

A graph database about Latin nominal compounds

Andrea Brunello – Alessandro Re – Giovanni Torresin

ABSTRACT

After collecting over the past years a substantial number of nominal compounds from Latin authors spanning from the archaic age to the late antiquity, a graph database has been created. This database is capable of classifying the types of compounds, showing the members they are composed of, and specifying in which works and by which authors they are used. The database can be interrogated through a series of queries that return structured data. The analysis of these data makes it possible to draw interesting insights not only about the specific occurrences of certain compounds within the examined corpus, but also to verify whether there are relationships linking literary genres to specific types of compounds.

In the last decades, a large amount of Latin linguistic material has been collected in order to develop an innovative theory of composition: these results have been presented primarily in two monographs (Oniga 1988; Re 2020) and in various essays (cfr. Oniga 1989; 1990; 1992; 1994; 2000; 2002; 2005; Oniga – Re 2017; Oniga 2020; Re 2021).

* A preliminary version of this essay was presented at the *23rd International Colloquium of Latin Linguistics*, held at the University of Udine on June 9-13, 2025. Although the database modeling and the writing of this article are the result of close collaboration among the authors, the individual sections can be assigned as follows: §§1-3 and §5 were written by Alessandro Re; §§3.1-4.1 by Andrea Brunello and Giovanni Torresin; in §4.2, the queries were formulated by Giovanni Torresin, while the interpretation of the results is attributed to Alessandro Re.

Linguistic data seem to confirm the general hypothesis – first proposed in Oniga (1988) and further developed in Oniga – Re (2025) – that there is a close correlation between the genre of literary works and the type of nominal compounds they employ.

Here, after briefly recalling the theoretical principles underlying our research, we are presenting a new database of Latin nominal compounds: this project has been developed in collaboration with Andrea Brunello and Giovanni Torresin.

1. TOWARDS A TYPOLOGICAL CLASSIFICATION OF LATIN COMPOUNDS

Over centuries of research on this topic, multiple classifications of Latin nominal compounds have been developed, reflecting the different linguistic paradigms adopted by the scholars (Lindner 1996; 2002; Lindner – Oniga 2005; Re 2020: 37-66).

First, Greek grammarians recognized that words can have different internal structures.

In the Hellenistic era, Dionysius Thrax introduced the concept of σχήματα ὀνομάτων (“figures of names”), which Latin grammarians translated as *figurae nominum*: names could be primitive (ὄνομα ἁπλοῦν – *nomen simplex*), compound (ὄνομα σύνθετον – *nomen compositum*), or derived from a compound (ὄνομα παρασύνθετον – *nomen decompositum*)¹.

Of particular interest is the observation that those *nomina simplicia*, which form complex semantic units, seldom appear in their ‘free’ form: they typically undergo derivational processes that significantly modify their external aspect.

¹ D. T. p. 29, 5-7 Uhlig. Latin grammarians generally spoke of only two *figurae*, as in Char. *gramm.* II (p. 194, 24-25 Barwick); for the definition of *decompositum* see Prisc. *gramm.* V, 56 (p. 177, 10-13 Keil).

Within Dionysius’ framework of διαφοραὶ ὀνομάτων (“varieties of names”) – a notion later adopted also by Latin grammarians – a fourfold combinatorial model is outlined, distinguishing between ‘complete’ words (ὀνόματα τέλεια – *nomina integra*) and ‘incomplete’ ones (ὀνόματα ἀπολείποντα – *nomina corrupta*). This model is summarized in Table 1, which compares passages from Dionysius Thrax and Priscian².

Dionysius Thrax		Priscian of Caesarea	
ἐκ δύο τελείων	Χειρίσοφος	<i>ex duobus integris</i>	<i>tribunus plebis, iusiurandum</i>
ἐκ δύο ἀπολειπόντων	Σοφοκλῆς	<i>ex duobus corruptis</i>	<i>benivulus, pinnirapus</i>
ἐκ τελείου καὶ ἀπολείποντος	Περικλῆς	<i>ex integro et corrupto</i>	<i>inimicus, extorris</i>
ἐξ ἀπολείποντος καὶ τελείου	Φιλόδημος	<i>ex corrupto et integro</i>	<i>efferus, impius</i>

Table 1: Word combination forms according to Dionysius Thrax and Priscian

Ancient grammarians understood that *nomina corrupta* are seldom visible on the surface of language; their major problem lies in merely stating these facts, without explaining why they occur. Moreover, they totally lack syntactic and semantic considerations regarding the relationship between the members of a *nomen compositum* and the meaning resulting from the combination.

A different approach emerged only at the end of the 18th century AD, when Pāṇini’s grammar began to circulate in the West: the importance of ancient Indian linguistic theory lies in the role it assigns to syntactic elements. Pāṇini’s classification can be sketched as in Figure 1.

Briefly, the classification is based on the overall meaning of the compound in relation to the syntactic connection between its members. First, compounds formed by coordination (*dvandva*) are separated from those formed by determination. Within the second group, an important distinction is made between *bahuvrīhi* and the others: in fact, a *bahuvrīhi* refers to an external element qualified by the compound itself. Among the remaining types, *avyayībhāva* include a preposition, while

² Cfr. D. T. p. 30, 1-4 Uhlig and Prisc. *gramm.* V, 58 (p. 178, 16-19 Keil).

tatpuruṣa consist of two nouns – or a noun and an adjective – in which the first member syntactically determines the second. Within *tatpuruṣa*, in *karmadhāraya*, the first element serves as attribute or apposition of the second. Finally, among *karmadhāraya*, a narrower category called *dvigu* includes those compounds in which the first element is a numeral³.

