Annotation guidelines for morphological and morphosyntactic annotation of Merovingian Latin

Reference document for the Latin corpus *PaLaFraLat*

Version 1.2

Elisabeth Reichle

October 2018

Version history

Contents

I. Introduction

The guidelines presented here describe the general morphological and morphosyntactic annotation principles of the Merovingian Latin corpus PaLaFraLat which was developed in the framework of the French-German DFG-ANR-funded project *PaLaFra* (Le PAssage du LAtin au FRAnçais – www.palafra.org) (2015-2018). One of the goals of the corpus project was the development of a morphosyntactically annotated multi-genre corpus which is regionally and temporally limited to the Merovingian Latinity providing fine-grained investigations on the transition from Medieval Latin to Old French.

Morphosyntactic annotation provides the relevant basis for further explorations such as a syntax layer, or semantic or pragmatic analyses of annotated Latin corpora. The lemmatization scheme and the tagset of the corpus (*lapos*) concerning part of speech (PoS) and full morphological tagging is – with some modifications[1] – based on the *CompHistSem* (cf. *CompHistSem* 2016) tagset. Most of the available tagsets for Classical or Medieval Latin are based on the grammar of Classical Latin – like the morphological tagsets of the *Index Thomisticus* Treebank (IT-TB), *Latin Dependency Treebank* (LDT) or *Pragmatic Resources of Old Indo-European Languages* (PROIEL)[2]. As Merovingian texts differ from Classical Latin in orthographical, lexical, morphological and syntactical respects (Stotz 1998, Väänänen [3]1981), our main objective was to adapt the annotation to Merovingian Latin texts.

II. Lemmatization

Lemmatization reduces every single word form – including inflected forms and different spelling variants – to a corresponding canonical lemma (cf. Jenset/McGillivray 2017:114). Merovingian Latin is extremely fluid and subject to continuous change. Even though there are tendencies toward new language 'norms', Merovingian Latin lacks an orthographical standard (cf. Buchner et al. 2017:137). Due to the absence of a homogeneous orthographical norm as well as practical reasons we reduce all word forms to a familiar Classical Latin dictionary entry (*classicallemma*) which makes corpus queries based on well-known lexemes possible. The classical orthography of Classical Latin lexemes and forms (*classicallemma* and *classicalword*) is added according to the lexicon entry of the Latin-German lexicon of Georges/Georges (1913/1918) which is available online.[3] Pure medieval word forms (e.g., *grafionem*, *graffione*) are reduced to a medieval '*classicallemma*' (e.g., *grafio*) according to the lexicon entries of the

---

[1] use (ADJECTIVAL, SUBSTANTIVAL), GREEK_DECLENSION; *classicalword*.
[2] There are no morphological guidelines for Latin treebanks.
[3] http://www.zeno.org/Georges-1913 (last visited October 2018).

Glossarium Mediae et Infimae Latinitatis (cf. Du Cange et al. 1883-1887)[4], the *Mediae latinitatis Lexicon minus* (cf. Niermeyer/Van de Kieft/Burgers ²2002) or the *Mittellateinisches Wörterbuch* (MLW)[5].

The system of lemmatization presented here accounts for the variability of Merovingian Latin with a complex hierarchy of different abstraction levels, precisely a *token*, *word*, *lemma*, *classicallemma* and a newly introduced *classicalword* level.

- **token**: Tokens are strings of contiguous characters, punctuation marks or numbers without any normalization. Tokens contain case (e.g., *INCIPIT*), editors' conjectures (e.g., *aec*(*lisiae*), [*ecclesia*]) and abbreviations (e.g., *ann*) as they occur in the edition of the *dMGH* (https://www.dmgh.de/). Thus, by displaying the token properties in a concordance of a given query, the lexical units are displayed as they appear in the edition.

- **word**: Words are full word forms in lowercase (apart from proper names or named entities) (e.g., *incipit*), including editors' emendations (e.g., *aeclisiae*, *ecclesia*) and expanded abbreviations (e.g., *annos*). A query on word level matches every occurrence of this word, including those containing uppercases, abbreviations and conjectures. *Word* is the default query form in PaLaFraLat.

- **lemma**: All inflected forms of a certain word stem (cf. *word*) are assigned to a lemma which preserves a certain spelling variant in its stem, for example *incipio*, *annus*, *aeclisia* (for *aeclisiae*) and *ecclesia* (for *ecclesia*). Verbal word forms are assigned to the first-person singular of the indicative active present of a word variant, inflected nominal, adjectival and pronominal word forms are assigned to the nominative masculine singular of the word variant. A query on lemma level for *aeclisia* matches all inflected forms of this specific spelling variant (e.g., *aeclisiae*, *aeclisia*, *aeclisiarum*), but not *ecclesia*, *ecclesiae* or *ecclesiarum.*

- **classicallemma**: The *classicallemma* is a lemma with Classical Latin spelling uniting all spelling variants in the stem, inflected word forms and variations in declension, conjugation, gender and voice (cf. *word*), for example *incipio*, *annus*, *ecclesia* (for *aeclisiae* and *ecclesia*). Verbal word forms are assigned to the first-person singular of the indicative active present, inflected nominal, adjectival and pronominal word forms are assigned to the nominative masculine singular of the Classical Latin lemma. A query

---

[4] http://ducange.enc.sorbonne.fr/ (last visited October 2018).
[5] https://woerterbuchnetz.de/?sigle=MLW#0 (last visited October 2018).

on *classicallemma* level for *ecclesia* matches all inflected forms and medieval spelling variants of *ecclesia* (e.g., *ecclesiam*, *eclisias*, *ecclesie*, *aecclesiarum*, *aecclesie*, *aeclesiam*, *aeclisia etc.*).

- **classicalword**: *Classicalwords* are full wordforms with Classical Latin spelling of non-Classical Latin words (cf. *word*), e.g., *incipit*, *annos*, *ecclesiae* (for *aecclisae*). A query on *classicalword* level matches all spelling variants and grammatical variations (e.g., changes in declension, conjugation, gender and voice) of a certain word form.

The predominant principle of lemmatization is that all word forms are converted to their base form (*lemma* and *classicallemma*), which is the nominative singular for nouns, the nominative masculine singular for pronouns, adjectives, ordinal and distributive numbers, and the first-person singular of indicative active present for verbs. Pluralia tantum such as *liberi* or *Kalendae* keep the plural form and are not reduced to a non-existent singular form, e.g., *word*: *Kalendae*, *lemma*/*classicallemma*/*classicalword*: *Kalendae*. Comparatives and superlatives are redressed to their neutral base forms (e.g., *word*: *maior*, *cassicallemma*: *magnus*), gerunds and participles to their first-person singular verb form. Adverbs, adpositions, conjunctions, and interjections retain their original form.

III. Morphosyntactic annotation

The morphosyntactic annotation involves part-of-speech tagging (PoS-tagging) and annotation of other morphological features. The choice of the features depends on the context. The *lapos* tagset is structured in different fields: 15 core parts of speech categories and 13 different morphological features with 65 subcategories.

Part of speech tags (pos)
1. ADJ (adjective)
2. ADV (adverb)
3. NUM (cardinal number)
4. CON (conjunction)
5. DIST (distributive number)
6. FM (foreign material)
7. ITJ (interjection)
8. XY (non word)

9. NN (normal noun)

10. ORD (ordinal number)

11. NP (personal name)

12. AP (adposition)

13. PRO (pronoun)

14. NE (named entity)

15. V (verb)

Morphological features (+ subcategories)

1.	casus	NOMINATIVE
		GENITIVE
		DATIVE
		ACCUSATIVE
		VOCATIVE
		ABLATIVE
		LOCATIVE
		OBLIQUE
		INDECLINABLE
2.	comparisondegree	POSITIVE
		COMPARATIVE
		SUPERLATIVE
3.	conjugationtype	FIRST_ CONJUGATION
		SECOND_ CONJUGATION
		THIRD_CONJUGATION
		FOURTH_CONJUGATION
4.	declensiontype	FIRST_DECLENSION
		SECOND_DECLENSION
		THIRD_DECLENSION
		FOURTH_DECLENSION
		FIFTH_DECLENSION

INDECLINABLE

GREEK_DECLENSION

5. genus   MASCULINE

FEMININE

NEUTER

6. mood   GERUND

GERUNDIVE

IMPERATIVE

INDICATIVE

INFINITIVE

PARTICIPLE

SUBJUNCTIVE

SUPINE

7. numerus   SINGULAR

PLURAL

8. person   FIRST_PERSON

SECOND_PERSON

THIRD_PERSON

9. pronountype   CORRELATIVE

DEMONSTRATIVE

INDEFINITE

INTENSIVE

INTERROGATIVE

PERSONAL

POSSESSIVE

REFLEXIVE

RELATIVE

10.	tense	FUTURE
		FUTURE PERFECT
		IMPERFECT
		PERFECT
		PLUPERFECT
		PRESENT

11.	verbtype	DEPONENT
		IMPERSONAL
		INTRANSITIVE
		SEMIDEPONENT
		TRANSITIVE
		VERBA_ANOMALA
		VERBA_DEFECTIVA

| 12. | voice | ACTIVE |
| | | PASSIVE |

| 13. | use | ADJECTIVAL |
| | | SUBSTANTIVAL |

IV. Challenges in annotation of Medieval Latin texts

Orthographic variation

The main challenge concerning annotation of medieval texts is the orthographic variation of words, which often affects the case-endings of words. The *classicallemma* and the *classicalword* take all different types of spelling deviations into account by normalizing the non-Classical spelling variations in the stem and the word endings according to the classical norm (see II).

In Classical Latin, a special word form can be assigned to a special function. In Medieval Latin, however, the equivalence between form and function is often not transparent as a result of phonological and morphological slips (Väänänen [3]1981). Orthographical variations which follow common morphological and phonological changes from Latin to the Roman languages and which are clearly related to the corresponding Classical Latin word form used in the same

function, are annotated according to their functions in the text (cf. Korkiakangas/Passarotti 2011).

- casus=NOMINATIVE: *Chrodoaldus* […] *ditatos* (for Classical Latin *Chridoaldus* […] *ditatus*)
- casus=GENITIVE: *maiores domus* (for Classical Latin *maioris domus*)
- casus=DATIVE: *virtute nostrae non evadit* (for Classical Latin *virtuti nostrae non evadit*)
- casus=ACCUSATIVE: *altercationis* (for Classical Latin *altercationes*), *ducis* (for Classical Latin *duces*)
- casus=ACCUSATIVE: *aureus* (for Classical Latin *aureos*), *infra murus* (for Classical Latin *infra muros*)
- casus=ABLATIVE: *de gente nobile* (for Classical Latin *de gente nobili*), *de perfecta quieti* (for Classical Latin *de perfecta quiete*)

All missing final *-m* of Medieval Latin accusative forms are labelled as casus=OBLIQUE:

- casus=OBLIQUE: *post anno tercio* (for Classical Latin *post annum tertium*), *contra Adalulfu* (for Classical Latin *contra Adalulfum*), *ad baseleca* (for Classical Latin *ad basilicam*), *Franci vero Leudesio, filio Erchonoldo, nobile in maiorum domatum statuunt* (for Classical Latin *Leudesium, filium Erchonoldum, nobilem*).

Deviations in gender, declension, conjugation, voice

If there is no relation between the Medieval Latin word form and the functionally required Classical Latin form, the Medieval Latin word form is given a purely formal annotation (according to the form as it appears in the text) (cf. Korkiakangas/Passarotti 2011) and is assigned to a non-Classical deviation (see 17).

We count among grammatical variations in Medieval Latin texts for example non-Classical Latin case use (*Iudaeus quidam Priscus* **nomen** for *nomine*), non-Classical Latin use of gender (e.g., *iudicius* for *iudicium*), mode (e.g., *ut*/*ne*/*si* clauses or indirect speech without subjunctive), voice (e.g., *coortaret* for *cohortaretur*; *provocaret* for *provocaretur*; *persequere* for *persequi*), number (e.g., *omnis populus ibidem commanentes*), declension (e.g., *villabus*, *certaminem* for *villis, certamen*), conjugation (e.g., *cadebis* for *cades*) and the use of absolute constructions such as accusative absolute or nominative absolute.

