

Formae reformandae: for a reorganisation of verb form annotation in Universal Dependencies illustrated by the specific case of Latin

Flavio Massimiliano Cecchini

Università Cattolica del Sacro Cuore / Largo Gemelli 1, 20123 Milan, Italy

flavio.cecchini@unicatt.it

Abstract

Nonfinite verb forms are a crosslinguistically widespread phenomenon that poses a challenge to universal annotation formalisms like Universal Dependencies (UD), often clashing with traditionally established, language-specific conventions and terminologies. This paper, using Latin as a concrete case study, aims to give a survey on the `VerbForm` feature distribution among UD treebanks and to suggest a restructuring thereof in a universal perspective.

1 Introduction

The project of Universal Dependencies (UD) aims to harbour syntactically and morphologically annotated treebanks of any language, on the basis of universal, crosslinguistically valid tagsets for parts of speech (UPOS), morpholexical features and syntactic dependency relations (de Marneffe et al., 2021). The latest release, v2.9 of November 2021 (Zeman et al., 2021), of its second version (Nivre et al., 2020) sees 122 languages with at least one treebank each for a total of 217 treebanks, among which four Latin treebanks, variously distributed on diachronic, diastratic and diaphasic axes: Classical and Late Latin works, from prose (e. g. *De bello gallico* by Caesar), poetry (e. g. *Metamorphoses* by Ovid) and religious sources (e. g. Bible) in Perseus (Bamman and Crane, 2011) and PROIEL (Eckhoff et al., 2018), Early Medieval notarial Latin from Tuscia in the Late Latin Charter Treebank (LLCT) (Korkiakangas and Passarotti, 2011; Cecchini et al., 2020a), polished Late Medieval Latin of philosophical-theological texts by Thomas Aquinas in the Index Thomisticus (IT-TB) (Passarotti, 2019; Cecchini et al., 2018) and treatises (e. g. *De vulgari eloquentia*), poetry and personal correspondence by Dante Alighieri in UDante (Tavoni, 2011; Cecchini et al., 2020b), for a total of nearly 1 million tokens. Together, they testify the livelihood of Latin as a political, scientific and literary *lingua franca* in the centuries, well up into the modern age, that followed the fall of the Western Roman Empire, where it was the administrative and everyday language of a continent-spanning dominion (Waquet, 2001; Clackson and Horrocks, 2007; Leonhardt, 2009).

Variation over the whole of UD treebanks is even greater, most importantly with respect to linguistic phenomena and grammatical traditions that have to converge into the universal formalism. This does not always take place without conflicting or incoherent annotational choices and styles, both inter- and intralinguistically: these often arise, on the one side, from the direct translation of language-specific conventional denominations into misleadingly homonymous UD labels, and on the other side from some unfortunate naming choices in UD itself borrowed from historical, Latin-influenced traditional grammars of (mostly) European languages. This paper focuses on one of such “front lines”: the values of the morpholexical feature `VerbForm`, which in UD defines the different kinds and behaviours of forms in a verbal paradigm, and in particular of those values for nonfinite forms. While the discussion is centered around the UD formalism, it is also of a more general nature about the tension in the identification of language-specific and universal classes,¹ and the paper also puts forward new criteria for the annotation of Latin nonfinite verb forms as a sort of practical example for its conclusions. We notice that, especially

¹On this tension, cf. the discussion in (Croft, 2001, §2) about the identification of universal parts of speech, and in (Haspel-math, 2009) about terminology for morphological case.

from a terminological point view, Latin has a particular position in this context, given the wide-reaching authority of its historical grammatical works.

In §2, the framework of this paper is briefly expanded upon; in §3, the particular case study of the supine in UD is detailed for Latin and other languages; in §4, a concise survey of `VerbForm` values' distribution in UD is given, commented in more generic terms in §5; in §6, a new system for Latin nonfinite `VerbForms` in UD is proposed; finally, §7 closes this work with some minor proposals.