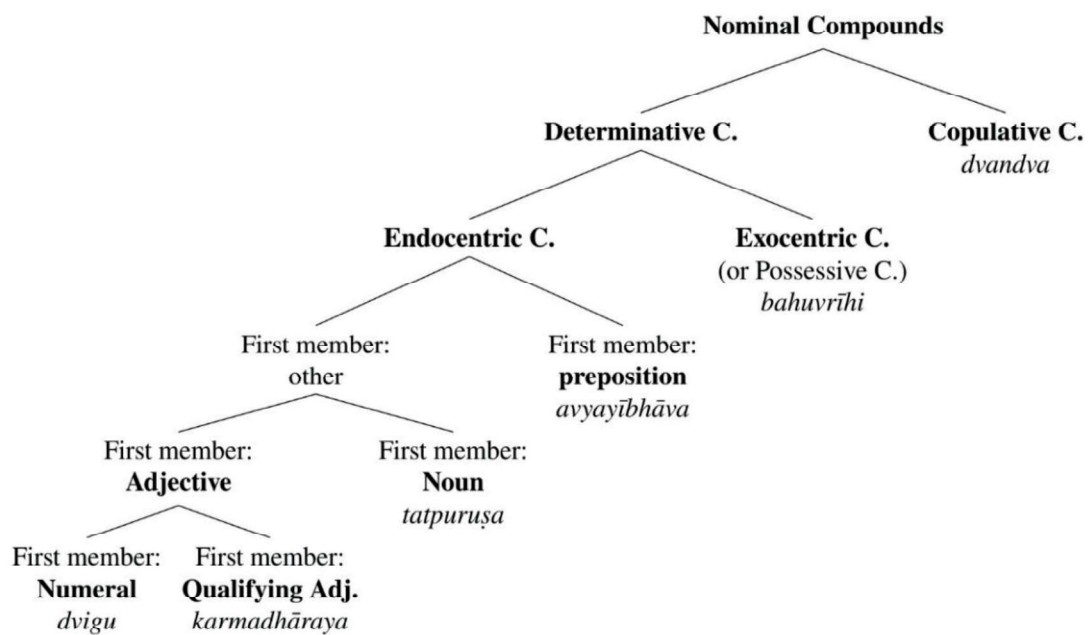


Figure 1: Pāṇini's classification of nominal compounds.

Later scholars adapted this classification to ancient and modern languages, including Latin: this can be clearly seen also in (Leumann 1977: 393-403), where the Pāṇinian theoretical framework is expanded into a complex system of categories and subcategories in order to maximize the descriptive accuracy.

³ For a more detailed discussion, refer to Re (2020: 77-85), with further specific bibliography.

2. THE DEVELOPMENT OF A TYPOLOGY

On the contrary, we argue that it is appropriate to elaborate a typological classification of Latin nominal compounds in which the following aspects are taken into consideration.

2.1. The members of the compound

First, particular attention is given to the morphological category of the compound members. An example of such a classification is already found among ancient grammarians: for instance, in Priscian's *Instituta artium*, this classification is presented⁴.

Type	Examples
Noun + Noun	<i>omniparens, paterfamilias</i>
Noun + Verb	<i>armiger, lucifer</i>
Noun + Participle	<i>senatusdecretum, plebisscitum</i>
Noun + Pronoun	<i>huiuscemodi, illiusmodi</i>
Noun + Adverb	<i>satisfactio, beneficus, maledicus</i>
Noun + Preposition	<i>impudens, perfidus</i>
Noun + Conjunction	<i>uterque, quisque, nequis, siquis</i>

Table 2: Typology of nominal compounds according to Priscian.

Priscian correctly recognized that nouns could be compounded with seven of the eight parts of the speech identified by Latin grammarians, excluding only interjections.

These examples above clarify this typology and also relate to our above comments regarding the combinatorial possibilities of *nomina integra* and *nomina corrupta*.

However, a more careful analysis raises some questions.

Are the second members *-fer* and *-ger* in the “Noun + Verb” compounds truly verbs? It seems difficult to deny, but none of these forms strictly appear in the conjugation of *fero* or *gero*.

⁴ Cfr. Prisc. *gramm.* V, 59 (p. 179, 11-17 Keil).

Likewise, in “Noun + Adverb” compounds, although *factio* is well attested, what about *-dicus* and *-ficus*? These are not recorded as standalone entries in lexicons: they are more appropriately derived from the verbs *dico* and *facio* – just as *-fer* and *-ger* derive from *fero* and *gero* – rather than being independent nouns.

These considerations clearly relate to the earlier discussion about *nomina integra* and *corrupta*: the ‘incompleteness’ frequently observed in compound members results from the fact that, beneath the inflected form, lies a root that seldom appears independently in everyday use. These “abstract words” – to use Mark Aronoff’s definition⁵ – largely correspond to the theme in Latin. Such simple entities are abstract, but really exist in the speaker’s mind, who applies specific rules to embed them into the active language.

2.2. Derivation through suffixation

Second, the phenomenon of derivation cannot be separated from the definition of *nomen corruptum*. In our view, a more accurate explanation consists in recognizing that *-fer*, *-ger*, *-dicus*, and *-ficus* are *nomina agentis* formed through suffixation (cfr. Oniga 1988, 105–11).

Adopting the principles of Construction Grammar⁶, this can be represented by the following construction schema (A).

$$(A) \quad < [[X]_{Va} \text{ Suf}]_{Nb} \leftrightarrow [\text{one who performs SEM}_a]_b >$$

Thus, the verbal roots $\sqrt{dic}$, $\sqrt{fac}$, $\sqrt{fer}$, and $\sqrt{ger}$ yield the nominal derivatives *-dicus*, *-ficus*, *-fer*, and *-ger* via the suffix *-o-*, like *factor* is derived from $\sqrt{fac}$ via *-tor*. However, if *factor* exists as an independent word, *-dicus*, *-ficus*, *-fer*, and *-ger* appear only in compounds such as *maledicus*, *beneficus*, *lucifer*, and *armiger*. This pattern is sketched in this construction schema (B).

⁵ See Aronoff (1976; 1983; 1994).

⁶ We adapt to Latin the theoretical framework of Booij (2015).

(B) $\langle [[x]_{Na} [[y]_{Vb} \text{Suf}]_N]_{Nc} \leftrightarrow [\text{one who performs SEM}_b \text{ on SEM}_a]_c \rangle$

In other cases, suffixation follows composition, as Priscian clearly explains with regard to *nomina decomposita*⁷: these processes can be sketched as in (C) and (D).