Word forms which show changes of case, gender, voice, declension and conjugation are reduced to the corresponding Classical Latin lemma and form (*classicallemma, classicalword*),

even if, for example, different conjugations, voices or declensions are grouped under one *classicallemma* (cf. Philippart De Foy 2012), e.g., *cadebis* (*lemma: *cadeo*; *classicallemma*: *cado*, *classicalword*: *cades*), *thesauribus* (*lemma*: *thesaurus*, *classicallemma*: *thesaurus*, *classicalword*: *thesauris*), *coortaret* (*lemma*: *coorto*, *classicallemma*: *cohortor*, *classicalword*: *cohortaretur*), *persequere* (*lemma*: *persequo*, *classicallemma*: *persequi*, *classicalword*: *persequor*).

Infinitives of Latin deponent verbs which end in *-re* as opposed to the Classical Latin form *-ri* (e.g., *largire*, *imitare*, *admirare* for *largiri*, *imitari*, *admirari*) are considered orthographic variations of the Classical Latin ending *-ri* (Bonnet 1968:403), therefore they are not annotated as changes of conjugation (*lemma*: *largior*, *classicallemma*: *largior*, *classicalword*: *largiri*).

Personal names (NP)

For personal names, it is difficult to determine correct and normalized Classical Latin spellings because of the wide range of spelling variations throughout the whole corpus and also within a single text (e.g., *Sigebertus*, *Siggibertus*, *Sighibertus*, *Sigibertus*, *Sigyberthu*, *Sigybertus*, *Sygiberthus*, *Sygibertus*). Likewise, there are many personal names of Germanic origin (e.g., *Waratto*, *Richarid*, *Vulfos*) without any established standard Latin spelling. In addition, etymologically identical personal names appear in different declension types in the corpus (e.g., *Drocus*, *i* vs. *Drogo*, *onis*; *Radegunda* vs. *Radegundis*; *David*, *idis* vs. *David* indecl.). For these reasons, it is difficult to decide which spelling variant is more reasonable or justified than the other and to determine a standardised *classicallemma* and *classicalword*. Since there is a lack of standard forms for personal names, we created some general lemmatization-guidelines to guarantee at least a uniform lemmatization of names throughout the corpus:

- The normalised form of the prefix *Vulf-* is ***Wulf-*** (e.g., *Vulfoaldus* → *Wulfowaldus*). The spelling variations *-ulfus/-vulfus* are subsumed under the suffix *-**wulfus*** on the *classicallemma* level (e.g., *Berulfus* → *Berwulfus*).
- The normalised spelling of the suffixes *-oaldus/-valdus/-aldus* is *-**waldus*** (e.g., *Ansovaldus/Ansoaldus* → *Ansowaldus*).
- The normalised spelling of *-ghisil-/-gisil-* is *-**g**isil-* (without *h*) (e.g., *Godeghiselus* → *Gode**g**isilus*).
- The standardised suffix of *-vera* is *-**wera*** (e.g., *Audovera* → *Audowera*).

- The normalised form of *Berth-/Bert-* is *Bert-* (e.g., *Berthefredus/Berthefredus* ➔ *Bertefridus*); the suffixes *-berthus/-bertus* are subsumed under the suffix *-bertus* (without *h*).

- The suffixes *-vechus/-vius/-veus* are subsumed under *-wicus* on *classicallemma* level (e.g., *Merovechus/Merovius* ➔ *Merowicus*).

- The standardised form of *Fred-* is *Frid-*. The normalised suffix of *-fredus* is *-fridus* (e.g., *Ermenfredus* ➔ *Ermenfridus*).

- Personal names of Greek origin which follow the Greek declension are labelled as declensiontype=GREEK_DECLENSION (e.g., *Iesus*, *Aristarchos*).

V. Principles of annotation

1. Definition of the tagset

The annotation concerns the parts of speech and morphological features. The PoS categories correspond with the classic parts of speech. Nine different inflected or variable parts of speech are distinguished (NN, NE, NP, ADJ, PRO, NUM, DIST, ORD, V) and seven uninflected or invariable parts of speech (ADV, AP, CON, PTC, ITJ, XY, FM). In principle, each word form is functionally labelled with its specific morphological features. Depending on the context, the adjectival and substantival use is distinguished for ADJ, PRO, NUM, DIST and ORD.

2. ADJ: adjective

2.1 Morphological features

Adjectives modify in six different morphological categories: **casus**, **genus**, **numerus**, **declensiontype**, **comparisondegree** and **use**.

2.2 casus, genus, numerus

Adjectives' endings (such as nouns, pronouns and participles) alter to show grammatical case. Non-Classical Latin accusative forms (apocope of the final *-m*) are classified into the category OBLIQUE. Adjectives are annotated with their gender (MASCULINE, FEMININE or NEUTER and number (SINGULAR, PLURAL).

morphological feature	subcategory	example
casus	NOMINATIVE	*bonus, omnis*
	GENITIVE	*atrocis*
	DATIVE	*acri*
	ACCUSATIVE	*veterem*
	VOCATIVE	*grave*
	ABLATIVE	*aspero*
	OBLIQUE	*bona(m)*
	INDECLINABLE	*frugi, nequam*
genus	MASCULINE	*liber*
	FEMININE	*misera*
	NEUTER	*tenerum*
numerus	SINGULAR	*vetus*
	PLURAL	*atroces*

2.3 declensiontype

There are first (*a* stems), second (*o* stems) and third declension (consonant stems) adjectives. Adjectives in the first group are declined like nouns of the *a*- and *o*-declension, whereas adjectives of in the second group are declined like nouns of the third declension.

morphological feature	subcategory	example
declensiontype	FIRST_DECLENSION	*usque ad* **vespertina** *ora* (Visio Baronti)
	SECOND_DECLENSION	**beatus** *Petrus* (Visio Baronti)
	THIRD_DECLENSION	*Quem* **gravis** *molestia pene usque ad mortem perduxerat* [...]. (Visio Baronti) e.g., *frugi, nequam*
	INDECLINABLE	*Recedite, spiritus* **nequam***;* (Visio Baronti)

2.4 comparisondegree

An adjective is annotated with its degrees of comparison: the category POSITIVE is the form of the adjective as given in the dictionary, the COMPARATIVE, indicating an increased quality or quantity, and the SUPERLATIVE indicates the utmost degree.

→ […] *fratres **karissimi** […].* (Viso Baronti)

→ *Deinde ad locum designatum pariter eos pergere ortans, ipse pro proditore **prior** cunctis accedit […].* (Vita Eligii)

→ *Erat enim ceteris **humilior**, **quietus**, ut supra dixi, in sermone, **hilaris** in vultu [...].* (Vita Wandregiseli)

Remark

In the lemmatization process, comparatives and superlatives of adjectives are redressed to their neutral base forms (positive). For the lemmatization of irregular comparative and superlative adjectives see annex 4.

2.5 use

Adjectives are annotated according to their (semantico-)syntactic function. Depending on the context, adjectives may be used substantively (and thus function as nouns) or as adjectives in a specific context.

morphological feature	subcategory	example
use	ADJECTIVAL	[…] *fratres **karissimi** […].* (Visio Baronti) ***beatus** Petrus* (Visio Baronti) *usque ad **vespertina** ora* (Visio Baronti)
	SUBSTANTIVAL	**Masculine adjectives used as nouns:** → […] *fuit igitur largitor indigencium, refocilator **pauperum**, sustentator orfanorum, consolator viduarum, adminiculum obpressorum, misertus **inopum**, monasteriorum visitator, pater **miserorum.*** (Vita Hugberti) → ***Beatus** qui intelligit super **egenum** et **pauperem** [...].* (Vita Austrigisili)

14

→ **Stultus** *in risu exaltat vocem suam* [...]. (Vita Wandregiseli)

→ […] ***mitissimus*** *ac* ***modestus*** *cum summa reverencia prae ceteris adolescentulis videbatur* [...]. (Vita Pardulfi)

→ *Habent* **mali** *in ea quod metuant,* **boni** *quod gratulent,* **superbi** *ut humilient,* **luxoriosi** *exemplum pudicitiae veraecunda mente custodiant,* **garroli** *ut quiescant, cupidi ut arescant,* **avari** *ut relinquentis mundum cum omni inpedimenta eius* [...]. (Vita Wandregiseli)

Neuter singular adjectives used as abstract nouns:

→ *Respondit: 'Sic est* **verum***'.* (Vita Austrigisili)

→ *Vae nobis miseris, qui nullum* **bonum***, quando potuimus, fecimus!* (Visio Baronti)

Substantival use of neuter singular adjectives with prepositions:

→ *Alter pastor e* **contrario** *respondit* […]. (Vita Genovefae)

→ [...] *in eius se consilio commendavit ex* **integro** [...]. (Vita Filiberti)

Neuter plural adjectives used as abstract nouns, signifying *things*:

→ *Ista et alia* **bona** *faciendo refulserunt eius digita et brachia.* (Visio Baronti)

→ *Sufficit* **pauca** *dixisse de pluribus;* [...] *inveniet ita, ut narravimus,* **vera** *dixisse: nusquam conprobabit* **falsa***.* (Vita Audoini)

2.6 Special cases

- Some **adjectival participles** (marked as **PAdi.** in the German-Latin dictionary of Georges/Georges (1913/1918)) partially lost the connection to the verb and may share properties of adjectives. In spite of the adjective properties, adjectival participles are treated as participles and are tagged as pos=V mood=PARTICIPLE, such as *abstinens*, *arrogans*, *diligens*, *florens*, *excellens*, *nocens*, *patiens*, *ardens*, *praeceps*, *libens*, *egens*, *flagrans*, *imminens*, *pollens*, *sapiens*, *situs*, *diversus*, *contemptus*, *doctus*, *perditus*, *rectus*, *tutus*, *aptus*, *certus* etc. (except for *sanctus*, *beatus*)

  → [...] *humilis atque **sapiens** [...].* (Vita Austrigisili)

  → *Erat quidem corpore pulcher, sed pulchrior mente **pollens**que castitatem.* (Vita Boniti)

  → *Nam infra urbem nullus fere **egens** erat* [...]. (Vita Desiderii)

- Some words function both, as adjectives or as nouns, such as *adolescens, iuvenis, infans, senex* etc. These words are tagged according to their use as pos=NN if they are used as nouns and as pos=ADJ if they are used as adjectives.

  → *Alio quoque tempore **adolescens** quidam nomine Ruccolenus* [...]. (Vita Desiderii)

- The Latin **pronominal adjectives** *ullus, nullus, uter, neuter, alter, alius, solus, totus* are categorized as indefinite pronouns (pos=PRO pronountype=INDEFINITE) (see 15.3).

- *Qualis, quantus, talis* and *tantus* are marked as correlative pronouns (pos=PRO pronountype=CORRELATIVE) (see 15.3).

- **Names of months** (e.g., *mensis Ianuarius/Febrarius/Martius/Aprilis/Maius/Iunius* etc.) are adjectives in Latin, they are tagged as ADJ.

3. ADV: adverb

3.1 Morphological features

Adverbs constitute a diverse group of words that may modify verbs, adjectives and other adverbs. All adverbs (also non-gradable adverbs) modify in one morphological category: **comparisondegree**.

  → [...] ***feliciter** remeavit in Galliis.* (Vita Amandi)

  → *Qui **multum** iurat, non effugiet peccatum.* (Vita Austrigisili)

  → *Rursumque post multos dies **crebro** ad virum Dei veniens* [...]. (Vita Pardulfi)

  → *Coepit **paulatim** liquor guttatim succrescere et in pavimento rivolus dilabere.* (Vita Eucherii)

→ [...] *coepit eum **bis terque** appellare.* (Visio Baronti)

→ ***Praeterea** vir beatus habebat amicum fidelissimum quendam inclausum nomine Arnanum.* (Vita Desiderii)

→ *Sic demum, **funditus** morbo expulso, sanitatem pristinam indeptus, ad locum proprium rediit.* (Vita Desiderii)

Adverbs are used to refer to a reason (e.g., *quid* 'why', *cur, quare, quamobrem, quapropter* 'for what reason'), to a degree of truth (e.g., *certe, profecto, vero, fortasse, forsitan, videlicet*), to a direction or goal (e.g., *intus, nusquam, quatenus, obviam, foris, intra, eo*), to an extend of space or time (e.g., *longe; diu, breviter, perpetuo*), to a path (e.g., *illac, istac*), to a position in space (e.g., *contra, intus, ubi, ibi, alibi, aliubi, ibidem, illic, istic, ubicumque, utrimque, passim, nusquam*), to a position in time (e.g., *heri, noctu, modo, ilico, primo, postremo, tum, dudum, aliquando, quando*), to a purpose (e.g., *idcirco, ideo, propterea, cur, quamobrem*), to a quantity or degree (e.g., *paene, valde, abunde, quantum*), to a source (e.g., *hinc, dehinc, exinde, extrinsecus, unde*(*cumque*)) or function as adjuncts of manner (e.g., *sic, ita, clam, palam, ut, quomodo, quemadmodum* 'in what manner' 'how', *qualiter* 'in what kind of manner') (see annex 1).

