; e.g. *tunicopallium* lit. 'tunic+pallium' or *masculofemina* lit. 'man+woman', 'hermaphrodite'), which are widely considered very rare both in Latin and in RLs. Coordination also occurs in A+A=A compounds which express the coexistence of two properties, e.g. *dulcamarus* (lit. 'sweet+bitter') 'bittersweet' or *sacrosanctus* (lit. 'sacred-holy') 'sacrosant'.

The V+V pattern, through which Italian creates nouns (e.g. *dormiveglia* 'half-sleep', lit. 'to sleep-to stay awake'), in Latin mainly forms new verbs, such as *patefacio* 'to reveal' (*pateo* 'to be evident' + *facio* 'to do').

In addiction to other patterns already identified as productive in previous literature (i.e. A+N=A, N+N=N, N+N=A), it is interesting to highlight the presence of a significant number of compounds consisting of two invariable forms (e.g. *etiamtum*, *etiam+tum* 'even then, yet') or two pronouns (e.g. *aliquis*, *alis+quis* 'anyone, someone') which are generally not mentioned in studies on Latin word-formation.

Moreover, it is worth noting that, in almost all cases, Latin compounds are made up of

[9] As explained above, in WFL numbers higher that four are labeled as N, due to the Lemlat categorisation, although they are generally considered as indeclinable adjectives. This implies that among N+N=N compounds one can find forms (namely 49 compounds) which are rather made up of an adjective and a noun linked by an attributive relation. For this reason, only the 86 compounds formed by two true nouns are considered here.

two constituents. There are only very few cases in which the compound is made of three elements, e.g. *turpilucricupidus* (turpis 'vile' + lucrum 'gain' + cupidus 'desirous'; WFR: A+N+N=N) or *suovetaurilia* (sus 'pig' + ovis 'sheep' + taurus 'bull'; WFR: N+N+N=N).

4.2 Input and output lexical categories

As already pointed out by (Brucale 2012), verbs and nouns are the most frequent input elements in Latin compounds. While nouns can be found both as first and/or second constituents, verbs show a clear tendency to appear in second position. Data collected in WFL confirms these observations.[10]

Table 2
Input and output lexical categories in WFL compounds

Lexical cat.	1° const.	2° const.	Output
A	519	101	1108
I	137	55	67
N	1105	547	513
PR	64	32	53
V	177	1266	273

Table 2 shows the quantitative distribution of the lexical categories (i.e. how many times adjectives are present as input or output PoS) in WFL compounds. More than half of the compounds in WFL (i.e. 1257 forms, 63%) have a verbal second element (e.g. compounds with *-facio* or a related stem, such as *aedifico* 'to build' or *candefacio* 'to whitewash'). On the other hand, nouns tend to occur as first constituent.

As far as the output of compounding WFRs is concerned, it is worth noticing that LC creates mostly adjectives (e.g. compounds with *-fer* as second constituent, such as *alifer* 'winged'), followed by nouns and verbs. Conversely, in RLs, compounding is exploited to create primarily nouns and less frequently adjectives. In Italian, there are very few cases of verbs obtained through compounding, which are made up of a noun and a verb (e.g. *manomettere* 'to tamper with').[11] The formation of pronouns and invariable forms through compounding does not seem to be productive anymore.

4.3 Constituent productivity

Data collected in WFL allowed to go beyond the description of LC based on PoS and to examine compound constituents in more detail.

As far as the first position is concerned (Table 3), the most frequent constituent is *multi-* (from the adjective *multus* 'much, many') which forms 66 compounds (e.g. *multicolor* 'many-colored', *multiformis* 'having many shapes', etc). Among adjectives, *aequus-* 'equal' (e.g. *aequimanus* 'ambidextrous'), *omnis-* 'every' (e.g. *omnigenus* 'of every kind') and the numerals *septem-* 'seven' (e.g. *septicollis* 'seven-hilled') and *quinque* 'five' (e.g. *quinquennis 'five-year-old'*) show a significant productivity. It is interesting to notice that

[10] However, as reported below in section 3.3, in order to interpret correctly the data in Table 2, a distinction should be made between adjectives and adjectival participles, which are categorised here as V.

[11] The first occurrence of *manomettere* goes back to 1219

multi-, *omni-* and *aequi-* are still used productively to create new words in Contemporary Italian as *multi-*, *onni-* and *equi-*: they are generally considered as prefixoids which can combine with adjectives or nouns, e.g. *multiculturale* 'multicultural', *equipotenziale* 'equipotential', *onnicomprensivo* 'all-embracing'.
Bene- and *male-* are the most frequent adverbs occurring in LC: interestingly, both of them, especially *male-*, are also widely used in Italian compounding. In LC, *male-* combines with verbs (e.g. *maledico*, lit. 'to say evil', 'to curse'), adjectives (e.g. *malecastus* lit. 'little, not chaste', 'immoral') and present participle (e.g. *malevolens*, lit. 'wanting evil', 'envious'). In Italian, it can bind to verbs (e.g. *maltrattare* 'to ill-treat'), adjectives (e.g. *malsano* 'insane'), present participle (e.g. *malvivente*, lit. 'bad living', criminal'), past participle (e.g. *malfrequentato* lit. 'badly attended') and adverbs (e.g. *malvolentieri* 'unwillingly').