(C) $\langle [[x]_{Aa} [[y]_{Nb} + \text{Suf}]_A]_{Ac} \leftrightarrow [\text{one characterized by SEM}_b, \text{ defined by SEM}_a]_c \rangle$

(D) $\langle [[[x]_{Aa} [[y]_{Nb} + \text{Suf}^1]_A]_{Ac} \text{Suf}^2]_{Nd} \leftrightarrow [\text{the condition of [one characterized by SEM}_b, \text{ defined by SEM}_a]_c]_d \rangle$

For example, *magnanimus* results from the compounding of *magnus* and *animus* (C), while *magnanimitas* is derived from *magnanimus* through the suffix *-itas* (D). A ‘native speaker’ like Priscian firmly rejects the existence of *animitas*, though it could be derivated from *animus*, just as *felicitas* derivates from *felix*.

This also confirms the model of *nomina agentis* presented above: although *-cida* does not occur by itself, Priscian indirectly acknowledges its derivation from $\sqrt{caed}$, using the suffix *-a*, and its compounding with *pater* – or *parens*, as Priscian prefers – to form *parricida*.

2.3. Value of the members and of the compound

Last, there is a close interconnection between the syntactic relationship linking the members and the overall semantic of a compound: this reflects the perspective earlier introduced with reference to Pāṇini’s *samāsa* doctrine.

So, it is not sufficient to simply ‘add together’ the first and second members to determine a compound’s classification; rather, a comprehensive analysis must consider morphological, syntactic, and semantic dimensions to explain the compound’s global meaning.

⁷ Cfr. Prisc. *gramm.* V, 56 (p. 177, 10 – p. 178, 3 Keil).

Therefore, the typology here proposed is sensitive to both structural and semantic distinctions. At the same time, the model must avoid the excessive fragmentation that characterizes some of ‘neogrammarian’ classifications above mentioned (§1).

3. FROM THEORIZATION TO TYPOLOGICAL CLASSIFICATION

The theoretical framework expresses a typology that first distinguishes Latin compounds on the base of the morphological category of the second member.

First, the largest and most productive group consists of formations whose second member is a verbal derivative.

This can be further subdivided into the two major categories of *nomina agentis* (E) and *nomina actionis* (F)⁸.

(E) $\langle [[x]_{Na} \ [[y]_{Vb} + \text{Suf}]_{N}]_{Nc} \leftrightarrow [\text{one who performs SEM}_b \text{ on SEM}_a]_c \rangle$

(F) $\langle [[x]_{Na} \ [[y]_{Vb} + \text{Suf}]_{N}]_{Nc} \leftrightarrow [\text{result of SEM}_b \text{ in relation to SEM}_a]_c \rangle$

In both cases, derivation precedes composition: a verbal root is first transformed into a noun by suffixation, and then compounded with a nominal theme. Moreover, the type of suffix enables grouping into further subclasses based on morphological or semantic features.

The other class of Latin compounds includes those whose second member is nominal.

A key distinction must be made between compounds in which the second member undergoes derivation, and those where no derivation occurs.

⁸ For detailed discussions, see Oniga (1988: 76-103) and Re (2020: 142-149); on the distinction between *nomina agentis* and *nomina actionis*, see also Benedetti (1988: 18-21) and Fruyt (1996).

Certain suffixes transform the second member into an abstract noun, whose meaning is then defined in relation to the first member, as in (G)⁹.

$$(G) \quad < [[X]_{Na} [[y]_{Nb} + \text{Suf}]_N]_{Nc} \leftrightarrow [\text{something characterized by SEM}_b, \text{specified by SEM}_a]_c >$$

Although structurally close to *nomina agentis* and *nomina actionis*, they differ in the semantic value: these compounds can be compared to Sanskrit's *karmadhāraya*.

Also *bahuvrīhi* have a similar structure but a different meaning: the suffix transforms the second member into an adjective, as in (H)¹⁰.

$$(H) \quad < [[X]_{N/Aa} [[y]_{Nb} + \text{Suf}]_A]_{Ac} \leftrightarrow [\text{one who possesses SEM}_b, \text{defined by SEM}_a]_c >$$

Although some such compounds exhibit visible suffixation, many involve 'zero' suffixes.

Other compounds have a nominal second member that undergoes no derivation, as in (I)¹¹.

$$(I) \quad < [[X]_{N/Aa} [y]_{N/Ab}]_{N/Ac} \leftrightarrow [\text{SEM}_b \text{ defined by SEM}_a]_c >$$

This type is less frequently attested: the definition of the *Compound- ing Parameter* explains why in Latin these compounds are less common in comparison, for instance, with the Germanic languages¹². This category may be further subdivided according to the morphological category (noun or adjective) of the two members.

⁹ See Oniga (1988: 113-116); Re (2020: 150-152).

¹⁰ See Oniga (1988: 116-127); Re (2020: 152-154).

¹¹ See Oniga (1988: 128-130); Re (2020: 154-155).

¹² See Snyder (2001; 2016).

Structurally similar to the preceding type (I) but semantically distinct are copulative compounds, in which the two members are linked by co-ordination¹³.

This typology is summarized in the Table 3. In the database, each main type is labeled with a number from 1 to 7; subtypes, where relevant, are identified by capital letters. These labels ensure consistent encoding and allow flexible querying based on both morphological structure and semantic value.