→ ***Unde** venis, inique, **quo** pergis, quid facere obtas?* (Vita Austrigisili)

→ *Erat enim ceteris humilior, quietus, **ut** supra dixi, [...]* (Vita Wandregiseli)

→ ***Exinde** processit partibus Spaniae, **ubi** Dominus per eum operatus est grande miraculum.* (Vita Audoini)

→ *Contigit **aliquando**, ut iter ageret per pagum Andegavo.* (Vita Audoini)

<u>Remarks</u>

- **Interrogative particles** such as *-ne, num, nonne, si* (*se*), *utrum* and *an* are considered as interrogative adverbs in sentence questions (pos=ADV).

  → ***Nonne** vides eam in manibus suis?* (Vita Austrigisili)

- Some words function both as **adverb** and as **preposition**: *ante, circum, clam, contra, extra, foras/foris, infra, intra, intus, palam, post, propter, supra, super* etc. are annotated as pos=ADV if they are used as adverbs, as pos=AP if they are used as prepositions (see annex 1/2).

  → […] *ut **supra** dictum est* […]. (Vita Desiderii)

  → *De morte Flavadi ab eo **ante** praedicta.* (Vita Eligii)

- ***Ubi***, ***ut***, ***si*** and ***vel*** are annotated according to their function in the text as pos=ADV or pos=CON (see annex 2).
- **Numerical adverbs** (*semel*, *bis*, *ter*, *miliens*, *centiens* etc.) are annotated as pos=ADV.

3.2 comparisondegree

All adverbs are annotated with their degrees of comparison. POSITIVE, COMPARATIVE and SUPERLATIVE forms are possible for gradable adverbs derived from adjectives. All adverbs without comparative or superlative forms are annotated as comparisondegree=POSITIVE.

→ […] *sed exigua, quas didici a narrantibus, silere **minime** audebo.* (Vita Eucherii)

→ *Sed sanctus vir **saepius** Dominum deprecabat [...].* (Vita Audoini)

4. NUM: cardinal number

4.1 Morphological features

The NUM-tag covers all cardinal numbers that represent a countable quantity. The NUM-tag is used for cardinal digits (II/2) and cardinal word numerals (*duo*). Digits (II/2) receive no morphological annotation.[6] Cardinal word numbers are annotated with their **casus**, **genus**, **numerus**, **declensiontype** and **use**.

→ *Tunc **unus** ex illis affatus ait* […]. (Vita Hugberti)

→ *Tunc sanctus Rafahel arcessivit **unum** ex angelis.* (Viso Baronti)

→ ***Tres** mulieres habuit, quod ei non licuerat* […]. (Viso Baronti)

→ *haec est praetiumdivitis et pauperis **XII** solidi.* (Visio Baronti)

4.2 casus, genus, numerus, declensiontype, use

Most cardinal word numerals behave as indeclinable adjectives (INDECLINABLE), the exceptions are *unus, duo* (FIRST_DECLENSION or SECOND_DECLENSION), *tres* (THIRD_DECLENSION) and multiples of *centum* (FIRST_DECLENSION or SECOND_DECLENSION), all of which decline. The numerals from two thousand up are not adjectives but are forms by means of the plural declinable noun *milia, milium* (THIRD_DECLENSION).

---

[6] In the Excel sheet we distinguish between NUM_WORD for word numerals (*duo*) which require a morphological annotation, and NUM_DIGIT for cardinal digits (II/2).

morphological feature	subcategory	example
casus	NOMINATIVE	*unus*
	GENITIVE	*duarum*
	DATIVE	*tribus*
	ACCUSATIVE	*unam*
	ABLATIVE	*uno*
	OBLIQUE	*una(m)*
	INDECLINABLE	*undecim*
genus	MASCULINE	*duobus*
	FEMININE	*trecentae*
	NEUTER	*unum, nongenta*
numerus	SINGULAR	*unus*
	PLURAL	*ducenti*
declensiontype	FIRST_DECLENSION	*una, duae, ducentae*
	SECOND_DECLENSION	*unus, duorum, ducenti*
	THIRD_DECLENSION	*tribus*
	INDECLINABLE	*quinquaginta*
use	ADJECTIVAL	***Tres*** *mulieres habuit, quod ei non licuerat* […]. (Viso Baronti)
	SUBSTANTIVAL	*Tunc* ***unus*** *ex illis affatus ait* […]. (Vita Hugberti)

5. CON: conjunction

The CON-tag belongs to the uninflected or invariable parts of speech. Coordination conjunctions join two expressions or two clauses at the same level. Subordinating conjunctions connect a clause at a lower level to a clause one level deeper.

Coordinating conjunctions are tagged as pos=CON such as *aut*, *vel*, *sive*, *sed*, *autem*, *tamen*, *nam*, *enim*, *et…et*, *neque…neque*, *non solum…sed etiam*, *-que*, *-ve* etc. (see annex 1).

→ […] *aperte cognoscite, non est vester socius,* ***sed*** *noster.* (Visio Baronti)

→ *Fratres, nolite diligere mundum* ***neque*** *ea quae in mundo sunt.* (Visio Baronti)

Subordinating conjunctions are tagged as pos=CON such as *antequam*, *cum*, *donec*, *dum*, *dummodo*, *etsi*, *postquam*, *priusquam*, *antequam*, *quamvis*, *quoad*, *quia/quod* 'because; that', *ut*, *ubi* etc. (see annex 1,2)

→ *Obsecro te, ut largiaris benedictionem tuam ex ore tuo super me, **priusquam** discedas a me.* (Vita Eucherii)

Remark

Ubi, ut and ***vel*** are annotated according to their function in the text as pos=ADV or pos=CON (see annex 2).

6. DIST: distributive number

6.1 Morphological features
A distributive number denotes a number that belongs to each of several groups. Distributive numbers are annotated with the grammatical features **casus**, **genus**, **numerus**, **declensiontype** and **use**.

→ *Dumque ita stantes **singuli** obstupescerent eventum rei* […]. (Vita Eligii)

→ *Quater **sena** centra decora, inferius superius conexa, surgent celsaque fastigia micant.* (Vita Boniti)

6.2 casus, genus, numerus, declensiontype, use
Distributive numbers are declined like the standard first and second declension plural adjectives.

morphological feature	subcategory	example
casus	NOMINATIVE	*quini*
	GENITIVE	*singulorum*
	DATIVE	*septenis*
	ACCUSATIVE	*novenos*
	VOCATIVE	*singuli*
	ABLATIVE	*singulis*
genus	MASCULINE	*denos*
	FEMININE	*singulae*
	NEUTER	*novena*

numerus	PLURAL	*singuli*
declensiontype	FIRST_DECLENSION	*novenas*
	SECOND_DECLENSION	*septingenos*
use	ADJECTIVAL	**bini** *boves*
	SUBSTANTIVAL	**singuli** *divino sacerdotio honorati et in clerico ministerio constituti*

7. FM: foreign material

All foreign material is sorted into this category. The FM-tag is used for all words that don't concern the linguistic analysis of the corpus (e.g., ancient Greek words).

→ *anthropos*

→ *cagaton*

→ *homostitios*

→ *amen*

8. ITJ: interjection

An interjection is a word that is often used as an exclamation and express an emotional reaction, such as *o*, *en*, *vah*, *heu*, *vae* or *eia*.

→ **O** *virtus abstinentiae et mortificatio carnis, quae ad caelos mittis milites Christi!* (Vita Audoini)

→ **En**, *unus illorum est in pelago, cui numquam nocere potero* […]. (Vita Galli Wettino)

→ **'Eia'**, *inquit Eligius, 'eia, probabis hoc in die illa, cum te illuc coram populis et principibus visurus sum in palam'.* (Vita Eligii)

Remark

Ecce is labelled as pos=ADV (see annex 1).

9. XY: non-word

In the *CompHistSem* tagset, a non-word is a word that is not recognized or accepted as legitimate, as one produced by a spelling or typographical error (OCR error). It is a word whose meaning or use is not approved in a given circumstance.

→ *H20*

→ *n*men*

In the *lapos* tagset, the tag XY is used for incomprehensible tokens that can't be assigned a real part of speech category, and that may not be annotated as foreign material (pos=FM see 7).

→ *ava* (aus *ab* + *via*) (Epistolae Austrasicae)

→ *aso* (Formulae Andecavenses)

Sometimes, especially in the genre of *charters*, additional, abbreviated information is placed between brackets. For avoiding any quantitative distortion, we set '#' on word- and lemma-level and we consider it as a non-word (XY) in terms of PoS-category.

(combination)	token	word	lemma	CL	CW	POS
(N. T.)	N	#	#	.	.	XY
(N. T.)	T	#	#	.	.	XY
(C.)	C	#	#	.	.	XY
(L. S.)	L	#	#	.	.	XY
(L. S.)	S	#	#	.	.	XY
(S. I.)	S	#	#	.	.	XY
(S. I.)	I	#	#	.	.	XY
(S. R.)	S	#	#	.	.	XY
(S. R.)	R	#	#	.	.	XY
(M.)	M	#	#	.	.	XY

10. NN: normal noun

10.1 Morphological features

The category of nouns is further split into normal nouns, personal names (see 13) which refer to an individual person and named entities (see 16) which refer to a river, region, nation or hill. Normal nouns – common nouns which don't denote an individual entity – modify in four different morphological categories: **casus**, **genus**, **numerus** and **declensiontype**.

10.2 casus, genus, numerus

As adjectives, pronouns and participles, nouns' endings alter to show grammatical case: ABLATIVE, ACCUSATIVE, DATIVE, GENITIVE, NOMINATIVE or VOCATIVE. Furthermore, indeclinable nouns are tagged as INDECLINABLE in the categories **casus** and **declensiontype**, locatives are categorized as LOCATIVE and non-Classical Latin accusative forms (apocope of the final -*m*) are classified into the category OBLIQUE.