2 Finite and nonfinite verb forms and `VerbForms`

In UD, inflectional, verbal features include `VerbForm`, glossed as “form of verb or deverbative”.² At a universal level, the definition of `VerbForm` is thus left somewhat underspecified and vague, and the single values for this feature are only sketchily illustrated by few examples. A first fundamental distinction is between the value `Fin` and all others. The notion of “finite” verb form is not unquestionable, as it is originally anchored in the Latin verbal system: there, it is a verb form expressing an agreement in person (`Person`) with the subject (therefore “conjugated” with it, from *coniugo* ‘join together’), a tense (`Tense`) and a mood (`Mood`), and is the only kind of `VERB` form that can stand as head of an independent, unmarked clause without auxiliaries (`AUX`). This definition is not applicable crosslinguistically (Koptjevskaja-Tamm, 1994), as can already be seen by the world-wide distribution of verbal person marking (Siewierska, 2013).³ We will however refrain from attempting a direct universal definition of “finite” verb form, due to the non-trivial connected issues that go beyond the scope of this paper, and will leave it as a “primitive notion”, contenting ourselves of its operative definition for Latin, for which it may make sense. Still, we note that a characterisation of finiteness “by negation” emerges from the identification of cardinal nonfinite forms in §5: then, finite forms are all the others. In a typological perspective, this conclusion would actually suggest to abandon the notion of “finiteness” altogether, as it does not so much correspond to an actual linguistic category as it is a language-specific co-occurrence of factors (in Latin, the expression of person, tense and mood being compulsory for main, independent predicates).⁴ So, we observe a correlation between finite forms and independent clauses on one side, and nonfinite forms and embedded or subordinated clauses, see (Pinkster, 1990, §7, §8), on the other. The latter are situated in a grey area where verbs assume morphosyntactic behaviours which are usually ascribed to other parts of speech (see §5), and this characteristic makes nonfinite forms difficult to pinpoint, hence the many nonfinite values for `VerbForm` in UD in contrast to only one finite value `Fin`, and the even more detailed record of cases and denominations by traditional grammars.

Nonfinite forms are a complex phenomenon that warrants attention and that so to speak puts a crosslinguistic annotation formalism like UD to the test, hence they have been chosen as the main subject of this paper. Further, they are significantly present in Latin, so that Latin itself can serve as a valid proving ground for claims and proposals on the matter.

3 The supine as case study of a nonfinite `VerbForm`

The implementation of the `VerbForm` label `Sup` for the so-called supine across UD treebanks is a very specific, but nonetheless good case study for the misunderstandings that arise when confronting a crosslinguistic formalism like UD with those of traditional grammars of single languages, and for the confusion of focus between particular and more general phenomena at different annotational layers (cf. §4).

²In the following the general reference is the documentation found on the UD website (<https://universaldependencies.org/>), in particular the resumptive page for all morpholexical features (<https://universaldependencies.org/ext-feat-index.html>) and the page about `VerbForm` (<https://universaldependencies.org/ext-feat-index.html#verbform>), with all related universal and language-specific documentation. All data is also retrievable through each treebank's hub page, and queries on them can be performed by means of different tools, e. g. online with Grew (Guillaume, 2021).

³Hence we note that this terminological choice is unfortunate from a universal point of view, since it cannot be based on the original Latin notion of (*in*)*finitus* ‘(in)definite (form)’; cf. §4 about other `VerbForm` denominations.

⁴We note that a similar reasoning might be put forth, at least for Latin, for the “positive” degree (`Degree=Pos` in UD) of an adjective or similar element, since its definition actually corresponds to the *absence* of a comparative or absolute degree.

To address the issue, we will first give a rather detailed presentation of the Latin supine, proposing a new, more typologically grounded way to annotate it, before turning to so-called supines in other languages and their representation in UD.