Second member	Derivation	Type	Subtype	Example
verb	yes	1 : <i>nomen agentis</i>	1A : Suf = <i>-a</i>	<i>agricola</i>
			1B : Suf = <i>-o/-a</i>	<i>aliger</i>
			1C : Suf = zero	<i>artifex</i>
			1D : Suf = <i>-nt</i>	<i>omnipotens</i>
			1E : Suf = <i>-t</i>	<i>sacerdos</i>
			1F : Suf = <i>-tor</i>	<i>vitisator</i>
			1G : Suf = <i>-ulo/a</i>	<i>damnigerulus</i>
			1G : Suf = <i>-āc</i>	<i>maledicax</i>
			1G : Suf = <i>-ida</i>	<i>oculicrepida</i>
			1G : Suf = <i>-bulo/a</i>	<i>nucifrangibulum</i>
			1G : Suf = <i>-(i)ōn</i>	<i>legirupio</i>
			1H : Suf = <i>-io/a</i>	<i>carnificius</i>
		2 : <i>nomen actionis</i>	2A : Suf = <i>-io</i>	<i>solstitium</i>
			2B : Suf = <i>-ia</i>	<i>vindemia</i>
			2C : Suf = <i>-īna</i>	<i>argentifodina</i>
noun	yes	1 : <i>nomen agentis</i>	1Z : Suf = <i>-io/a</i>	<i>fulclopedia</i>
	yes	3 : nominal abstract	3A : Suf = <i>-io</i>	<i>aequinocetium</i>
			3B : Suf = <i>-io</i>	<i>triennium</i>
			3C : Suf = <i>-ia</i>	<i>misericordia</i>
	yes	4 : <i>bahuvrīhi</i> (A+N)	4A : Suf = zero	<i>falsiparens</i>
			4B : Suf = <i>-i</i>	<i>bilinguis</i>
			4C : Suf = <i>-o/-a</i>	<i>tripectorus</i>
			4D : Suf = <i>-io/a</i>	<i>caldicerebrius</i>
	no	5 : <i>bahuvrīhi</i> (N+N)	Suf = zero	<i>auricomus</i>
		6 : determinative c.	6A : A+N	<i>semihora</i>
			6B : A+A	<i>mediterraneus</i>
			6C : N+N	<i>caprificus</i>
		7 : copulative c.	—	<i>duodecim</i>

Table 3: Types and subtypes of Latin nominal compounds and their main characteristics.

¹³ See Oniga (1988: 131-133); Re (2020: 155).

Two further categories comprehend Juxtapositions and Grecisms.

Juxtapositions (Ju) are phrases composed of two juxtaposed words with a strong syntactic boundary¹⁴. They include:

- Noun + Adjective: *respublica* “commonwealth, state”, *iusiurandum* “oath”, etc.;
- Noun + Noun in fixed case: *paterfamilias* “head of a family, householder”, *tribunus militum* “tribune of the soldiers”, etc.;
- Adverb + Participle: *benedictus* “blessed”, *malefactus* “evil-doing”, etc.

Grecisms (Gr) are compound words borrowed from Greek. They are generally formed by two Greek elements (e.g. *allophylus* “of another race, foreign”, *protoplastus* “the first man”); less frequent is the case in which a Greek and a Latin element are put together (e.g. *bacciballum* “a bit of stuff”, *bilychnis* “lamp having two lights”)¹⁵.

3.1. Collecting linguistic material

In light of the developed theory, we proceed to collect the linguistic material.

First, for each work examined, a Microsoft Excel spreadsheet is created with the following structure:

- cell A1 contains the Latin name of the author;
- cells B1 and C1 respectively contain the author’s century of birth and death¹⁶;
- cell A2 contains the full title of the work;

¹⁴ See Oniga (1988: 139-144); Re (2020: 156).

¹⁵ See Re (2020: 157); for hybrid compound, refer to Magni (2016; 2017).

¹⁶ Positive numbers refer to AD centuries and negative ones to BC centuries.

- cell B2 includes the author/work citation form, following the criteria of the *Thesaurus linguae Latinae*¹⁷;
- cells A3 and B3 contain the literary genre and any subgenre, respectively;
- from cell A6 downward: the alphabetical list of nominal compounds attested in the work;
- from cell B6 downward: the number of occurrences.

Figure 2 shows an example of this structure.

	A	B	C	D
1	Ammianus Marcellinus Antio	4	4	
2	<i>Rerum gestarum quae exstan</i>	AMM.		
3	Prose	Historiography		
4				
5	Compounds	Occurrences		
6	aedificium	15		
7	aequinoctialis	1		
8	aequinoctium	3		
9	aequisonus	1		
10	alienigenus	1		
11	armiger	7		
12	armipotentia	1		
13	artifex	14		
14	haruspicus	1		
15	auceps	1		
16	aucupium	1		
17	auspex	2		

Figure 2: Table of compounds attested in Ammianus Marcellinus.

The linguistic material included in these tables generally derives from printed concordances¹⁸. If not available, the list of word forms is obtained using TextStat, a cross-platform program developed by the Department for Dutch Linguistics at Freie Universität Berlin¹⁹: in such cases, the Latin text is taken from digital editions available online

¹⁷ The *conspectus siglorum* is available at this URL: <https://thesaurus.badw.de/en/tll-digital/index/a.html>.

¹⁸ Employed concordances are listed in Oniga (1988: 169); Re (2020: 388-390).

¹⁹ TextStat can be downloaded from this URL: <https://neon.niederlandistik.fu-berlin.de/en/textstat/>; for further information see Milanese (2020: 184-191).

through various databases, both open-access (*Perseus Project*, *PHI*, *Musisque Deoque*, *digilibLT*, *Corpus Corporum*, etc.) and subscription-based (Brepols' *Library of Latin Texts*).

Subsequently, the compounds of all the examined authors are inserted into another Excel file containing a twofold table.

The first one (*COMPOUNDS*, Figure 3) is structured as follows:

- column A contains the list of lemmata according the standard citation form used in Latin dictionaries (nominative singular);
- column B, the lexical category of the compound (verb, noun, adjective, numeral, adverb);
- columns C and D, respectively, the type and subtype of the compound;
- columns E and F, the first member and its lexical category (verb, noun, adjective, numeral, adverb);
- columns G and H, the same for the second member;
- columns I and J, likewise for a possible third member;
- columns K and L, likewise for a possible fourth member.