The **genus** category may be MASCULINE, FEMININE or NEUTER, the category **numerus** may be modified in SINGULAR or PLURAL.

morphological feature	subcategory	example
casus	NOMINATIVE	*Ubi est illa **puella**, quam ospes noster accepit?* (Vita Austrigisili)
	GENITIVE	[...] *ita ut magnarum **fluminum** alvei **siccitatis** squalore torrerent* [...]. (Vita Boniti)
	DATIVE	*Mox, absolutis membris, pristinae **sanitati** ac **usui** cotidiano restitutus probatur.* (Vita Austrigisili)
	ACCUSATIVE	*Lugdunensem pervenisset in **urbem*** [...]. (Vita Boniti)
	VOCATIVE	*Venite, **fili**, audite me;* (Visio Baronti)
		*Gloria tibi, **Deus**!* (Visio Baronti)
		*Redime te, **frater**!* (Visio Baronti)
	ABLATIVE	*Eodem **tempore** sub Theoderico **principe** Pippinus regni primatum tenens atque curam palacii gerens* [...]. (Vita Boniti)
	LOCATIVE	*nullus monachus **domi** resedit, qui non obviam pastori procederet.* (Vita Desiderii)
		*Et usque ad tertia **ora** cedentes me audaciter* [...]. (Visio Baronti)
	OBLIQUE	[...] *ad superna **patria** festinabat.* (Vita Audoini*)*
		[...] *Attela Chunorum regem sevitia superatum Gallia **provintia** coepisse vastare* [...]. (Vita Genovefae)
		[...] *qua ultione percussi sunt illi, qui hoc **nefas** perpetrarunt.* (Vita Trudonis)
	INDECLINABLE	*Si vis vivere vel usque **mane**, presentialiter redde quod inde habes*. (Vita Austrigisili)
		opus (*est*)
genus	MASCULINE	*nauta*
	FEMININE	*puella*
	NEUTER	*periculum*
numerus	SINGULAR	*iter*
	PLURAL	*gratias*

- INDECLINABLE

  Indeclinable are all indeclinable neuter nouns such as *cepe*, *git*, *gummi*, *stibi*, *pondo, nefas*, *fas*, *vale*, *sinapi*, *mane* or *gausape.*

- LOCATIVE

  Nouns that use the locative instead of a preposition. The locative is confined to a few Latin nouns: e.g., *domi*, *ruri* and *humi.*

- OBLIQUE

  Non-Classical Latin accusative forms (apocope of the final -*m*) are classified into the category OBLIQUE: e.g., (*infra*) *monition**e***

10.3 declensiontype

Normal nouns are tagged with their specific decensiontype on the basis of their inflectional endings. The declension types are subdivided into first declension (-*a* stem declension), second declension (-*o* stem declension), third declension (consonant stems and vowel stems), fourth declension (-*u* stems), fifth declension (-*e* stem declension) and indeclinable nouns.

morphological feature	subcategory	example
declensiontype	FIRST_DECLENSION	*Illa summo rerum creatori domino Iesu Christo **gratias** referens* [...]. (Vita Sadalbergae)
		***Nautae** vero, misso in pelago rete, eundem piscem coeperunt.* (Vita Amandi)
	SECOND_DECLENSION	*quod **periculum** nos deinceps vidimus.* (Vita Sadalbergae)
		*de tectis **domorum*** (Vita Austrigisili)
	THIRD_DECLENSION	*quem postea et comitem **itineris** et socium **laboris** habuit.* (Vita Sadalbergae)
	FOURTH_DECLENSION	*Mox, absolutis membris, pristinae sanitati ac **usui** cotidiano restitutus probatur.* (Vita Austrigisili)
	FIFTH_DECLENSION	*Mira **res** et vehementer supenda*! (Vita Desiderii)

INDECLINABLE	*Fraudolenus presbyter, qui* **dies** *suos bene custodivit;* (Visio Baronti)
	All indeclinable neuter nouns such as *cepe, git, gummi, stibi, pondo, nefas, fas, vale, sinapi, mane* or *gausape* (see above 10.1).

10.4 Special cases

- The singular and plural forms of *Deus/deus, Filius/filius, Pater/pater, Dominus/dominus* and *Spiritus/spiritus* are tagged as NN even if these nouns are written in majuscule like personal names. On word level, these forms are written in lowercase.

  → *Et fratres gratias* **Deo** *referent pro eius oboedientia.* (Visio Baronti)

  → […] *sed si* **Pater** *caelestis voluerit, adhuc illuc remeavit.* (Visio Baronti)

  → *Clamat et admonet* **Dominus** *per prophetam.* (Visio Baronti)

- Some words function both as adjective and as noun, such as *adolescens*, *iuvenis*, *infans, senex* etc. These words are marked as pos=NN if they are used as nouns, they are categorized as pos=ADJ if they are used as adjectives (see 2.6).

  → *Alio quoque tempore* **adolescens** *quidam nomine Ruccolenus* [...]. (Vita Desiderii)

11. ORD: ordinal number

Ordinal numbers determine in what order a series of things is. The ORD-tag is used for ordinal digits (II./2.) and ordinal word numerals (*secundus*). The section numbering of the texts is marked as ordinal number. Ordinal digits (II./2.) receive no morphological annotation.[7]

  → *Mox, expleta oratione, venimus ad* **secundam** *portam paradysi* […]. (Viso Baronti)

  → *Mox, aperto euangelio, hoc* **primum** *verbum invenerunt* […]. (Vita Hugberti)

  → *Acta sunt haec omnia* **VIII.** *Kal. April. in* **sexto** *anno regnante Theoderico regem Francorum.* (Visio Baronti)

11.1 Morphological features

Ordinal word numbers modify in five different morphological subcategories: **casus**, **genus**, **numerus**, **declensiontype** and **use**. Ordinal numbers are declined like the standard first and second declension adjectives.

---

[7] In the Excel sheet we distinguish between ORD_WORD for ordinal word numerals (*secundus*) which require a morphological annotation, and ORD_DIGIT for ordinal digits (II/2).

morphological feature	subcategory	example
casus	NOMINATIVE	*quartus*
	GENITIVE	*primorum*
	DATIVE	*millesimo*
	ACCUSATIVE	*undecimos*
	VOCATIVE	*quarti*
	ABLATIVE	*sextis*
	OBLIQUE	*tertia*(*m*)
genus	MASCULINE	*vicesimus*
	FEMININE	*tertia*
	NEUTER	*quintum*
numerus	SINGULAR	*quingentesimus*
	PLURAL	*ducentesimi*
declensiontype	FIRST_DECLENSION	*secundas*
	SECOND_DECLENSION	*septimorum*
use	ADJECTIVAL	*venimus ad* **secundam** *portam paradysi* […]. (Viso Baronti)
	SUBSTANTIVAL	**primi** *hostium* *Childevertus rex Leudovaldum episcopum cum* **primis** *regni sui dirixit.* (Gregor VI)

11.2 Special cases

Non-Classical Latin accusative forms (apocope of the final -*m*) are classified into the category casus=OBLIQUE.

→ *Et usque ad* **tertia ora** *cedentes me audaciter* […]. (Visio Baronti)

12. NP: personal name

The NP-tag is used for all anthroponyms, names or part of the names of human beings. Personal names are annotated with their **casus**, **genus**, **numerus** and **declensiontype**.

→ *Pater eius **Severus**, mater **Geroncia** vocitata est.* (Vita Genovefae)

→ *Ita demum dominus noster **Iesus** Christus per servum suum, innovata gratia, replevit aquas Spania.* (Vita Audoini)

→ *Cum esset gentiles **Childericus** rex Francorum [...].* (Vita Genovefae)

morphological feature	subcategory	example
casus	NOMINATIVE	*Gunthchramnus*
	GENITIVE	*Sygiberthi*
	DATIVE	*Chlothario*
	ACCUSATIVE	*Geronicam*
	VOCATIVE	*Gai, Marce*
	ABLATIVE	*Egidio*
	OBLIQUE	*Genovefa(m)*
	INDECLINABLE	*Adam, Elisabeth, Judith*
genus	MASCULINE	*Severus*
	FEMININE	*Mariae*
numerus	SINGULAR	*Leudovaldus*
declensiontype	FIRST_DECLENSION	*Cornelia*
	SECOND_DECLENSION	*Childebertus*
	THIRD_DECLENSION	*Cicero, Moyses*
	INDECLINABLE	*Adam, Elisabeth, Judith*
	GREEK_DECLENSION	*Aeneas, Thomas, Iesus*

Remarks

- *Deus/deus, Pater/pater, Dominus/dominus* and *Filius/filius* are annotated as normal nouns (pos=NN) even if these nouns are written like personal names in majuscule.

  → *Et fratres gratias **Deo** referent pro eius oboedientia […].* (Visio Baronti)

  → *sed si **Pater** caelestis voluerit, adhuc illuc remeavit.* (Visio Baronti)

  → *Clamat et admonet **Dominus** per prophetam […].* (Visio Baronti)

- Indeclinable personal names are annotated as casus=INDECLINABLE and declensiontype=INDECLINABLE (such as *Adam*, *Elisabeth*, *Judith etc.*).

→ […] *a quo* **Adam** *olim victus est in paradyso* [...]. (Vita Richarii)

- Personal names may consist of other parts of speech (such as *Alexander Magnus*, *Iesus Christus*). Words that belong to other categories (adjectives, verbs etc.) are not annotated as personal names. In the aforementioned examples *Magnus* or *Christus* are pos=ADJ.

    → *Hoc ita* **Alexander Magnus** *insigni volumine ad matrem suam scribit* […]. (004_6069.xml)

- For the lemmatization of the various spellings of personal names see III.

13. AP: adposition

Adposition is a cover term for prepositions and postpositions. A prepositional phrase in Latin is made up of a preposition followed by a noun or a noun phrase in an oblique case. There are few postpositions (such as *causa*, *gratia*, *tenus* etc.) that follow a noun or a noun phrase. Prepositions and postpositions are marked as pos=AP.

→ […] *quadragisimo die* **post** *gloriosam Domini resurrectionem* […]. (Vita Audoini)

→ […] *die dominico ante lẹtanias* **inter** *catervas populi,* **inter** *agmina clericorum,* **inter** *choros psallentium consecrati sumus gratis* **ab** *episcopis pariter episcopi* […]. (Vita Eligii)

→ […] *rogat sibi deferri* **ex** *eo oleo quod medendi* **gratia ante** *pignora praefati antestitis dependebat.* (Vita Eligii)

→ *Deinde venimus* **ad** *tertiam portam paradysi* […]. (Visio Baronti)

→ […] *ad predictam parochiam manendi vel maxime orandi* **gracia** *se contulerunt.* (Vita Genovefae)

Remark

Some words (such as *ante*, *circum*, *clam*, *contra*, *extra*, *foras/foris*, *infra*, *intra*, *intus*, *palam*, *post*, *propter*, *supra*, *super* etc.) may be annotated as either ADV or AP depending on whether they are used as adverbs or prepositions.

→ […] *ut* **supra** *dictum est* […]. (Vita Desiderii)

→ *De morte Flavadi ab eo* **ante** *praedicta.* (Vita Eligii)

14. PRO: pronoun

14.1 Morphological features

Pronouns substitute nouns or noun phrases. A pronoun contains its **casus**, **genus**, **numerus**, **declensiontype**, **pronountype** and **use**.

14.2 casus, genus, numerus

Pronouns modify in **casus** (NOMINATIVE, GENITIVE, DATIVE, ACCUSATIVE, ABLATIVE, OBLIQUE and INDECLINABLE; VOCATIVE if possessive pronoun), **genus** (MASCULINE, FEMININE, NEUTER) and **numerus** (SINGULAR, PLURAL).

14.3 pronountype, use

Pronouns are annotated with a pronoun type that further differentiates pronouns into CORRELATIVE, DEMONSTRATIVE, INDEFINITE, INTENSIVE, INTERROGATIVE, PERSONAL, POSSESSIVE, REFLEXIVE and RELATIVE pronouns. Apart from personal pronouns that are only used independently as pronouns, pronouns can occur both as determiners (attributes) (use=ADJECTIVAL) and as pronouns (use=SUBSTANTIVAL).

- CORRELATIVE

  Correlative pronouns are closely related to each other (e.g., *talis*, *qualis*, *tantus*, *quantus*, *tantum*, *quantum*, *tot*, *quot*, *aliquantus*, *aliquot*, *quantuscumque*, *quotquot*). Each series contains an interrogative, an indefinite, a demonstrative, a relative, and an indefinite relative pronoun/adjective.

  → *Reversus igitur mensumque lignum repperit **tot** largius, **quot** dudum fuerat brevius* […]. (Vita Amati)

  → *Astra non sunt munda in conspectu Dei, **quanto** magis nos quorum vita temptatio est* […]. (Vita Desiderii)

  → ***Quantus** sit in calicibus decor, ex distinctione gemmarum nec ipsos intuencium obtutos facile diiudicare reor.* (Vita Desiderii)

morphological feature	subcategory	example
casus	NOMINATIVE	*quantacumque*
	GENITIVE	*qualis*
	DATIVE	*aliquanto*
	ACCUSATIVE	*tantumdem*
	ABLATIVE	*qualibus*
	OBLIQUE	*tale*(*m*)
	INDECLINABLE	*tot, quot, totidem, aliquot, quot*
genus	MASCULINE	*quantus*
	FEMININE	*tanta*
	NEUTER	*quantum*
numerus	SINGULAR	*quanto*
	PLURAL	*talibus*
declensiontype	FIRST_DECLENSION	*tanta, quanta, aliquantus*
	SECOND_DECLENSION	*tantus, quantus*
	THIRD_DECLENSION	*talis, qualis*
	INDECLINABLE	*tot, quot, totidem, aliquot, quot*
use	ADJECTIVAL	***Quantus*** *sit in calicibus* ***decor***, *ex distinctione gemmarum nec ipsos intuencium obtutos facile diiudicare reor;* (Vita Desiderii)
	SUBSTANTIVAL	*Quoniam,* ***qualis*** *in imperiis fuerit, exposui, referam nunc interiorem ac familiarem vitam.*

Remark

There is no **genus** and **numerus** for indeclinable correlative pronouns (such as *tot*, *quot*, *totidem*, *aliquot*, *quot*).