3.1 The Latin supine

In traditional Latin grammars, the supine form, or “mood”,⁵ is described among others as “a verbal abstract of the 4th declension [...], having no distinction of tense or person” (Greenough et al., 2014, §508)⁶ or more explicitly as “a defective verbal noun” (Barbieri, 1995, §150).⁷ In such works, cf. e. g. (Palmer, 1988, p. 324f.), it is noted that the supine only appears in the singular number, either in the accusative or ablative case (but dative might also be attested), assigning to the former an alleged “active” and to the latter an alleged “passive” voice, however on no clear morphosyntactic grounds. Its use is limited as the complement of a) verbs expressing “directionality” (motion, giving/taking, sending), e. g. *frumentatum missa* fetch.grain-SUP-ACC⁸ send-PFV.PASS.PTCP-ABL.SG.F ‘sent to fetch grain’ (PROIEL, 53469),⁹ in the accusative, and b) of adjectives of evaluation, e. g. *difficile factu* difficult-NOM.SG.N do-SUP-ABL ‘difficult to do’ (PROIEL, 86346), in the ablative. Also recorded, even if marginally, but simply a subcase of the “active” use, is the periphrastic construction of the so-called “future”¹⁰ passive infinitive, formed with the passive (with impersonal meaning) present infinitive *iri* from *eo* ‘to go’, as in e. g. *has tibi redditum iri [putabam]* this-ACC.PL.F you-DAT give.back-SUP-ACC go-IPFV.INF-PASS [think-IPFV.IND.PST-1SG] ‘[I thought] these would be returned to you’ (PROIEL, 225189). This supplies the supposed lack of a construction with the auxiliary *sum* ‘to be’ as in the very infrequent (cf. §6, Table 2) active equivalent [*eum*] *has tibi redditurum esse [putabam]* (constructed), using the so-called future participle instead.

The uses of the supine alternate with some infrequent constructions with the infinitive, such as the so-called infinitive of purpose as in *meridie bibere dato* noon-ABL.SG drink-IPFV.INF-ACT give-IMP.FUT-2SG ‘give (them) to drink at noon’ (Greenough et al., 2014, §460f.) (cf. the supine *potum dedi* drink-SUP-ACC give-PFV.IND.PRES-1SG ‘I have given (you) to drink’ PROIEL, 224782), which might be reminiscent of the origin of this verb form (Palmer, 1988, p. 319f.). In modern Romance languages the supine has been indeed replaced by the infinitive in all contexts, e. g. *ha mandato a dire* ‘(he/she/it) sent to tell (i. e. let know)’ (Italian-VIT, VIT-8312; Alfieri and Tamburini (2016)) or *difficile da raggiungere* ‘difficult to reach’ (Italian-VIT, VIT-242),¹¹ and a direct descendant seems to survive only in Rumanian, as in *e de mirat cum trăiește* ‘it’s amazing how he lives’¹² (Mallinson, 1988, §4).

The use of supine is already very sparse in the Latin data at our disposal: across all Latin UD treebanks, it occurs a mere 17 times¹³ (5 times in the ablative), 16 of which are found in PROIEL, and it is so totally absent in corpora representing later varieties, but for one unusual case in UDante (Mon-644). However, morphologically identical abstract deverbative nouns of the 4th declension, annotated as NOUNs, are very common. In the UDante treebank (the only one where this information is available as of UD v2.9) more than 75% of fourth-declension noun (NOUN/PROPN) lemmas, i. e. 90 out of 120 (for a total of 421 out of 522 occurrences), are traceable back to supine forms,¹⁴ e. g. *spiritus* ‘spirit’ from *spiro* ‘to breath’. These

⁵In Latin grammars, the term “mood”, beside indicative, subjunctive and imperative, often also encompasses nonfinite verb forms, even if these do not actually express a MOOD in UD’s sense.

⁶Quite uniquely, this grammar lists the supine under other participles, probably in the absence of a better choice.

⁷The excerpts from this grammar are presented here in the translation by the author.

⁸here and thereafter, the gloss SUP stands for “supine”

⁹All quotations from UD corpora report the respective corpus code and sentence id (*sent_id*), while constructed examples are labelled as such. Forms under discussion are bold-faced, while arguments relevant to the discussion are underlined. Only for Latin samples, a linear gloss is given in the Leipzig formalism (see <https://www.eva.mpg.de/lingua/resources/glossing-rules.php>). A general reference for Latin is the classic grammar by Allen and Greenough (2014), or Barbieri’s (1995).

¹⁰The aspectual notion of prospectivity would be probably more fitting here; cf. §6, Table 1.

¹¹Translatable into Latin with a supine as *mittit dictum* and *difficile peruentu* (constructed).

¹²Where *de mirat* would correspond to Latin ablative supine *miratu*, or also dative *miratui*, from *miror* ‘to marvel at’. We note that Rumanian treebanks do not use the value Sup, but seem to prefer Part or a treatment as NOUN instead.

¹³Tokens with UPOS VERB and VerbForm=Sup.