	A	B	C	D	E	F	G	H
	Compound	Lexical category of the compound	Type	Sub-type	First member	Lexical category of the first member	Second member	Lexical category of the second member
1								
2	acersecomes (ἀκερσεκόμης)	Noun	Grecism	Gr	ἀκερσος	Adjective	κόμη	Noun
3	acoenoetetus (ἀκονοετήτης)	Noun	Grecism	Gr	ἀκονοετός	Adjective	νοέω	Verb
4	acroama (ἀκρόαμα)	Noun	Grecism	Gr	ἀκρόα	Adjective	οὔς	Noun
5	acrufolius	Noun	Bahuvrihi	4A	acer	Adjective	folium	Noun
6	acupediis	Adjective	Bahuvrihi	4D	acer	Adjective	pes	Noun
7	aedifex	Noun	Nomen agentis	1C	aedes	Noun	facio	Verb
8	aedificium	Noun	Nomen actionis	2A	aedes	Noun	facio	Verb
9	aediliscurulis	Noun	Juxtaposition	Ju	aedilis	Noun	curulis	Adjective
10	aedilisplebei	Noun	Juxtaposition	Ju	aedilis	Noun	plebs	Noun
11	aedilisplebi	Noun	Juxtaposition	Ju	aedilis	Noun	plebs	Noun
12	aedilisplebis	Noun	Juxtaposition	Ju	aedilis	Noun	plebs	Noun
13	aeditimus	Noun	Nomen agentis	1B	aedes	Noun	tueor	Verb
14	aeditua	Noun	Nomen agentis	1A	aedes	Noun	tueor	Verb
15	aeditualis	Adjective	Nomen agentis	1B	aedes	Noun	tueor	Verb
16	aedituens	Noun	Nomen agentis	1D	aedes	Noun	tueor	Verb
17	aedituus	Noun	Nomen agentis	1B	aedes	Noun	tueor	Verb
18	aegisonus	Adjective	Nomen agentis	1B	aegis	Noun	sono	Verb
19	aenipes	Adjective	Bahuvrihi	4A	aeneus	Adjective	pes	Noun
20	aequaevus	Adjective	Bahuvrihi	4A	aequus	Adjective	aevum	Noun
21	aequanimis	Adjective	Bahuvrihi	4B	aequus	Adjective	animus	Noun
22	aequanimitas	Noun	Bahuvrihi	4A	aequus	Adjective	animus	Noun
23	aequanimiter	Adverb	Bahuvrihi	4A	aequus	Adjective	animus	Noun

Figure 3: Comprehensive table of compounds attested in the authors.

For simplicity, the compound members are given according the dictionary citation form. For Greek borrowings, the original Greek word is placed in round brackets next to the Latin lemma in column A.

The second table (*DUPLICATES*, Figure 4) consists of only two columns: A contains the lemma as in *COMPOUNDS* table, while B its spelling variant.

	A	B
1	Compound	Duplicate
2	aeditimus	aedituus
3	aequipar	aequiper
4	angiportus	angiportum
5	Apennincola	Appennincola
6	architecton	architectus
7	arcitenens	arquitenens
8	benevolens	benivolens
9	benevolentia	benivolentia
10	benevolus	benivulus
11	bidens	duidens
12	bipennis	bipinnis
13	domumitio	domuitio
14	dupondium	dupundium
15	forceps	formucapes
16	halophanta	halofanta
17	idololatra	idololatres
18	luniter	luniter

Figure 4: Table containing duplicates of nominal compounds.

The cross-referencing of these tables and the population of the database are managed by a Java script²⁰.

4. DATABASE IMPLEMENTATION

A graph database is a software system that supports CRUD operations (*Create, Read, Update, Delete*) on a graph-based data model. It enables the representation of a problem domain and its interrogation at runtime through specific queries.

²⁰ All the material used for creating the database is freely available through the GitHub repository at this URL: <https://github.com/AI4CH-UniUD/Genus-Compositium/>.

One of its main advantages over relational databases lies in the handling of complex relationships. In graph database, relationships are treated as first-class entities and stored directly within it, significantly enhancing query performance.

A second major advantage concerns schema flexibility: whereas relational databases require predefined and rigid schemas, graph database allow nodes, relationships, or properties to be added without altering a fixed schema. This adaptability is beneficial in contexts where the data model rapidly evolves.

Another notable strength is the intuitive nature of graph-based modeling: data structures can be shaped by directly mirroring real-world connections between entities.

For the implementation of the present database, Neo4j was adopted²¹.

The database can be freely accessed at this URL: <http://158.110.146.222:7475/>. Upon connection, users will find a web server interface, asking for the login data:

```
database = compostinominali  
username = nominalCompoundsPublic  
password = pub_Comp_Nom_25
```

4.1. Conceptual model

An Entity-Relationship model must be defined in order to conceptually represent the domain under investigation. This modeling approach employs entities, attributes, and relationships²². Entities refer to real-world ‘things’ that exist independently; each is characterized by attributes, i.e. specific properties that describe it; relationships denote the connections that one entity has with another.

²¹ The software is accessible at this URL: <https://neo4j.com/>.

²² See Elmasri – Navathe (2016).

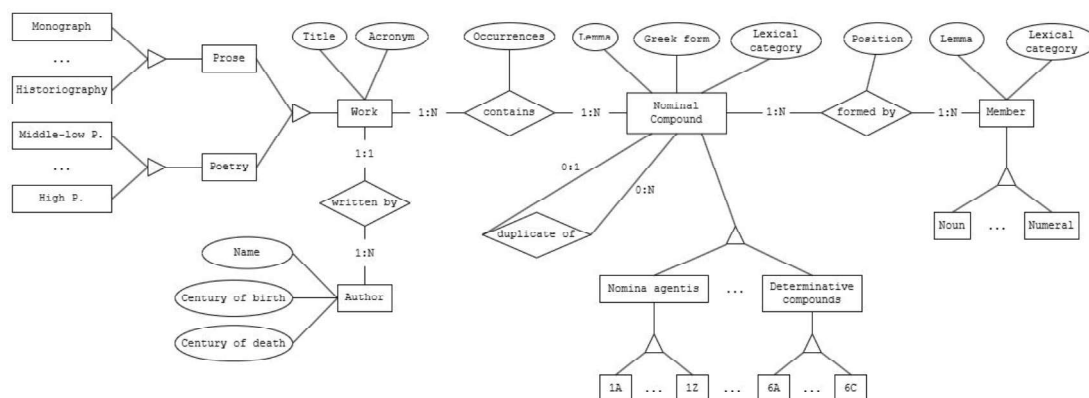


Figure 5: Schema of the ER model of the graph database.