- DEMONSTRATIVE

There are five demonstrative pronouns (used attributively or independently) indicating a place with respect to the speaker: *hic/haec/hoc, is/ea/id, iste/ista/istud, ille/illa/illud, idem/eadem/idem.* The demonstrative pronouns are declined like adjectives of the first and second declensions.

→ […] *iam non potestis nocere ad **istam** animolam, dum signum sonavit super ipsam ecclesiam, quia fratres congregantur, ut orarent pro **illam** […].* (Visio Baronti)

morphological features	subcategory	example
casus	NOMINATIVE	*hic*
	GENITIVE	*illius*
	DATIVE	*isti*
	ACCUSATIVE	*eandem*
	ABLATIVE	*istis*
	OBLIQUE	*illa(m)*
genus	MASCULINE	*ille, hic, is, iste, idem*
	FEMININE	*illa, haec, ea, ista, eadem*
	NEUTER	*illud, hoc, id, istud, idem*
numerus	SINGULAR	*huic*
	PLURAL	*has*
use	ADJECTIVAL	[…] *iam non potestis nocere ad **istam** animolam* […]. (Visio Baronti)
	SUBSTANTIVAL	*dixit **hoc**:* […]

Remarks

o Latin uses the possessive genitive (*eius*, *eorum*, *earum*) of the demonstrative pronoun *is*, *ea*, *id* as a **non-reflexive possessive**. In this case *eius*, *eorum* and *earum* are marked as possessive pronouns (pos=PRO pronountype=POSSESSIVE).

→ *Filius decessit. Quotiens cubiculum **eius** intraret, vivere filium simulabat.*

- For lack of a **non-reflexive personal pronoun** of the third person, Latin uses the demonstrative pronoun *is*, *ea*, *id* to refer to the just mentioned person. These word forms are annotated as personal pronouns (pos=PRO pronountype=PERSONAL use=SUBSTANTIVAL).

  → *Pater filium vocat. Iulus autem **eum** non audit.*

- INDEFINITE

  Indefinite pronouns such as *aliquis/aliquid* (independent), *aliqui/aliqua/aliquod* (attibutive), *alecus/alicus*, *ullus/ulla/ullum*, *quidam/quaedam/quiddam*, *quidam/quaedam/quoddam*, *quilibet/quaelibet/quidlibet*, *quivis/quaevis/quidvis*, *quisquam/quicquam*, *uterque/utraque/utrumque*, *quispiam*, *neuter/neutra/neutrum*, *nemo*, *nihil* (*nil*), *quidquis/quaeque/quidquid* and *quodquod*, *quicumque/quaecumque/quodcumque*.

  The Latin **pronominal adjectives** *ullus*, *nullus*, *nonnullus*, *uter*, *neuter*, *alter*, *alius*, *solus* and *totus* are annotated as indefinite pronouns (pos=PRO, pronountype=INDEFINITE).

  <u>Remark</u>
  - *Unus* is annotated as a cardinal number (pos=NUM).
  - *Ambo* 'both' is annotated as pos = PRO pronountype=INDEFINITE

  → *Nam sicut non iustificabitur **aliquis** aliena laude, ita nec ledetur infamia.* (Vita Genovefae)
  → [...] ***cuique** ut vocabulum nominis, ita et virtus animi et caelestium gratiarum magnificentia concordabat.* (Vita Eligii)
  → *Hoc miraculo **nemo** ambigit hodie usque degere.* (Vita Eucherii)
  → *Respondit ei **quaedam** soror* [...]. (Vita Bertilae)
  → [...] *quod **unusquisque** secundum suam personam desolveret* [...]. (Vita Austrigisili)

morphological feature	subcategory	example
casus	NOMINATIVE	*quidam, altera*
	GENITIVE	*cuiusque, totius*
	DATIVE	*nemini, ulli*
	ACCUSATIVE	*nihil, aliam*
	ABLATIVE	***ullo** modo*
genus	MASCULINE	unusquisque, unus
	FEMININE	quaevis, *neutra*
	NEUTER	*quod, aliud*
numerus	SINGULAR	***quavis** ratione*
	PLURAL	*ab **utrisque** militibus auditus est*
use	ADJECTIVAL	*Respondit ei **quaedam soror** […]. (Vita Bertilae)*
		*Fratres vero, requirentes eas, in **alteram domum** invenerunt prefatum presbiterum desuper lacmentantem […]. (Vita Pardulfi)*
	SUBSTANTIVAL	*[...] quod **unusquisque** secundum suam personam desolveret [...]. (Vita Austrigilili)*
		Uterque** nostrum; quorum **uterque
		*[...] **nullus** ex nobis ista credebat. (Vita Hugberti)*

Remark

After short conjunctions (such as *ne*, *ut*, *nisi*, *si*, *num*) *aliqui*(*s*) is shortened to *qui*(*s*). The *classicallemma* is either *aliquis/aliquid* or *aliqui/aliqua/aliquod* depending on whether the pronoun is used independently or attributively.

- INTENSIVE

  The intensive pronoun *ipse*, *ipsa*, *ipsum* is used to intensify the meaning of a noun, but it may also appear alone. *Semetipse*, *semetipsum*, *semetipsi* and *semetipsos* are also annotated as intensive pronouns.

  → *Tunc **ipse** cursu veloci habiit et vocavit sanctum Petrum.* (Visio Baronti)

  → *Et ego, ut audivi, respexi et vidi **ipsum** monasterium [...].* (Visio Baronti)

morphological feature	subcategory	example
casus	NOMINATIVE	*ipse, semetipse, semetipsi*
	GENITIVE	*ipsorum*
	DATIVE	*ipsi*
	ACCUSATIVE	*ipsas, semetipsum, semetipsos*
	ABLATIVE	*ipso*
	OBLIQUE	*ipsa(m)*
genus	MASCULINE	*ipse*
	FEMININE	*ipsa*
	NEUTER	*ipsum*
numerus	SINGULAR	*ipsium*
	PLURAL	*ipsis*
use	ADJECTIVAL	*Et ego, ut audivi, respexi et vidi **ipsum monasterium** [...].* (Visio Baronti)
	SUBSTANTIVAL	*Tunc **ipse** cursu veloci habiit et vocavit sanctum Petrum.* (Visio Baronti)

- INTERROGATIVE

  The interrogative pronouns are used strictly for asking questions (e.g., *quis*, *quid*, *uter*, *qui*, *quae*, *quod*, *quisnam*, *quidnam*, *quaenam*, *quondam*, *cuius*). These words can be used in the same way in indirect questions. They are distinct from relative pronouns.

  → ***Quid** possum, bone pastor, donare, quia hic prae manibus nihil habeo.* (Visio Baronti)

  → ***Quis** diceret, eum umquam vel de cibo aut potum esse satiatum?* (Vita Audoini)

  → ***Quod** huic monacho habetis **crimen** obponere?* (Visio Baronti)

  → ***Quisnam** ille est […]?* (Visio Baronti)

morphological feature	subcategory	example
casus	NOMINATIVE	*quis?*
	GENITIVE	*utrius?*
	DATIVE	*utri?*
	ACCUSATIVE	*quem?*
	ABLATIVE	*quo?*
	OBLIQUE	*que(m)?*

genus	MASCULINE	*qui*
	FEMININE	*utra*
	NEUTER	*utrum*
numerus	SINGULAR	*quem*
	PLURAL	*quorum*
use	ADJECTIVAL	***Quis civis*** *creatur?*
	SUBSTANTIVAL	***Quid*** *possum, bone pastor, donare, quia hic prae manibus nihil habeo.* (Visio Baronti)

- PERSONAL

  The personal pronoun is used only independently as a pronoun.

  → *Quando* **me** *vidistis externa nocte matutinis laudibus Christi* **vobiscum** *incolomem consumasse* [...]. (Visio Baronti)

  → *Nisi claritas Dei eum a* **nobis** *auferat,* **tu** *non potes auferre.* (Visio Baronti)

  → *Hi omnes, ut* **nos** *viderunt et daemones latere meo sinistro vehementer adstringere, stupefacti sermocinare* **nobiscum** *voluerunt* [...]. (Visio Baronti)

  → *Et* **ego** *aio ad eos* [...]. *Et* **nos** *statim ut ingressi sumus* [...]. (Visio Baronti)

morphological feature	subcategory	example
casus	NOMINATIVE	*tu, ego, ea*
	GENITIVE	*mei, vestrum, eius*
	DATIVE	*tibi, nobis, ei*
	ACCUSATIVE	*me, nos, eum*
	VOCATIVE	*tu, vos*
	ABLATIVE	*tecum, nobiscum, eo*
	OBLIQUE	*sua(m), ea(m)*
genus	MASCULINE	*et ait ad eos beatissimus Rafahel:* **ego** *istam animam mecum hinc ante tribunal aeterni iudicis duco* [...] (Visio Baronti)
	FEMININE	*Quam cum in basilica sancti Anniani antestitis orantem repperisset, cadens ad pedes eius, ululans, hoc tantum*

		*deprecasse fertur: 'Redde **mihi**, domina Genovefa, filiam meam'.* (Vita Genovefae)
numerus	SINGULAR PLURAL	*mihi* *nobis*, *nos*
use	SUBSTANTIVAL	*Nisi claritas Dei eum a **nobis** auferat, **tu** non potes auferre.* (Visio Baronti) *Pater filium vocat. Iulus autem **eum** non audit.*

- POSSESSIVE

  Possessive pronouns are declined like the adjectives of the first and second declension (*a*- and *o*-declensions) with the exception that the vocative masculine singular of *meus* is *mi.* The possessive adjectives are used attributively (*mare nostrum* 'our sea'), the possessive pronouns are used independently (*nostri* 'our men').

  Latin lacks a non-reflexive possessive pronoun of the third person. Therefore, it uses the possessive genitive (*eius*, *eorum*, *earum*) of the demonstrative pronoun *is*, *ea*, *id*, that is classified in this case as a possessive pronoun (pos=PRO pronountype=POSSESSIVE use=ADJECTIVAL).

  → *Benedictus tu, **mi** pater, quia que desidero sciscitas, si ambio adipisci.* (Vita Genovefae)

  → *Amice, refrena cupiditatem **tuam**!* (Vita Eligii)

  → *Sed ille frater, manus **suas** ad latum **suum** extensas, oculos clusos, coepit semivivus iacere* [...]. (Visio Baronti)

  → *Quapropter tam cautos bellatores **nostrae** non superabunt versutiae* [...]. (Vita Galli Walahfrido)

  → [...] *per idem vero tempus germanus **eius** Avitus pontifex* [...]. (Vita Boniti)

morphological feature	subcategory	example
casus	NOMINATIVE	*suus*
	GENITIVE	*tuorum, eius, earum, eorum*
	DATIVE	*meis*
	ACCUSATIVE	*suos*
	VOCATIVE	*mi*
	ABLATIVE	*vestris*
	OBLIQUE	*mea*(*m*)
genus	MASCULINE	*meus, eorum*
	FEMININE	*tuam, earum*
	NEUTER	*nostrum*
numerus	SINGULAR	*suus, meus, eius*
	PLURAL	*suus, noster, eorum*
declensiontype	FIRST_DECLENSION	*tua*
	SECOND_DECLENSION	*suorum*
use	ADJECTIVAL	*Amice, refrena cupiditatem **tuam**!* (Vita Eligii)
		*Domus **eius**/**eorum***
	SUBSTANTIVAL	***meum** mihi reddatur*
		nostri

Remark

No **declensiontype is marked** for the non-reflexive possessive pronoun of the third person.

- REFLEXIVE

   The personal pronoun of the third person is reflexive. The reflexive pronoun refers to the subject of the clause in which it stands or to the subject of the governing clause: *sui, sibi, se, sese, a se, secum*.