¹⁴Corresponding in Word Formation Latin (WFL), a resource for Latin derivational morphology (Litta and Passarotti, 2019),

nouns appear marked for every possible case and nominal dependency relation.¹⁵ This observation could actually suggest, from a synchronic perspective,¹⁶ to give purely nominal interpretations to the occurrences of supines: on the one hand, parallel to the “active” supine we have the accusative of direction (Greenough et al., 2014, §388b) as in *ire Hyerosolimam* go-IPFV.INF-ACT Jerusalem-ACC.SG ‘go to Jerusalem’ (PROIEL, 13700), and on the other hand, parallel to the “passive” supine, the ablative of respect or specification, as in *uirtute praecedunt* virtue-ABL.SG excel-IPFV.IND.PRES-3PL ‘they excel in courage’ (Greenough et al., 2014, §418). While the latter interpretation seems justified, in the former case we can effectively find the supine taking arguments the same way as a corresponding finite predicate, e. g. *Tigranem ires salutatum* Tigranes-ACC.SG go-IPFV.SUB.PST-2SG greet-SUP-ACC ‘(so that) you would go to greet Tigranes’ (PROIEL, 76590), justifying its interpretation as a predicate, but then only limited to its accusative form.

The point of these observations is that the `VerbForm` value `Sup` appears, already on a language-internal point of view, not sufficiently focused: it seems to be based more on its etymological origin as a deverbative noun rather than on its synchronic function. But it would not be desirable to use `Sup` to represent a derivational process: derivational morphology is not really the focus of UD morphological features, and otherwise, for coherence, other deverbal nouns should also be marked for their derivation.¹⁷ At the same time, it is reasonable to assume that the function of a verbal form like the crystallised accusative of the supine in Latin can be found in other languages, and so, in a crosslinguistic perspective, another typological label would be better suited (cf. §6).

3.2 Supine in other languages and `Sup` in UD

The label of “supine” is variously used in grammars of languages other than Latin, but often the connection with the Latin supine does not appear fully motivated from a morphosyntactic point of view, a fact that corroborates the observations about the language-specificity of this label. As of UD v2.9, apart from Latin, 9 languages¹⁸ make use of this feature; unfortunately, a documentation page is available for only 3 of them: Old Church Slavonic, Slovene and Swedish. Slovene and Swedish are also cited as examples in the universal guidelines, while the language-specific documentation for Old Church Slavonic and Slovene is essentially identical, mirroring their close kinship. It follows an overview of these two supines, together with the Estonian one (the only non-Indo-European language together with North Sami) and their alleged relationship with the Latin supine:

Slovene (and some other current or old Slavic languages): it is an indeclinable verb form whose formation is linked to that of the infinitive (differing relatively to an *-i* suffix), apparently contrasting with it by expressing intentionality (Greenberg, 2006, §4.1.1.8), only appearing after verbs of motion “instead of infinitive” (UD guidelines) and capable of taking its own arguments: *grem domev sežgat dnevnik* ‘I’m going home to burn (my) diary’ (Priestly, 1993, §3.2.1, §4.5). If we ignore morphological differences, the communality on the syntactic level is thus partial, involving only the Latin accusative supine. We are not in the presence of a deverbative noun, but rather of a variant of the infinitive with lexically determined complementary distribution, so the `Inf` value might be a more fitting choice.¹⁹

mainly to rules 107 and 119 (conversions), 627 (suffix *it*), and 748 (suffix *at*). We notice that such figures seem to point out the fact that the prototypic Latin 4th-declension noun actually *is* a supine, as it were; however, a more thorough investigation over more extensive data is needed to support this claim.

¹⁵As a search for tokens satisfying `UPOS=NOUN`, `Gender=Masc`, `Inf1Class=IndEurU`, and, where present, `VerbForm=Sup`, in one of the Latin treebanks using these features can confirm.

¹⁶In Old Latin, contrarily, it is surely the case that deverbative nouns in general can take arguments like a corresponding finite predicate; cf. (Clackson and Horrocks, 2007, §4.2.3, c, iii).

¹⁷For example the extremely productive *(t)io(n)* suffix, as in *uisio* ‘vision’ from *uideo* ‘to see’, which according to WFL accounts for 2684 forms out of 14 418 recorded nouns (Litta and Passarotti, 2019).