The entity *Nominal compound* has the attributes *Lemma*, *Lexical category*, and *Greek form*; it is related to itself through the relationship *duplicate of*, which is given only in cases where duplicates exist.

The entity *Member* shares the same attributes as the previous.

These entities are connected by the relationship *formed by*, which includes the attribute *Position*, indicating the place each *Member* has within the *Nominal compound*.

To the left, the entity *Work* is characterized by the attributes *Title* and *Acronym*. The relationship *contains* links *Work* and *Nominal compound*, signifying that the latter appears in a specific text.

The entity *Author* has the attributes *Name*, *Century of birth*, and *Century of death*.

The relationship *written by* connects *Work* and *Author*.

On the far left, literary *Genre* and *Subgenre* of each *Work* are represented.

4.2. Querying the database

The following Cypher queries can be applied to the graph database in order to investigate the link between nominal compounds and their occurrences in the *corpus*.

4.2.1. Which works and authors attest a given compound?

Investigating which works attest a given compound is done by matching nominal compounds via the CONTAINS relationship to a work linked to an author through the WRITTEN_BY relationship, requiring that the *lemma* of the nominal compound matches the value specified in line 5 of this listing. The results are shown in a table and alphabetically sorted by author and work.

```
MATCH
  (c:NominalCompound)<-[r:CONTAINS]-(w:Work),
  (w)-[wr:WRITTEN_BY]->(a:Author)
WHERE
  c.lemma='aliger' // Compound
RETURN
  c.lemma AS Compound,
  w.acronym AS Author_Title,
  r.occurrences AS Occurrences,
  c.type AS Type,
  c.subtype AS Subtype
ORDER BY
  tolower(w.acronym)
```

The compound *aliger* is a *nomen agentis* (1B) typically used in high poetry, starting with Virgil (*Aen.* 1, 663. 12, 249); this model is then adopted by various 1st-century AD authors (Lucan 9, 684; Val. Fl. 2, 224. 5, 453. 7, 120. 7, 171; Sil. 2, 92. 3, 169. 7, 458. 14, 507; Stat. *Theb.* 2, 1. 5, 599. 10, 302; Stat. *silv.* 3, 3, 80) and subsequently in late antiquity by Ausonius (*Mos.* 300). The only attestations in prose are found in Pliny's *Naturalis historia* (2, 17. 12, 85. 36, 41).

As a root-based *nomen agentis* (1C), *particeps* is not as semantically marked as *aliger*. Its earliest attestation is in Livius Andronicus (*trag.* 3 Ribbeck³) but, from the 1st century BC onward, it appears predominantly in prose, whereas its use in poetry nearly disappears, with the exception of Lucretius (3, 462). This trend continues in the subsequent centuries.

From the 1st century BC onward, *alipes* is a *bahuvrīhi* (5) characteristic of high poetry. The earliest attestations are found in Lucretius' *De rerum natura* (6, 765) and in the *Aeneid* (7, 277. 12, 484), with greater frequency among 1st-century AD poets (Val. Fl. 5, 183. 5, 611. 6, 208; Sil. It. 3, 292. 7, 700. 15, 554; Stat. *Theb.* 3, 428. 4, 351. 5, 699. 6, 298. 9, 206. 9, 590. 11, 198). Later occurrences are found in Ausonius (*ecl.* 4, 5 = 22, 5 Green; *epit.* 33, 8).

4.2.2. Which compounds share the first member?

This query retrieves all nominal compounds that share a specific first member (line 6), identifying all compounds in which the FORMED_BY relation is assigned to the property "position=1" and matching them with other compounds exhibiting the same condition.

```
MATCH
  (c:NominalCompound)-[r1:FORMED_BY]->(m1:Member),
  (c)-[r2:FORMED_BY]->(m2:Member)
WHERE
  r1.position=1
  AND m1.lemma='aequus' // First member
  AND r2.position=2
RETURN
  c.lemma AS Compound,
  m1.lemma AS First_Member,
  m2.lemma AS Second_Member,
  c.type AS Type,
  c.subtype AS Subtype
ORDER BY
  toLower(c.lemma)
```

The adjective *aequus* combines with nine different second members, most of which are names (*aevum*, *animus*, *dies*, *latus*, *manus*, *nox*, *par*, *pes*, *sonus*): the majority of compounds it forms is *bahuvrīhi*, where the notion of “equality” is predicated of the referent denoted by the second member.

By reapplying the preceding query, one may also explore the literary contexts in which these forms occur, confirming that such compounds are characteristic of high poetry. However, more neutral lexical items are not absent: for instance *aequinoc-tium* and its derivative *aequinoc-tialis* are generally attested in prose. The only compound with a verbal second member is *aequilavium* (“a half of the whole, said of wool, when half of the weight remains after washing”), a *nomen actionis* (2A), attested only in the grammarian Festus (p. 22, 24 Lindsay).

4.2.3. Which compounds share the second member?

This query aims to find all nominal compounds that have the same second member (line 6): similar to the previous, in the FORMED_BY relationship the property "position=2" is specified.

```
MATCH
  (c:NominalCompound)-[r1:FORMED_BY]->(m1:Member),
  (c)-[r2:FORMED_BY]->(m2:Member)
WHERE
  r2.position=2
  AND m2.lemma='facio' // Second member
  AND r1.position=1
RETURN
  c.lemma AS Compound,
  m1.lemma AS First_Member,
  m2.lemma AS Second_Member,
  c.type AS Type,
  c.subtype AS Subtype
ORDER BY
  toLower(c.subtype)
```

The verb *facio* combines with a large number of first members (nouns and adjectives): among these, *nomina agentis* of subtype 1B are used predominantly in poetry; by contrast, *nomina agentis* of subtype 1C and *nomina actionis* are generally prosaic, with very few occurrences in high poetry.