Table 3
Most productive first constituents in WFL (type frequency > 15)

First constituent	Compounds
multi-	66
aequi-	27
septi-	26
omni-	24
quinque-	23
sesqui-	22
alte/i-	19
centi-	17
auri-	16
bene-	16
male-	16

As already pointed out in section 3.2, the second position is more frequently occupied by a verb: in particular, the most productive verbs are *-facio* 'to make', *-fero* 'to bring' and *-gero* 'to bear' (Table 4). Notably, *facio* can occur within a compound as *-facio* (e.g. *calefacio* 'to heat'), *-fio* (e.g. *liquefio* 'to become melted', *-fex* (e.g. *artifex* 'craftsman'), *-ficus* (e.g. *beneficus* 'beneficent'), *fico* (e.g. *damnifico* 'to injure'). In Italian, the verbal root *-fic-* functions as a productive verbalizing suffix (e.g. *plastificare* 'to laminate') or as an adjectival suffix (e.g. *immaginifico* 'highly imaginative') (Brucale and Mocciaro 2016).
On the other hand, *fero* can occur just as *-fer* (e.g. *aurifer* 'gold-bearing') or seldom as *-lator* (e.g. *legislator* 'law-giver'); *gero* as *-gero* (e.g. *flammigero* 'to blaze') or, more often, as *-ger* (e.g. *naviger* 'ship-bearing').
The most frequent noun occurring as second constituent is *pes*: it displays three bound forms, i.e. *-pes* (e.g. *flexipes* 'crooked-footed'), *-pedus* (e.g. *aequipedus* 'having equal feet') and *-pedal* (e.g. *palmipedalis* 'a foot and a palm in height').
Finally, it is worth pointing out the productivity of the present participle *-potens* 'powerful, capable': it occurs in 28 compounds, such as *omnipotens* 'omnipotent', *altipotens* 'very mighty'. The presence of a present participle as second constituent represents an element of continuity between LC and Italian compounding which has been neglected so far in the literature devoted to LC (cf. (Brucale 2012), (Fruyt 2002)). In Italian, compounds made up of a present participle as second constituent are attested from the earliest stages (e.g. *verodicente* 'that tells the truth' and *malparlante*

Table 4
Most productive second constituents in WFL (type frequency > 15)

Second constituent	Compounds
-facio	284
-fero	190
-gero	76
-pes	64
-loquor	57
-gigno	46
-fluo	28
-colo	28
-potens	28
-caedo	24
-annus	21
-sono	20
-vagus	20
-dico	19
-capio	18
-vir	18
-cano	18
-color	17
-plico	17

'talebearer' are attested from the 13th century), and still used in Contemporary Italian (cf. compounds made up of *-dipendente* 'addicted' as second constituent, e.g. *cibodipendente* 'food-addicted').

The case studies presented above have shown that data collected in WFL provide both quantitative and qualitative information which is helpful to fill gaps in the literature devoted to LC. In particular, it has been shown which were the most productive WFRs (i.e. N+V=N and N+V=A) and the most frequent constituents (i.e. *multus-*, *aequus*, *omnis-* as first constituents and *-facio*, *-fero* and *-gero* as second elements). Moreover, data extracted from WFL revealed that compounding in Latin allows to create not only adjectives, nouns and verbs, but also adverbs, conjunctions and pronouns.

5. Conclusions and future work

This paper has provided an overview of how compounding is represented in WFL, a derivational lexicon for Latin. This preliminary study, with its quantitative analysis in the field of LC, shows the potential for raising new questions and issues offered by a resource that for the first time collects all compounds used in Classical and Late Latin. For instance, representing all compounding rules into a network, as it has been already successfully done for the affixal rules listed in WFL, (Litta, Passarotti, and Ruffolo 2017), could lead to further research questions. These could be the investigation on constituent typologies or on the productivity of the different types of compounds. Future developments in WFL should consist in finding a way of searching through constituents by original lemma (as opposed to only PoS), and implementing a way of marking those PoS that appear differently in the resource's lexical basis, such as past and present participles that are included in dictionaries as independent lemmas. This would also allow for a more precise quantitative investigation on constituent typologies.

References

Benedetti, Marina. 1988. *I composti radicali latini. Esame storico e comparativo*. Giardini, Pisa.

Benigni, Valentina and Francesca Masini. 2009. Compounds in russian. *Lingue e linguaggio*, 8(2):171–194.

Bisetto, Antonietta and Chiara Melloni. 2008. Parasynthetic compounding. *Lingue e linguaggio*, 7(2):233–260.