¹⁸Estonian, Faroese, Icelandic, Marathi, North Sami, Old Church Slavonic, Old East Slavic, Slovene, Swedish.

¹⁹Diachronically, though, it is true that the Slavonic supine seems to be derived from the accusative case of a *u*-stem deverbative noun, too (Schenker, 1993, §3.2.2). This might be a further reason for its denomination.

Swedish : it is an indeclinable variant of the past (i. e. perfective) participle²⁰ used in the periphrastic construction of the composite past, based on the auxiliary verb *ha* ‘to have’, while in the passive construction (which has a predicative origin similarly as in Latin) the participle agrees in gender and number with the subject: *Jag har ätit maten* ‘I have eaten dinner’ vs. *Maten är äten* ‘Dinner is eaten’ (UD guidelines). As such, no connection whatsoever can be found with the Latin supine. Judging from the data, the notion of supine in Faroese and Icelandic is the same as in Swedish.

Estonian (not documented in UD): it essentially appears to be the inflected form of the infinitive under a different name, i. e. the infinitive fulfilling oblique roles, as opposed to the proper infinitive, used for core arguments (Viitso, 1998, p. 139). In this context, the infinitive in the illative case, representing motion to a place, comes closest to the “active” supine of Latin, as in *lähen malet mängima* ‘I’m going (somewhere) to play chess’; but also *olin klubis malet mängimas* ‘I was in the club playing chess’ (inessive), with no parallels. The similarity with the Latin supine is again only syntactic and partial, and, representing a paradigmatic variation, like in Slovene it seems better captured by *Inf*, especially since it seems part of a full inflectional paradigm (with possible suppletive forms), and not defective like the Latin supine.

To sum it up, the uses of the *Sup* label outside Latin treebanks tendentially seem to rest upon vague parallels between the syntax of generic infinitival constructions and that of the Latin supine in the “active” construction (see §3.1): resemblances are however at best only partial (most often they do not include the “passive” usage of Latin), or are only part of a more extensive paradigm (like in Estonian). Beside that, none of these forms appear to be used in non-predicative constructions as is the equivalent Latin noun. Finally, we note that the generic status of “verbal noun” does not *per se* justify a preference for *Sup* with respect to other possible labels like *Inf* or *VNoun* (or even *Conv*), especially when these have a better appeal in the respective languages. As a comparison, it is interesting to notice that the grammatical tradition of Finnish, a language very closely related to Estonian, just uses the denomination “(third) infinitive” for the exact equivalent (including inflection, cf. Abondolo (1998)) of the Estonian “supine”, as reflected by the use of the value *Inf* and absence of *Sup* in Finnish UD treebanks.

The real main reason for these traditional denominations thus seems to lie in the language-internal need to terminologically differentiate similar and correlated forms, e. g. infinitive in Slovene and Estonian, and participle in Swedish. Unfortunately, this brings along all the problems of excessive specificity already discussed for Latin in §3.1, and, on a typological level, is further misleading in that it establishes very specific, but not really grounded, parallels with the Latin form (the “original one”, as it were), which are wanting also from a purely morphosyntactic point of view.

4 Distribution of nonfinite *VerbForm* values in UD

The picture that emerges from the discussion in §3.2 is that of a label, *Sup*, employed in UD not so much on the basis of morphosyntactic consideration, as for assonance, in deference to prior grammatical traditions; such traditions are themselves based on simultaneously superficial and too focused syntactic resemblances to phenomena originally studied for Latin, whose grammar, from ancient times, has long represented the “ideal grammar” in Europe.²¹ With this, it is not meant that such distinctions are not useful or motivated internally to the given language but that, regrettably, these more or less successful attempts at following in the footsteps of Latin grammar terminology do not allow for meaningful inter-linguistic comparisons, and often even contribute to the establishing of inaccurate analogies. A similar state of things appears also from the use of other nonfinite *VerbForm* values UD-wide, of which an overview follows (as of v2.8; where not specified, quotations are from universal guidelines) :

²⁰Originally, its neuter singular form (Andersson, 1994, p. 284f.).

²¹While for a long time Latin grammarians themselves resorted to Greek grammar canon to frame Latin, see (Clackson and Horrocks, 2007, §6), also (Law, 2003, §4); see the pioneering work of Priscianus (Keil, 1855), and also cf. e. g. how the diverging phenomenon of *ablativus absolutus* was approached, as detailed in (Sluiter, 2000).