Repeating the same query with the noun *pes*, a very different situation emerges. The vast majority of compounds are *bahuvrīhi*, whose use – just as previously noted for *alipes* – is characteristic of epic poetry. Among the exceptions, however, there are some cases worth additional comment.

Semipes is one of many nominal compounds (6A) characterized by the first member *semi-*, an adjective that ‘degraded’ into a prefix to indicate the concept of ‘half’, and it combines with a large number of second members: although it is typically prosaic (e.g. 7 occurrences in Cato’s *De agri cultura*), it also appears in Ausonius (*cento*, p. 133, 30, 31 Green; *epist.* 12, 87) and Prudentius (*perist.* 2, 150).

In *fulcipedia* (Petron. 75, 6) the verbal element derived from the verb *fulcio* precedes the nominal one: this is likely the first attestation in Latin of a type of nominal compound that would later become widespread in the Romance languages (cfr. *poggiapiedi* ‘footrest’, *asciugacapelli* ‘hairdryer’, etc.).

4.2.4. Which types and subtypes recur according to work, author, and genre?

These queries aim to investigate the types and subtypes of nominal compounds in order to determine whether there exist ‘linguistic constants’ showing a correlation between literary genre and the use of nominal compounds.

This listing queries the database to display which types of nominal compounds are attested in Statius’ *Thebaid* (line 4): the results are sorted from the most frequent to the least frequent subtype.

```
MATCH
  (w:Work)-[r:CONTAINS]->(n:NominalCompound)
WHERE
  w.acronym='STAT. Theb.' // Acronym of the work
```

```

RETURN
  n.type AS Type,
  n.subtype AS Subtype,
  COUNT(*) AS Occurrences
ORDER BY
  Occurrences DESC,
  n.subtype

```

In an epic poem, the preferred compositional types are *nomina agentis* of subtypes 1B and 1A, along with *bahuvrīhi*. These data confirm the situation previously outlined and provide preliminary evidence that the use of ‘marked’ forms such as nominal compounds is not arbitrary, but rather stems from a deliberate expressive intention by the author.

If we analyze Tacitus’ *Annals*, results are quite dissimilar in comparison with the previous ones: the most widespread compounds are *nomina actionis* (particularly 2A), radical *nomina agentis* (1C), juxtapositions, and nominal abstracts; *nomina agentis* of subtype 1B are less frequent, as well as *bahuvrīhi*. This demonstrates that historiography employs a lexicon radically different in comparison with that of epic poetry. Beyond the individual stylistic choices of each author, a clear division is evident between prose and poetry.

This listing investigates the types of nominal compounds attested in Virgil (line 5).

```

MATCH
  (w:Work)-[r:CONTAINS]->(n:NominalCompound),
  (w)-[r2:WRITTEN_BY]->(a:Author)
WHERE
  a.name = 'P. Vergilius Maro' // Name of the author
RETURN
  n.type AS Type,
  n.subtype AS Subtype,
  COUNT(*) AS Occurrences
ORDER BY
  Occurrences DESC,
  n.subtype

```

Despite the variety of Virgil's *opera omnia*, results confirms a situation not so dissimilar to that previously described for Statius' *Thebaid*. *Nomina agentis* and *bahuvrīhi* are the most frequent: among the former, subtype 1B predominates, while 1C, 1A, and 1D are much less frequent; among the latter, there is greater heterogeneity, though subtypes 4A and 4B show a slight predominance over the others. In this regard, one can reasonably state that Statius' *Thebaid* clearly positions itself within the epic literary tradition: thus, we have further confirmation of what traditional philology states about the persistence of certain expressions characteristic of this literary genre.

The next queries pertain to literary genres and subgenre (line 4).

```
MATCH
  (w:Work)-[r:CONTAINS]->(n:NominalCompound)
WHERE
  w.subgenre='Historiography' // Subgenre
RETURN
  n.type AS Type,
  n.subtype AS Subtype,
  COUNT(*) AS Occurrences
ORDER BY
  Occurrences DESC,
  n.subtype
```

By cross-referencing the results about Historiography with those referring to Tacitus' *Annals*, one can assert that these belong to the historiographical genre not only for thematic content but also in terms of linguistic features. The most frequent types are *nomina actionis* (particularly 2A), radical *nomina agentis* (1C), and juxtapositions; other types are less frequent.

On the contrary, the distribution of nominal compound subtypes in the Latin Novel differs significantly in comparison with historiography. Despite the marked thematic and lexical differences between Petronius' *Satyricon* and Apuleius' *Metamorphoses*, it is evident that the language of these two authors diverges notably from that of historiography. The

prevalence of Greek borrowings and thematic *nomina agentis* (1B) clearly contrasts with the situation previously described.

5. CONCLUSION

This database extracts various types of information concerning the lexical choices of Latin authors, such as occurrences of individual entries (§4.2.1), elements involved in the formation of compound nouns (§§4.2.2-4.2.3), types and subtypes of compounds employed either in individual works, or across the entire *corpus* of authors, or within the various genres and subgenres (§4.2.4)²³.

We believe that this tool will increasingly allow us to explore in detail how morphological constructions determine not only the form of compounds, but also anticipate their stylistic value. Although morphological constructions are abstract entities that exist only in the speakers' mind, their effect has practical stylistic consequences, clearly observable in Latin authors. Thus, it is not enough to state that *nomina agentis* and *bahuvrīhi* are compounds more common in high poetry, while *nomina actionis* or nominal abstract are typically prosaic: morphological constructions imply that, even within the same type, certain groups are perceived as more refined and so more suitable for epic poetry, while others carry a less solemn connotation, making them more fitting for prose.

Language and style are therefore not two separate levels, but rather support one another through their mutual interactions.

Università degli Studi di Udine
andrea.brunello@uniud.it
alessandro.re@uniud.it
torreggio@gmail.com

²³ Further queries may be developed and will be made public through the GitHub repository mentioned in note 20. The syntax of Cypher queries is outlined in Robinson, Webber, and Eifrem (2015).

REFERENCES

Aronoff, M.

1976 *Word Formation in Generative Grammar*, M.I.T. Press, Cambridge (MA).