   → […] *ipse **sibi** fervidus milis institerat.* (Vita Audoini)

   → *Tunc ipse puer coepit cum eiulato magno currere et vocato fratre **secum** adducere.* (Visio Baronti)

morphological feature	subcategory	example
casus	GENITIVE	*sui*
	DATIVE	*sibi*
	ACCUSATIVE	*se*
	ABLATIVE	(*a*) *se*, *secum*
genus	MASCULINE	*Hoc audito, rex arcessivit ad **se** Austrigisilum* […] (Vita Austrigisili)
	FEMININE	*Ad illa, que in cassum **se** Christi sponsa profitebatur, conscientia convicta protinus ad pedes Genovefe corruit.* (Vita Genovefae)
numerus	SINGULAR	*vocans eam ad **se**, ait ei* […] (Vita Genovefae)
	PLURAL	*in Brittaniam* […] ***se** contulerunt.* (Vita Genovefae)
use	SUBSTANTIVAL	*Tunc ipse puer coepit cum eiulato magno currere et vocato fratre **secum** adducere.* (Visio Baronti)

- RELATIVE
    - → *Leodoaldus lector, **quem** Deus proprio benedixit.* (Visio Baronti)
    - → *Omnis enim **quicumque** invocaverit nomen Domini salvus erit* [...]. (Vita Adelphii)
    - → […] *in ecclesiam **quisquis** ignarus ingressurus adesset* [...]. (Visio Boniti)

morphological feature	subcategory	example
casus	NOMINATIVE	*quisquis*
	GENITIVE	*quorum*
	DATIVE	*cuicumque*
	ACCUSATIVE	*quemquem*
	ABLATIVE	*quoquo*
	OBLIQUE	*que*(*m*)
genus	MASCULINE	*quicumque*
	FEMININE	*quaecumque*

	NEUTER	*quod*
numerus	SINGULAR	*quidquid*
	PLURAL	*quibus*
use	ADJECTIVAL	***quoquo*** *modo*
	SUBSTANTIVAL	[…] *in ecclesiam* ***quisquis*** *ignarus ingressurus adesset* [...]. (Vita Boniti)

14.4 Special cases

- Non-Classical Latin accusative forms (apocope of the final -*m*) are marked as casus=OBLIQUE.

    → [...] *procedens Genovefa ad* ***cellola sua*** [...]. (Vita Genovefae)

15. NE: named entity

Named entities or proper names are names or parts of the names of nations, peoples, cities, rivers or hills. Proper names are annotated with their **casus**, **genus**, **numerus** and **declensiontype**.

→ [...] *inde vero procedente tempore orta est discordia inter gentem* ***Francorum*** *et* ***Austrasiorum*** [...]. (Vita Genovefae)

→ ***Pannonia, Noricus et Rhetia*** *habent ab oriente* ***Moesiam***, *a meridie* ***Histriam***, *ab africo* ***Alpes*** *Penninas, ab occasu* ***Galliam*** *Belgicam, a circio* ***Danubii*** *fontem, a septentrione* ***Danubium et Germaniam***. (Vita Galli Walahfrido)

morphological feature	subcategory	example
casus	NOMINATIVE	*Noricus*
	GENITIVE	*Danubii*
	DATIVE	*Moesiae*
	ACCUSATIVE	*Germaniam*
	VOCATIVE	*Roma*
	ABLATIVE	*Galliis*
	LOCATIVE	*Erat autem eo tempore* ***Romae*** *praesul beatissimus papa Martinus* […]. (Vita Eligii)
	OBLIQUE	*Roma*(*m*)

	INDECLINABLE	*Parisius*, *Jerusalem*
genus	MASCULINE	*Danubius*
	FEMININE	*Moesia*
	NEUTER	*Lugdunum*
numerus	SINGULAR	*Francus*
	PLURAL	*Samnites*
declensiontype	FIRST_DECLENSION	*Pannonia*
	SECOND_DECLENSION	*Romanus*
	THIRD_DECLENSION	*Alpes*
	FOURTH_DECLENSION	*Idus*
	INDECLINABLE	*Parisius*, *Jerusalem*

Remarks

- The **locative or the ablative of location** is confined to cities and small islands: *Romae*, *Corinthi*, *Cypri*, *Carthagine*, *Athenis* and *Delphis.*
- Words that belong to other categories (adjectives, nouns etc.) are not annotated as proper names. In the aforementioned examples *Penninas* and *Belgicam* are tagged pos=ADJ.
- **Indeclinable proper names** are tagged casus=INDECLINABLE and declensiontype=INDECLINABLE (such as *Jerusalem*, *Parisius* etc.).
  → [...] in **Parisius** urbem migravit. (Vita Genovefae)
- All **names of ethnic groups** are converted to their base form, which is the nominative singular (*token*: *Gothorum* → *lemma*: *Gothus*)
- **Named entities** may consist of other parts of speech (such as *Lugdunum Clavatum*, *mare Rubrum*, *mare Cespium*). Words that belong to other grammatical categories (adjectives, verbs etc.) are not annotated as named entities. In the aforementioned example, *Clavatus* is labelled as pos=V mood=PARTICIPLE.
- **Names of months** (e.g., *mensis Ianuarius/Febrarius/Martius/Aprilis* etc.) are adjectives in Latin, they are labelled as pos=ADJ (see 2).

16. V: verb

16.1 Morphological features

A verb contains its **casus**, **genus**, **numerus**, **person**, **declensiontype**, **comparisondegree**, **mood**, **tense**, **voice**, **verbtype** and the **conjugationtype**.

morphological feature	mood
casus	PARTICIPLE, SUPINE, GERUND, GERUNDIVE
genus	PARTICIPLE, SUPINE, GERUND, GERUNDIVE
numerus	GERUND, GERUNDIVE, IMPERATIVE, INDICATIVE, PARTICIPLE, SUBJUNCTIVE and SUPINE
person	IMPERATIVE, INDICATIVE and SUBJUNCTIVE
declensiontype	PARTICIPLE, GERUNDIVE
comparisondegree	PARTICIPLE
tense	all moods except GERUND, GERUNDIVE and SUPINE
voice	all moods except GERUND, GERUNDIVE and SUPINE
verbtype	all moods
conjugationtype	all moods

16.2 casus, genus, numerus, person

Finite verbs are annotated with their **numerus** and **person.**

The personal forms of finite verbs refer to an expressed or understood subject of first person (FIRST_PERSON), second person (SECOND_PERSON) and third person (THIRD_PERSON). Furthermore, they express a numerus (SINGULAR, PLURAL).

Non-finite verbs are annotated with their **casus**, **genus**, **numerus**.

In the compound tenses (perfective passive tenses and the periphrastic conjugation), the nominal part of the verb agrees with the subject in casus, genus and numerus: the three participles (present active participle, present passive participle, future active participle, the gerund(ive) and the supine) contain a casus, genus and numerus.

16.3 mood

All verbs are annotated with their mood.

- GERUND

    The gerund is a verbal noun that is annotated with its **casus** (GENITIVE, DATIVE, ACCUSATIVE, ABLATIVE), **genus** (NEUTER), **verbtype** and **conjugationtype** and **declensiontype.**

    → *Predicando iugiter et multos et innumerabiles in peccatis delapsos ad viam salutis revocabat.* (Vita Hugberti)

→ [...] *quodque magis impedirent* **contradicendo**, *quam consentirent* **annuendo.** (Vita Bertilae)

→ [...] *ad predictam parochiam* **manendi** *vel maxime* **orandi** *gracia se contulerunt.* (Vita Genovefae)

morphological features	subcategory	example
casus	GENITIVE	*dicendi*
	DATIVE	*scribendo*
	ACCUSATIVE	*ad* **dimicandum**
	ABLATIVE	*inconsulte* **largiendo**
	OBLIQUE	*ad* **audiendo**
genus	NEUTER	*ad* **docendum**
numerus	SINGULAR	*vivendo*
declensiontype	SECOND_DECLENSION	*conradicendo*

- GERUNDIVE

  The gerundive is a passive verbal adjective. It is annotated with its **casus**, **numerus**, **genus**, **declensiontype**, **verbtype** and **conjugationtype**.

  → *Illud etiam non est* **praetermittendum** [...]. (Vita Amandi)

  → [...] *ut vehicula ad vinum in usibus fratrum monasterio* **deferendum** *praeparare deberet* [...]. (Vita Amandi)

morphological feature	subcategory	example
casus	NOMINATIVE	*Epistula* **scribenda** *est*
	GENITIVE	*Iniuriae* **obliviscendae** *sunt*
	DATIVE	*rei* **gerendae** *diem dicere*
	ACCUSATIVE	*ad bella* **suscipienda**
	ABLATIVE	*in libro* **legendo**
	VOCATIVE	**diligenda** *sponsa mea,* **diligende** *domne*
	OBLIQUE	*ad strepitum* **audiendo**
genus	MASCULINE	*ex hominibus* **tuendis**
	FEMININE	*consilium* **relinquendae** *Italiae*
	NEUTER	*consilii* **habendi**

morphological feature	subcategory	example
numerus	SINGULAR	*In **conservanda** patria*
	PLURAL	*criminibus **inferendis** delectari*
declensiontype	FIRST_DECLENSION	*consilium **relinquendae** Italiae*
	SECOND_DECLENSION	*in libro **legendo***

- IMPERATIVE

  The imperative modifies in **numerus**, **person**, **tense** (present imperative and future imperative), **voice**, **verbtype** and **conjugationtype**.

  → *'Christe, **adiuva** me! Christe, **adiuva** me!'* (Vita Amandi)

  → *'**Vade** et **dic** monacho illi [...].* (Vita Amandi)

  → ***Nolite**, cives, tantum admittere facinus.* (Vita Genovefae)

morphological feature	subcategory	example
numerus	SINGULAR	*adiuva!*
	PLURAL	*nolite!*
person	SECOND_PERSON	*abi!*
	THIRD_PERSON	*verentor!*
tense	PRESENT	*laudate!*
	FUTURE	*hortator!*
voice	ACTIVE	*monete!*
	PASSIVE	*monentor!*

- INDICATIVE

  Verbs in indicative modify in **numerus**, **person**, **tense**, **voice**, **verbtype** and **conjugationtype**.

morphological feature	subcategory	example
numerus	SINGULAR	*moneo*
	PLURAL	*prohibent*
person	FIRST_PERSON	*audior*
	SECOND_PERSON	*ageris*
	THIRD_PERSON	*partiuntur*
tense	PRESENT	*vereor*
	IMPERFECT	*agebatis*

	PERFECT	*laudavi*
	PLUPERFECT	*auxerat*
	FUTURE	*hortabor*
	FUTURE PERFECT	*audiverimus*
voice	ACTIVE	*agis*
	PASSIVE	*partiebar*

- INFINITIVE

  The infinitive is a verbal noun with properties of both verbs and nouns. It is annotated with its **voice**, **tense**, **verbtype** and **conjugationtype**.

  → *Longum est **enarrare** de tantis miraculis.* (Vita Audoini)

  → ***Memorare** vobis volo [...].* (Visio Baronti)

  → *Sufficit pauca **dixisse** de pluribus [...].* (Vita Audoini)

  → *Quis diceret, eum umquam vel de cibo aut potum **esse** satiatum?* (Vita Audoini)

  → *Cuius intercessione meritis **equari** non possumus [...].* (Vita Austrigisili)

morphological feature	subcategory	example
tense	PRESENT	*hortari*
	PERFECT	*suscepisse*
	FUTURE	*fore*
voice	ACTIVE	*fuisse*
	PASSIVE	*laudari*

- PARTICIPLE

  The participle is a verbal adjective. As for adjectives, pronouns and nouns, it declines for **casus**, **numerus** and **genus**. It is annotated with its **declensiontype, comparisondegree**, **tense**, **voice** and **verbtype**. There are three types of participles: the present participle, the past passive participle and the future active participle. Participles that are used as attributive adjectives are also considered as pos=V mood=PARTICIPLE.