Conv employed by 36 languages for a verb form “which shares properties of verbs and adverbs”, consequently appearing in an adverbial function, and so identified principally at a syntactic level (following the UD definition of **ADV** as “words that typically modify verbs for such categories as time, place, direction or manner”). Despite this value, some languages like Slovene and Latin opt to annotate possible candidates directly as **ADVs** (and consequent relation `advmod` instead of `advcl`), e. g. *sufficienter* suffice-IPFV.ACT.PTCP-ADV ‘sufficiently’.²² We note that the term “converb” first appeared in the field of Altaic studies (Haspelmath, 1995, §7) and has never been part of traditional terminologies of European languages.

Gdv employed by 4 languages (including Latin). The universal guidelines briefly state “used in Latin and Ancient Greek”. While in Latin the gerundive is a kind of participle (see §6), the documentation for Armenian defines it as “a nonfinite verb form that shares properties of verbs and nouns”, which would rather fit with **VNoun**. So, also this label appears to arise from traditional denominations, without being supported by a morphological definition; it is highly language-specific, and as such has not spread beyond these few languages (3 of which, Latin, Ancient Greek and Sanskrit, represent ancient phases of modern Indo-European languages, to which Armenian also belongs).

Ger employed by 21 languages (including Latin), even if deprecated by the universal guidelines. Here, the difference with **Conv** is not clear: e. g. the Italian gerund *Arrivando tardi si perde il treno* ‘Arriving late you miss the train’ (Italian guidelines) looks equivalent to Czech transgressive *udělavši večeři, zavolala rodinu ke stolu* ‘having prepared the dinner, she called her family to the table’ (universal guidelines), where it is labelled as **Conv**. Notably, while the latter inflects for gender, number and aspect, the former is invariable. The term derives from the Italian *gerundio* being the direct descendant of Latin *gerundium*, cf. again (Haspelmath, 1995, §7), itself a *gerundivum* (**Gdv**) in a particular syntactic context (see §6), of which it has kept the name despite radical morphosyntactic changes. In English treebanks, the use of **Ger** is contextual (but it is probably the case that we are dealing with two different homographic forms here): the same form is labelled as **Part** instead when preceded by the auxiliary verb *to be*; it is described (also universally) as “shar[ing] properties of verbs and nouns”, which would rather lead to **VNoun**.

Inf employed by 75 languages (including Latin), it is together with **Part** the most universally used value, and at the same time the most undefined. Neither the universal guidelines nor any language-specific documentation put forward any true definition; the wide-spread identification as a citation form is of course purely conventional and extremely language-specific. Infinitive seems to be treated as a linguistic “primitive notion”, self-evident for the languages it is applied to. However, the documentation for Irish, stating that “[t]he infinitive verb form is the same as the verbal noun”, lets one question if **Inf** and **VNoun** are not actually referring to the same entity (see **VNoun** and cf. §6).

Part employed by 75 languages (including Latin), with a general agreement on it representing a verb form “shar[ing] properties of verbs and adjectives”. This identification is bound to happen principally on a syntactic (but possibly also semantic) level, as for the UPOS **ADJ** itself, defined in the guidelines as “words that typically modify nouns and specify their properties or attributes”. In some languages (as in Latin) morphological criteria might also be applied, but this is not a universal fact, and it is more often than not a consequence of a word being an adjective rather than the opposite. There can be contradictions, however: notwithstanding that the Latin gerundive shares morphology and syntax with other participial forms (see §6), the historical difference in naming convention (**Gdv** vs. **Part**) has been carried over into UD treebanks.

VNoun employed by 15 languages, it stands for “[v]erbal nouns other than infinitives”; however, being the value **Inf** undefined (see **Inf**), this leaves place for arbitrariness, and the distinction is not motivated. Indeed, despite the cases (cf. §3.2) in which this label could be appropriate, there seems to be

²²Which is potentially accompanied by *suffecte* suffice-PFV.PASS.PTCP-ADV and *suffecture* suffice-PROSP.ACT.PTCP-ADV, thus showing a paradigmatical variation in aspect/voice, all from the same verb.

a general preference for `Inf`, probably influenced by Western naming traditions. Only 9 languages use both labels:²³ this complementarity goes into the direction of a factual equivalence of the two labels. Indeed, the Turkish documentation explicitly mentions this fact, claiming a preference for `VNoun`. In UD, `NOUNs` are defined as “a part of speech typically denoting a person, place, thing, animal or idea”, pointing to mainly semantic criteria for their identification.