1983 *A Decade of Morphology and Word Formation*, in «Annual Review of Anthropology», 12, pp. 355-375.

1994 *Morphology by Itself. Stems and Inflectional Classes*, M.I.T. Press, Cambridge (MA).

Benedetti, M.

1988 *I composti radicali latini. Esame storico e comparativo*, Orientamenti Linguistici 22, Giardini, Pisa.

Booij, G.

2015 *Word-Formation in Construction Grammar*, in Müller, P.O. – Ohnheiser, I. – Olsen, S. – Rainer, F. (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 40/1, De Gruyter Mouton, Berlin – München – Boston, vol. 1, pp. 188-202, <https://doi.org/10.1515/9783110246254-014>.

Elmasri, R. – Navathe, S.B.

2016 *Fundamentals of Database Systems*, Pearson, Boston.

Fruyt, M.

1996 *Noms de procès en latin: Évaluation des positions benvenistiennes dans “Noms d’agent et Noms d’action en Indo-Européen”*, in Rosén, H. (ed.), *Aspects of Latin. Papers from the Seventh International Colloquium on Latin Linguistics (Jerusalem, April 1993)*, Institut für Sprachwissenschaft der Universität Innsbruck, Innsbruck, pp. 193-208.

Leumann, M.

1977 *Lateinische Laut- und Formenlehre*, Lateinische Grammatik 1, Beck, München.

Lindner, T.

1996 *Lateinische Komposita. Ein Glossar, Vornehmlich zum Wortschatz der Dichtersprache*, Innsbrucker Beiträge zur Sprachwissenschaft 89, Institut für Sprachen und Literaturen der Universität Innsbruck, Innsbruck.

2002 *Lateinische Komposita. Morphologische, historische und lexikalische Studien*, Innsbrucker Beiträge zur Sprachwissenschaft 105, Institut für Sprachen und Literaturen der Universität Innsbruck, Innsbruck.

Lindner, T. – Oniga, R.

2005 *Zur Forschungsgeschichte der Lateinischen Nominalkomposition. Per una storia degli studi sulla composizione nominale latina*, in «Papers on Grammar», 9/1, pp. 149-160, <https://doi.org/10.1515/joll.2005.9.1.149>.

Magni, E.

2016 *Notha Verba: Interferenza e mutamento attraverso le formazioni ibride latino-greco*, in «Linguarum Varietas», 5, pp. 133-144.

2017 *Suffix borrowing and conflict through Latin-Greek hybrid formations*, in «Pallas», 103, pp. 283-291.

Milanese, G.F.

2020 *Filologia, Letteratura, Computer. Idee e strumenti per l'informatica umanistica*, Vita & Pensiero, Milano.

Oniga, R.

1988 *I composti nominali latini: una morfologia generativa*, Pàtron, Bologna.

1989 *Morphological Theory and Latin Morphology*, in Lavency, M. – Longrée, D. (eds.), *Actes du V Colloque de Linguistique Latine. Proceedings of the Vth Colloquium of Latin Linguistics*, Cahiers de l'Institut de Linguistique de Louvain, Louvain, pp. 333-344.

1990 *La composizione nominale in Sallustio*, in «Lexis», 5-6, pp. 147-196.

1992 *Compounding in Latin*, in «Rivista di Linguistica», 4, pp. 97-116.

1994 *Morfologia, sintassi e semantica nella composizione nominale latina*, in «Aufidus», 23, pp. 81-98.

2000 *La création lexicale chez Pétrone*, in Fruyt, M. – Nicolas, C. (eds.), *La création lexicale en Latin. Actes de la table ronde du IXème Colloque International de Linguistique Latine organisée par Michèle Fruyt à Madrid le 16 Avril 1997*, Presses de l'Université de Paris-Sorbonne, Paris, pp. 155-166.

2002 *La formazione delle parole per composizione in latino*, in «Paideia», 57, pp. 340-361.

2005 *Composition et préverbation en Latin: problèmes de typologie*, in Moussy, C. (ed.), *La composition et la préverbation en latin*, Presses de l'Université de Paris-Sorbonne, Paris, pp. 211-227.

2020 *Latin Synthetic Compounding and Distributed Morphology*, in «Lingue Antiche e Moderne», 9, pp. 205-224, <https://doi.org/10.4424/lam92020-9>.

Oniga, R. – Re, A.

2017 *L'analyse synchronique des composés nominaux du latin hier et aujourd'hui*, in «L'Antiquité Classique», 86, pp. 39-58.

2025 *Between Linguistics and Stylistics: Nominal Compounding and Literary Genre in Latin*, in Pultrová, L. – Vaníková, M. (eds.), *Exploring Latin: Structures, Functions, Meaning*, De Gruyter, Berlin – Boston, vol. I, pp. 121-138, <https://doi.org/10.1515/9783111332956-008>.

Re, A.

2020 Genus compositivum. *La composizione nominale latina*, Innsbrucker Beiträge Zur Sprachwissenschaft 165, Institut für Sprachwissenschaft der Universität Innsbruck, Innsbruck.

2021 *L'uso dei composti nominali nel Panegyricus Traiano Imperatori di Plinio il Giovane*, in «Paideia», 76, pp. 521-553, <https://doi.org/10.1400/285698>.

Robinson, I. – Webber, J. – Eifrem, E.

2015 *Graph Databases*, O'Reilly, Beijing – Cambridge – Farnham – Köln – Sebastopol – Tokyo.

Snyder, W.

2001 *On the Nature of Syntactic Variation: Evidence from Complex Predicates and Complex Word-Formation*, in «Language», 77/2, pp. 324-342, <https://doi.org/10.1353/lan.2001.0108>.

2016 *How to set the Compounding Parameter*, in Perkins, L. – Dudley, R. – Gerard, J. – Hitczenko, K. (eds.), *Proceedings of the 6th Conference on Generative Approaches to Language Acquisition – North America*, Cascadia Proceedings Project, Somerville (MA), pp. 122–30, <http://www.lingref.com/cpp/galana/6/abstract3212.html>.