  → *Protinus Genovefa **suspiciens** in caelum, manibus **expansis**, auxilium a Domino precabatur.* (Vita Genovefae)

  → *Ex illa hora, **fugato** demonio, puella illa pristine sanitati **restituta** est [...].* (Vita Austigisili)

→ *'Eia', inquit Eligius, 'eia, probabis hoc in die illa, cum te illuc coram populis et principibus* **visurus** *sum in palam'.* (Vita Eligii)

morphological feature	subcategory	example
casus	NOMINATIVE	*cernens*
	GENITIVE	*dicentis*
	DATIVE	*derepto*
	ACCUSATIVE	*intendentes*
	VOCATIVE	*amati*
	ABLATIVE	*sapiente,* **flagranti** *sudio*
	OBLIQUE	*suscepta(m)*
genus	MASCULINE	*praecipiens*
	FEMININE	*adpraehensa*
	NEUTER	*dicturum*
numerus	SINGULAR	*dicens*
	PLURAL	*dicentibus*
declensiontype	FIRST_DECLENSION	*nata*
	SECOND_DECLENSION	*dilecto*
	THIRD_DECLENSION	*tacentem*
comparisondegree	POSITIVE	*prudens*
	COMPARATIVE	*neglegentior*
	SUPERLATIVE	*amantissimus*
tense	PRESENT	*laudans*
	PERFECT	*amatus*
	FUTURE	*auditurum*
voice	ACTIVE	*audientis*
	PASSIVE	*acturus, compositum*

Remarks

o Some **adjectival participles** partially lost the connection to the verb and may share properties of adjectives. In spite of the adjective properties, adjectival participles are annotated as participles (pos=V mood=PARTICIPLE) (such as *ardens, egens, flagrans, imminens, pollens, sapiens, situs, diversus* etc.).

→ [...] *humilis atque* **sapiens** [...]. (Vita Austrigisili)

→ *Erat quidem corpore pulcher, sed pulchrior mente **pollens**que castitatem.* (Vita Boniti)

→ *Nam infra urbem nullus fere **egens** erat [...].* (Vita Desiderii)

→ *Et cum has easdem laudes **flagrans** amoris magnitudo exigeret [...].* (Vita Eligii)

- o **Past participles used as nouns** are annotated as pos=V mood=PARTICIPLE, for example *promissum* (*promitto*), *commissum* (*committo*), *decretum* (*decerno*), *gestum* (*gero*), *placitum* (*placeo*), *stratum* (*sterno*), *factum* (*facio*), *meritum* (*mereo*) etc.
- o All **deverbal nouns of the fourth declension** (e.g., *exerceo* → *exercitus, us*) are tagged pos=NN.

- SUBJUNCTIVE

  Verbs in subjunctive modify in **numerus**, **person**, **tense**, **voice**, **verbtype** and **conjugationtype**. There are four tenses: The present subjunctive, the imperfect subjunctive, the perfect subjunctive and the pluperfect subjunctive.

  → *Ne **timeas** venire ad me!* (Vita Galli Wettino)

  → *Et rogavit pontifices, ut per eorum canones **eligerent**, quem **voluissent**.* (Vita Galli Wettino)

  → *Alio vero tempore cum Eligius monasterium suum in territurio Lemoveceno **visitasset*** [...]. (Vita Eligii)

morphological feature	subcategory	example
numerus	SINGULAR	*moneam*
	PLURAL	*audiatis*
person	FIRST_PERSON	*audiamur*
	SECOND_PERSON	*agereris*
	THIRD_PERSON	*capiatur*
tense	PRESENT	*capias*
	IMPERFECT	*monerentur*
	PERFECT	*placueritis*
	PLUPERFECT	*fuissemus*
voice	ACTIVE	*sis*
	PASSIVE	*vereamur*

- SUPINE

  The supine is in form a noun of the fourth declension, annotated with its **casus** (ACCUSATIVE, DATIVE or ABLATIVE), **genus** (NEUTER), **verbtype** and **conjugationtype**.

  → *Sed mirum **dictu**!* (Vita Richarii)

morphological feature	subcategory	example
casus	ACCUSATIVE	***venatum*** *ire*
	DATIVE	*hoc facile est **intellectu***
	ABLATIVE	***memoratu*** *dignum*
genus	NEUTER	***nuptum*** *dare*
numerus	SINGULAR	*mirabile **dictu***

16.4 tense

The categories IMPERATIVE, INDICATIV, INFINITIVE, PARTICIPLE and SUBJUNCTIVE are annotated with their **tense**.

16.5 voice

The categories IMPERATIVE, INDICATIV, INFINITIVE, PARTICIPLE and SUBJUNCTIVE are annotated with their specific **voice** (ACTIVE or PASSIVE).

16.6 verbtype

All verbs are annotated with their **verbtype**. Seven different types of verbs are distinguished.

- DEPONENT

  Group of verbs that have passive forms but active meanings. These verbs are annotated as PASSIVE in **voice**.

  → […] *integram **adeptus** est sanitatem.* (Vita Genovefae)

  → *Verum post breve tempus **sequitur** effectus eius sermonem;* (Vita Eligii)

  → […] *qui ex uno **manantur** fonte vitis, uno sapore **sacientur** sicientes* (Vita Pardulfi)

- IMPERSONAL

  Third person singular verb forms without personal subject are conjugated only in the THIRD_PERSON singular and in the infinitive. The *token licuit* is reduced to the *lemma licet.*

→ […] *sed, quod dicere **pudet**, egestion naturae turpi impetus prorumpens* […]. (Vita Galli Walahfrido)

→ *Tres mulieres habuit, quod ei non **licuerat*** […]. (Viso Baronti)

→ *Post haec denique tempus **evenit**, ut* […]. (Vita Romarici)

Remark

Licet is labelled as pos=V mood=IMPERSONAL or as pos=CON depending on whether it is used as an impersonal verb or a as a concessive subordinator (see annex 2).

- INTRANSITIVE

Verbs that do not take any direct object and may not be classified into any other verbtype category (DEPONENT, IMPERSONAL, SEMIDEPONENT, VERBA_ANOMALA or VERBA_DEFECTIVA) are tagged INTRANSITIVE. Perfect passive participles of transitive verbs are annotated verbtype=INTRANSITIVE (except verba deponentia).

→ *Quidam ergo circumstantium partibus filii **favebant*** […]. (Vita Eligii)

→ *Imperiis tuis, beatissime papa Dydo, eloquio quidem exiguo, sed voto uberrimo **parere** disposui.* (Vita Amati)

- SEMIDEPONENT

Semi-deponent verbs have active forms in the first two principal parts, but deponent forms in the third principal part: *audere, fidere, gaudere, solere, reverti.*

→ *Qui cum cognovisset adventum venerabilis patris, **gavisus** est valde et honorifice suscepit illum* […]. (Vita Trudonis)

→ *Quapropter non **audeo** adgredi invalidus in sensu* […]. (Vita Pardulfi)

→ *Quem ipse locum utpute care coaemptum unice diligebat et plerumque ibique ad manendum divertere **solitus** erat.* (Vita Desiderii)

→ *Cui sacerdos, cui id officium commissum erat, orare innixius et **fidere** de merita praesulis monuit.* (Vita Desiderii)

- TRANSITIVE

Verbs that take its object in the accusative case, and which may not be classified into any other verbtype category are annotated as TRANSITIVE. Verbs (which may not be classified into any other verbtype category) followed by quotations, indirect questions

or subordinate clauses functioning as direct objects are also considered as TRANSITIVE.

→ *Omnia ista mala, quod **dicitis**, supercalcavit ista bona.* (Visio Baronti)

→ *Quis, inquid rex, hanc preceptionem **dedit**?* (Vita Austrigisili Biturigi)

- VERBA_ANOMALA

  Verbs that do not form their tenses according to the common rules: such as *esse, (com)edere, ferre, velle, nolle, malle, fieri, ire*. These word forms are annotated without conjugationtype.

  → […] ***ire** cum magistri non potuit, ut nobis viam veritatis ostenderet.* (Vita Galli Walahfrido)

  → […] *licet **nolens** quidem ad aulam regiam perrexit.* (Vita Wandregiseli)

- VERBA_DEFECTIVA

  Verbs that only possess forms in the perfect system, such as *odisse, meminisse, coepisse, aio* and *inquit*. These word forms are annotated without conjugationtype.

  → […] *et **memento** mei semper.* (Vita Amati)

  → *Tunc ergo **ait** ad illos sanctus Rafahel* […]. (Visio Baronti)

16.7 conjugationtype

All verbs are annotated with their conjugationtype except VERBA_ANOMALA and VERBA_DEFECTIVA. The regular verbs may be subdivided into four conjugations: first conjugation (*a*-stems), second conjugation (*ē*-stems), third conjugation (*ě*-stems), fourth conjugation (*i*-stems).

- FIRST_CONJUGATION (*a*-stems)

  → *Super quem sanctus Rafahel **oravit** prolixę et hunc capitulum aedidit ex ore, dicens* […]. (Visio Baronti)

- SECOND_CONJUGATION (*ē*-stems)

  → *Quid possum, bone pastor, donare, quia hic prae manibus nihil **habeo.*** (Visio Baronti)

- THIRD_CONJUGATION (*ě*-stems)

  → *Sed illi nequaquam **adquiescebant** [...].* (Visio Baronti)

- FOURTH_CONJUGATION (*i*-stems)

  → […] *quem iam nullus sperabat **audire** loqui.* (Visio Baronti)

Oriri is consistently annotated as a fourth conjugation verb because of its present infinitive (even though it can also be listed in the third *i*-stem conjugation).

17. anomaly

Grammatical variations in the use of cases, gender, number, prepositions, mode and diathesis, as well as anacoluthons and the use of absolute constructions such as nominative absolute and accusative absolute in Medieval Latin texts is marked as 'deviations' from Classical Latin grammar.

	description	examples
1.	ACCUSATIVE_ABSOLUTE[8]	***data susceptaque** de pace **sacramenta** pactionibusque firmatis*
2.	VARIATION_DIATHESIS	*coortaret* (*lemma: coorto, classicallemma: cohortor, classicalword: cohortaretur*); *manantur* (*lemma: manor, classicallemma: mano, classicalword: manant*); *sacientur* (*lemma: sacior, classicallemma: satio, classicalword: satiant*); *Cum asidue beatissime Gaugerici fama,* […]*, in bonis operibus **provocaret*** (*lemma: provoco, classicallemma: provoco, classicalword: provocaretur*); *persequere* (*lemma:*

---

[8] In the Excel sheet, German abbreviations are used for the different types of grammatical variation. After the export of the text files, the German abbreviations are substituted by the corresponding English terms:
ACCUSATIVE_ABSOLUTE for AA
VARIATION_DIATHESIS for ADG
VARIATION_GENDER for AGG
VARIATION_CASE for AKG
VARIATION_MODE for AMG
VARIATION_NUMBER for ANG
VARIATION_PREPOSITION for APG
VARIATION_DECLENSION for DU
LACK_CONGRUENCE for KF
VARIATION_CONJUGATION for KU
NOMINATIVE_ABSOLUTE for NA
ANACOLUTHON for SB.

		persequo, classicallemma: persequor, classicalword: persequi); *dilabere* (*lemma: dilabo, classicallemma: dilabor, classicalword: dilabi*)
3.	VARIATION_GENDER	*vestro quieti; iudicius* (*classicalword: iudicium*); *armae* (*classicalword: armorum*)
4.	VARIATION_CASE	*Iudaeus quidam Priscus* **nomen**
5.	VARIATION_MODE	e.g., *ut/ne/si* clauses or indirect speeches without subjunctive
6.	VARIATION_NUMBER	*omnis populus ibidem commanentes*
7.	VARIATION_PREPOSITION	**in** *ira maxima commoti;* **in** *exilio damnare*
8.	VARIATION_DECLENSION	*villabus; thesauribus; certaminem; accolabus; animabus*
9.	LACK_CONGRUENCE	*ex quibus* **quae melius fuit** *recipientes, ad Chilpericum retulerunt*
10.	VARIATION_CONJUGATION	*spondidit* (*classicalword: spospondit*), *cadebis* (*classicalword: cades*), *erupui* (*classicalword: erupi*)
11.	NOMINATIVE_ABSOLUTE	**hi contemnentes** *eum assurgere ei nemo voluit*
12.	ANACOLUTHON	*Lupus vero dux Companensis, cum iam diu a diversis fategaretur et spoliaretur assiduae et praesertim ab Ursione et a Berthefredo, ad extremum conventione facta ut occideretur,* **commoverunt exercitum contra eum**