5 Identification of cardinal `VerbForm`s in UD

From the overview in §4, an explicit distinction emerges between those values whose definition is oriented towards a specific UPOS with all respective morphosyntactic (and semantic) implications, i.e. `Conv`~`ADV`, `Part`~`ADJ` and `VNoun`~`NOUN`, and the remaining ones (`Gdv`, `Ger`, `Inf`, `Sup`), whose definitions are left undetermined and/or which stem directly from traditional, language-specific denominations; no less than 3 (`Gdv`, `Ger`, `Sup`) originally refer to entities or constructions extremely peculiar to the Latin language, and have been adopted with various degrees of consistency by other grammatical traditions (cf. §3.2 and §4). Another issue is that such very specific labels isolate peculiar syntactic constructions which are not necessarily related to morphology, and which obscure the more general picture. The three UPOS-oriented values `Conv`, `Part` and `VNoun` can instead be seen as cardinal choices that logically reflect all possibilities contained in the morphosyntactic system of UD. They follow straight from the intuition that a verb form that keeps its lexicality continues to be head of a clause that can be itself embedded in a matrix clause in the same way as a non-verbal, i.e. nominal, phrase: so, in the UD formalism, either “mimicking” an adjective (`ADJ`), an adverb (`ADV`) or a noun (`NOUN/PROPN`), i.e. each and every lexical nominal part of speech in UD.

In the end, if the feature `VerbForm` stands to represent the morphosyntactic (and to some extent also semantic) properties that a verbal stem can assume in its inflectional paradigm (cf. §2), all the while keeping the possibility to act as the equivalent predicate of a main, independent clause with respect to its arguments, then, in agreement e.g. with Haspelmath (1995), we argue that a set of values mirroring all possible logically corresponding parts of speech in the given annotation formalism should suffice: in the case of UD, then, those which are conventionally labelled as `Conv`, `Part` and `Inf/VNoun`.²⁴ Consequently, the other labels are not actually needed and, on the contrary, do not contribute to the goal of inter-linguistic comparison implicit in the UD project, since they usually arise from idiosyncratic, language-specific terminology that conflicts with universal labels; we also argue that they can be all traced back to the three cardinal categories, or to specific (syntactic) behaviours of other deverbative parts of speech (not being truly part of a verbal paradigm in such a case, cf. §3.1). This will be done for Latin in the next section. Finally, we note that such reorganisation of verb forms around cardinal, UPOS-oriented values would not alter the extant possibility to assign a `VerbForm` to a non-`VERB` token, as such an assignment is of etymological rather than morpholexical character, and points to the paradigmatic origin of the form in question, not to its synchronic use; given the part-of-speech label, there subsists no ambiguity about this double connotation of the `VerbForm` feature.

6 Reorganising nonfinite Latin verb forms

As seen in §4, UD Latin treebanks currently make use of five out of seven values for nonfinite `VerbForm`s, i.e. `Gdv`, `Ger`, `Inf`, `Part` and `Sup`. Their implementation is comparable between treebanks, as it more or less regularly follows traditional definitions. Below, Latin nonfinite forms are examined from a morphological and a syntactic point of view. Other considerations concerning when and whether some forms should be considered `VERBs` or else,²⁵ are out of the scope of this investigation.

²³Erzya, Irish, Komi Zyrian, Mbya Guarani, Moksha, Polish, Skolt Sami, Turkish, Turkish German. But: Irish claims their identity (see `Inf`); `Inf` does not appear in the Turkish language-specific documentation; some of these languages share common annotation principles (e.g. Uralic languages, under the code `urj`; cf. Partanen et al. (2018, §3)). Thus, actual figures are lower.