Bisetto, Antonietta and Sergio Scalise. 2005. The classification of compounds. *Lingue e linguaggio*, 4(2):319–0.

Brucale, Luisa. 2012. Latin compounds. *Probus*, 24:93–117.

Brucale, Luisa and Egle Mocciaro. 2016. 18 composizione verbale in latino: il caso dei verbi in facio, fico. *LATINITATIS RATIONES: Descriptive and Historical Accounts for the Latin Language*, page 279.

Budassi, Marco and Eleonora Litta. 2017. In trouble with the rules. Theoretical issues raised by the insertion of -sc- verbs into word formation latin. In *Proceedings of the Workshop on Resources and Tools for Derivational Morphology (DeriMo)*, pages 15–26, Milano, Italy, October 5-6.

Budassi, Marco and Marco Passarotti. 2016. Nomen omen. Enhancing the latin morphological analyser lemlat with an onomasticon. In *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2016)*, pages 90–94, Berlin, Germany, August 11.

Forcellini, Egidio. 1940. Lexicon Totius Latinitatis / ad Aeg. Forcellini lucubratum, dein a Jos. Furlanetto emendatum et auctum; nunc demum Fr. Corradini et Jos. Perin curantibus emendatius et auctius meloremque in formam redactum adjecto altera quasi parte Onomastico totius latinitatis opera et studio ejusdem Jos. Perin. Typis Seminarii, Padova.

Fruyt, Michèle. 2002. Constraints and productivity in latin nominal compounding. *Transactions of the Philological Society*, 100(3):259–287.

Fruyt, Michèle. 2011. Word-formation in classical latin. In James Clackson, editor, *A companion to the Latin language*. Wiley-Blackwell, pages 157–175.

Gaeta, Livio. 2008. Constituent order in compounds and syntax: typology and diachrony. *Morphology*, 18(2):117–141.

Georges, Karl Ernst and Heinrich Georges. 1972. *Ausführliches Lateinisch-Deutsches Handworterbuch*. Hahn, Hannover.

Glare, Peter G.W. 1982. *Oxford Latin Dictionary*. Oxford University Press, Oxford.

Gradenwitz, Otto. 1904. *Laterali Vocum Latinarum*. Hirzel, Leipzig.

Grandi, Nicola and Anna Pompei. 2010. Per una tipologia dei composti del greco. *La morfologia del greco tra tipologia e diacronia*, pages 209–232.

Grenier, Albert. 1912. *Ètude sur la formation et l'emploi des composès nominaux dans le latin archaique*. Berger-Levrault, Paris.

Grossmann, Maria. 2012. Romanian compounds. *Probus - International Journal of Latin and Romance Linguistics*, 4(1).

Hockett, Charles F. 1954. Two models of grammatical description. *Words*, 10:210–231.

Litta, Eleonora, Marco Passarotti, and Chris Culy. 2016. Formatio formosa est. Building a word formation lexicon for Latin. In *Proceedings of the Third Italian Conference on Computational Linguistics (CLiC–it 2016)*, pages 185–189, Napoli, December 5-6. aAccademia University Press.

Litta, Eleonora, Marco Passarotti, and Paolo Ruffolo. 2017. Node formation. Using networks to inspect productivity in affixal derivation in classical latin. In *Proceedings of DATeCH2017*, Göttingen, Germany, June 1-2.

Oniga, Renato. 1988. *I composti nominali latini: una morfologia generativa*. Patron, Bologna.

Oniga, Renato. 1992. Compounding in latin. *Rivista di linguistica*, 4(1):97–116.

Passarotti, Marco, Marco Budassi, Eleonora Litta, and Paolo Ruffolo. 2017. The Lemlat 3.0 package for morphological analysis of latin. In Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language, pages 24–31, Gothenburg, Sweden, May 22-24.

Passarotti, Marco and Francesco Mambrini. 2012. First steps towards the semi-automatic development of a wordformation-based lexicon of latin. In *LREC*, pages 852–859.

Puccioni, Giulio. 1944. *L'uso stilistico dei composti nominali latini*. Atti della Accademia d'Italia. Memorie della classe di scienze morali e storiche, Series 7. Accademia d'Italia.

Ševčíková, Magda and Zdeněk Žabokrtský. 2014. Word-formation network for czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, pages 1087–1093, Reykjavik, Iceland, May 26-31. ELRA.

Talamo, Luigi, Chiara Celata, and Pier Marco Bertinetto. 2016. Derivatario: an annotated lexicon of italian derivatives. *Word Structures*, 9(1):72–102.

Žabokrtský, Zdeněk, Magda Ševčíková, Milan Straka, Jonáš Vidra, and Adéla Limburská. 2016. Merging data resources for inflectional and derivational morphology in Czech. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1307–1314, Portorož (Slovenia), May 23-28. ELRA.

Zeller, Britta D., Jan Snajder, and Sebastian Padó. 2013. Derivbase: Inducing and evaluating a derivational morphology resource for german. *ACL*, 1:1201–1211.