Annex 1

List (incomplete) of adverbs and conjunctions.

adverb (pos=ADV)	conjunction (pos=CON)
acsi	**acsi**
adhuc	antequam
alibi	aut
aliubi,	autem
alicunde	cum
aliter	donec
aliquo	dum
aliquando	dummodo
antea	enim

cur	*et…et*
denuo	*etiam*
diu	*etsi*
ea	**licet**
ecce	*nam*
eiusmodi	*nec*
eo	*neque…neque*
ergo	*nisi*
exinde	*non solum…sed etiam*
foris/foras	*postquam*
forte	*priusquam*
fortuito, fortuitu	**quamdiu**
forsitan, fortasse	*quamvis*
huiusmodi, huiuscemodi	**quando**
iam	**quasi**
ibi	*-que*
ibidem	*quia*
ideo	*quin*
igitur	**quod**
illuc	*quodsi*
inde	*quominus*
inantea	*sed*
invicem	*seo, seu*
istac/illac	**si (se)**
ita	*simul ac/atque*
item	*sive*
itaque	*tamen*
merito	**ubi**
modo	**ut**
multum	*-ve*
nihil, **nil**	**vel**
non	
nusquam	
obviam	
passim	
paulum, **paululum**	
postergum	
postremo	
postremum	
postea	
potius	
praesertim	
primo	
profecto	
proinde	
qua	
qualiter	
quam	
quamdiu	
quamlibet	
quamobrem	

quantum
quapropter
quare
quasi
quid
quatenus
quoad
quomodo
quoque
sic
sicut
sicuti
tam
tantum
tum
ubi
ubicumque
unde
undecumque
undique
ut
utrimque
vel
vero
verum
videlicet

prepositions used as adverbs (incomplete):
ante, circum, contra, extra, infra, intra, intus, palam, post, propter, subter, super, supra, prae, prope

interrogative particles:
-ne, num, num, nonne, utrum…an, si (*se*)

Annex 2

List (incomplete) of words which are annotated differently depending on the context.

- ***ad invicem***
  - ***ad***: Adposition (pos=AP)
  - ***invicem***: Adverb (pos=ADV)
- ***ambo***: *pos=*PRO *pronountype=*INDEFINITE

- *ante*, *circum*, *clam*, *contra*, *extra*, *foris/foras*, *infra*, *intra*, *intus*, *palam*, *post*, *propter*, *supra*, *super*:
  - Adverb (pos=ADV)
  - Adposition (pos=AP)
- *is*, *ea*, *id*
  - Demonstrative pronoun (pos=PRO pronountype=DEMONSTRATIVE use=ADJECTIVAL): ***eum montem*** *ascendunt*; ***eo tempore***
  - Personal pronoun (pos=PRO pronountype=PERSONAL use=SUBSTANTIVAL): [...] *fingentes se illud in thesauro regis recondere, suis* ***eum*** *domibus intulerunt* [...]
  - Possessive pronoun (pos=PRO pronountype=POSSESSIVE use=ADJECTIVAL): *nunc autem filius tuus regnat,* ***regnum****que* ***eius*** *non tua, sed nostra tuitione salvatur*
- *licet*
  - Verb (pos=V mood=IMPERSONAL): *nos frui* ***liceret***
  - Conjunction (pos=CON): *quamvis* ***licet***
- *multum*
  - Adverb (pos=ADV): ***multum*** *loqui*
  - Adjective (pos=ADJ): *post* ***multum*** *tempus*
- *modo*
  - Adverb (pos=ADV): *semel* ***modo***
  - Noun (pos=NN): *hoc* ***modo*** *dicens*
- *nihil*, *nil*
  - Adverb (pos=ADV): ***nihil*** *nisi*; ***nihil*** *aliud quam*
  - Indefinite pronoun (pos=PRO pronountype=INDEFINITE): ***nihil*** *mali*; *aliquem* ***nihil*** *putare*
- *paululum*, *paulum*
  - Adverb (pos=ADV): *paululum*
  - Adjective (pos=ADJ): *paululus, a, um*

- *qamdiu*
  - Adverb (pos=ADV): *ut eundem locum adierint, et quid in conventu populi beatus Gallus fecerit, qualiter oratorium restauraverint, et* **quamdiu** *ibidem constiterint*
  - Conjunction (pos=CON): *ut nullus discipulorum eius,* **quamdiu** *vixerat, quid intus servaretur, cognoscere potuisset*
- *quamlibet*
  - Adverb (pos=ADV): **Quamlibet** *nos ad ministrandum gubernandumque rerum [...] cura constringat, nihil tamen tam principale quam principe dignum est, ut [...]* (Formulae Marculfi)
  - Indefinite pronoun (pos =PRO pronounype=INDEFINITE): [...] *si roget eos quispiam, num illa definitio possit in aliam rem transferri* **quamlibet**?
- *quando*:
  - Adverb (pos=ADV): **Quando**?
  - Conjunction (pos=CON): *tum* **quando** *legatos Tyrum misimus*
- *quasi*, *acsi*
  - Adverb (pos=ADV): **quasi** *per amicitiam monere*
  - Conjunction (pos=CON): *medico tria milia iugerum dedisti,* **quasi** *te sanasset*
- *quid*
  - Adverb (pos=ADV): **Quid** *id tu quaeris*? 'Why?'
  - Interrogative pronoun (pos=PRO pronountype=INTERROGATIVE): **Quid** *?* 'What ?'
- *quod*
  - Conjunction (pos=CON)
  - Relative pronoun (pos=PRO pronountype=Relative)
  - Interrogative pronoun (pos=PRO pronountype=INTERROGATIVE)
- *si* (*se*)
  - Conjunction (pos=CON)
  - Adverb (pos=ADV): as a question particle in sentence questions
- *singuli*
  - Adjective (pos=ADJ): *singulus, a, um*
  - Distributive number (pos=DIST): *singuli, ae, a*

- ***tantum***, ***quantum***
  - Adverb (pos=ADV)
  - Correlative pronoun (pos=PRO pronountype=CORRELATIVE)
- ***ubi***
  - Adverb (pos=ADV)
  - Conjunction (pos=CON)
- ***ut***
  - Adverb (pos=ADV): *gens,* ***ut*** *barbari credunt*
  - Conjunction (pos=CON): *constituit,* ***ut*** *ludi fierent*
- ***uter***
  - Indefinite pronoun (pos=PRO pronountype=INDEFINITE)
  - Interrogative pronoun (pos=PRO pronountype=INTERROGATIVE)
- ***vel***
  - Adverb (pos=ADV): *navem* ***vel*** *usque ad oceanum*
  - Conjuction (pos=CON): *mittere fortuna populi posita est in unius voluntate vel moribus*

Annex 3

List of tokens (incomplete) with different spellings or with changes of declension/conjugation/gender/diathesis which are subsumed under one *classicallemma.*

- ***abba***, ***abbas***: *classicallemma=abbas*
- ***arma*** (f., sing.), ***arma*** (n., pl.): *classicallemma=arma* (n., pl.)
- ***castra***, ***castrum***: *classicallemma=castrum*
- ***cinis***, ***cinus***: *classicallemma=cinis*
- ***dilabere***, ***dilabi***: *classicallemma=dilabor*
- ***equus***, ***eques***: *classicallemma=equus*
- ***filiis***, ***filiabus***: *classicallemma=filia*
- ***fortuitu***, ***fortuito***: *classicallemma=fortuito*
- ***honor***, ***honos***: *classicallemma=honor*
- ***idolatria***, ***idololatria***: *classicallemma=idolatria*
- ***iter***, ***itinerem***: *classicallemma=iter*
- ***iudicius***, ***iudicium***: *classicallemma=iudicium*
- ***Moses***, ***Moyses***: *classicallemma=Moyses*
- ***persequere***, ***persequi***: *classicallemma=persequor*

- *protinus*, *protenus*: *classicallemma=protinus*

- *sive*, *seu*: *classicallemma=sive*

- *Spania*, *Espania*, *Hispania*: *classicallemma=Hispania*

- *urgere*, *urguere*: *classicallemma=urgeo*

- *villis*, *villabus*: *classicallemma=villa*

Annex 4

The following table shows the lemmatization of irregular comparative and superlative adjectives. In some cases, the positive form has been lost over time. These adjectives are redressed to the comparative form.

positive/lemma	comparative	superlative
exter	*exterior*	*extremus*
infer	*inferior*	*infimus/imus*
superus	*superior*	*supremus/summus*
citer	*citerior*	-
potis	*potior*	*potissimus*
senex	*senior*	-
iuvenis	*iunior*	-

positive	comparative/lemma	superlative
-	*interior*	*intimus*
-	*prior*	(*primus* is tagged as pos=ORD)
-	*propior*	*proximus*
-	*deterior*	*deterrimus*
-	*ulterior*	*ultimus*

References

Bonnet, Max. 1968. *Le latin de Grégoire de Tours*. Hildesheim: Georg Olms Verlagsbuchhandlung.

Buchner, Sabina; Eller-Wildfeuer, Nicole; Eufe, Rembert; Greule, Albrecht; Hackl-Rössler, Sabine; Selig, Maria. 2017. „Die Legenden der merowingischen Münzen des Münzkabinetts Berlin aus sprachwissenschaftlicher Sicht. Historische, numismatische und philologische Untersuchung auf Grundlage des Bestandes im Münzkabinett der Staatlichen Museen zu Berlin". In: Greule, Albrecht; Kluge, Bernd; Jarnut, Jörg; Selig, Maria (edd.): *Die merowingischen Monetarmünzen als interdisziplinär-mediaevistische Herausforderung*. Paderborn, 125–154 (Mittelalterstudien des Instituts zur Interdisziplinären Erforschung des Mittelalters und seines Nachwirkens, Paderborn, Band 30).

CompHistSem. 2016. „Computational Historical Semantics. Beta", https://comphistsem.org/home.html (last visited October 2018).

Du Cange, Charles du Fresne; Bénédictins de Saint-Maur, Carpentier, Pierre; Henschel, G. A. Louis; Favre, Léopold. 1883–1887. *Glossarium mediae et infimae latinitatis*. Niort: L. Favre.

Georges, Karl Ernst; Georges, Heinrich. [1913/1918] [8]1998. *Ausführliches lateinisch-deutsches Handwörterbuch. Aus den Quellen zusammengetragen und mit besonderer Bezugnahme auf Synonymik und Antiquitäten unter Berücksichtigung der besten Hilfsmittel ausgearbeitet*. 2 Bände. Darmstadt: Wissenschaftliche Buchgesellschaft.

Jenset, Gard B.; McGillivray, Barbara. 2017. *Quantitative historical linguistics. A corpus framework*. Oxford: University Press.

Korkiakangas, Timo; Passarotti, Marco. 2011. „Challenges in annotating medieval Latin charters". In: *Journal of Language Technology and Computational Linguistics*. 26, 103–114.

Niermeyer, Jan Frederik; Van de Kieft, Co; Burgers, Johannes Wilhelm Jan. [2]2002. *Mediae Latinitatis Lexicon Minus* (édition remaniée). Darmstadt: Wissenschaftliche Buchgesellschaft.

Philippart De Foy, Caroline. 2012. „Lemmatiser un corpus de textes hagiographiques: enjeux et modalités pratiques". In: *Latin vulgaire – latin tardif IX. Actes du IXe colloque international sur le latin vulgaire et tardif*, Lyon 2-6 septembre 2009. Lyon: Maison de l'Orient et de la Méditerranée Jean Pouilloux, 481–490.

Pinkster, Harm. 2015. *The Oxford Latin Syntax. Volume I. The Simple Clause*. Oxford: University Press.

Stotz, Peter. 1998. *Handbuch zur lateinischen Sprache des Mittelalters. Vierter Band. Formenlehre, Syntax und Stilistik*. München: Beck.

TXM. 2007. „Projet Textométrie", http://textometrie.ens-lyon.fr/?lang=fr (last visited October 2018)

Väänänen, Veikko. 1981. *Introduction au latin vulgaire*, Paris: Éditions Klincksieck.