²⁴Noting that `VNoun` is a better choice, being less language-specific than `Inf`, and that `Part` should also be relabelled in

Denomination & example	VerbForm current	VerbForm proposed	Aspect	(Tense)	Voice	InflClass [nominal]	Gender	Number	Case
Perfect participle <i>actus/a/um</i>	Part	Part	Perf	(Past)	Pass	IndEurA/ IndEurO	*	*	*
Present participle <i>agens</i>	Part	Part	Imp	(Pres)	Act	IndEurI	*	*	*
Future participle <i>acturus/a/um</i>	Part	Part	Prosp	(Fut)	Act	IndEurA/ IndEurO	*	*	*
Present infinitive <i>agere, agi</i>	Inf	VNoun	Imp	(Pres)	Act/Pass	Ind	<i>Neut</i>	<i>Sing</i>	<i>Acc/Nom</i>
Perfect infinitive <i>agisse</i>	Inf	VNoun	Perf	(Past)	Act	Ind	<i>Neut</i>	<i>Sing</i>	<i>Acc/Nom</i>
Gerund <i>agendo/um/oi</i>	Ger	Part or VNoun	Prosp	(Fut/Pres)	Pass	IndEurO	<i>Neut</i>	<i>Sing</i>	<i>Abl/Acc/ Dat/Gen</i>
Gerundive <i>agendus/a/um</i>	Gdv	Part	Prosp	(Fut/Pres)	Pass	IndEurA/ IndEurO	*	*	*
Supine <i>actu/um</i>	Sup	Conv (or NOUN)	Prosp	-	Act	IndEurU	<i>Masc</i>	<i>Sing</i>	<i>Abl/Acc</i>

Table 1: Morphological properties of Latin nonfinite verb forms expressed in the UD formalism, with proposed `VerbForm` relabellings. Values for `Tense` are shown following their use in treebanks for legacy reasons only, since tense is not applicable to Latin nonfinite forms, cf. e. g. (Pinkster, 1990, §11.2.2). Legend: asterisk = all values possible; italics = inherent or contextual, not morphologically expressed values, i. e. not matched in the actual form (infinitives are indeclinable); hyphen = not observed; or = a different annotation might be possible for some contexts (see text). The example forms are limited to singular nominatives where possible, else all forms are listed. The value `Voice` is intended in a purely morphological, and not syntactic (clausal), sense.

Morphology Table 1 shows, in UD terms, the possible sets of values corresponding to the morphological features that are expressed by Latin nonfinite verb forms.²⁶ Notably, `Mood`, `Tense` and `Person` are absent, as in Latin they are a prerogative of so-called finite forms (see §2); `Degree`, being only optional, is also not shown. The split between two groups is evident: irrespective of different combinations of `Aspect` and `Voice`, one group (participles and gerundives) follows (prototypic) adjectives in not having an inherent, but only a relational²⁷ `Gender/Number`, inflecting for `Case` according to so-called 1st- (“*a* & *o* stems”) and 2nd-class (specifically, “*i* stems”) adjectival paradigms, and the possibility of being marked for `Degree` (e. g. *ardentiori* burn-IPFV.ACT.PTCP-CMPR-DAT.SG ‘more burningly’, UDante, Mon-283);²⁸ conversely, the other group (infinitives and supine) is similar to nouns, in that its members are bound to one given inflectional paradigm and/or possess a fixed, inherent gender and number, while case varies (even if defectively), and cannot express degree. Therefore, from this point of view, we have a natural partition into `Part`-forms and `VNoun`-forms, as discussed in §5. This means that both `Ger` and `Gdv` would be superseded by `Part`; morphologically, the identity of these two forms, specifically of the gerund as a particular case of gerundive, seems to be out of question (Haspelmath, 1987; Miller, 2000; Jasanoff, 2006). These choices are in fact already substantiated by traditional grammars: gerundive is considered a participle in (Greenough et al., 2014, §500), which “expresses the action of the verb in the form of an Adjective” (Greenough et al., 2014, §488), and is “a verbal adjective” according to

this sense.

²⁵This problem becomes particularly relevant for later varieties of Latin. For example, cf. the treatment of *agens*, the present participle from *ago* ‘to drive, to act’, in the IT-TB (13th c. CE): either ‘driving, acting’, UPOS `VERB` (862 occurrences), or ‘agent’, UPOS `NOUN` (353 occurrences, including one incorrectly annotated as `VERB`). Conversely, no tokens with lemma *agens* are found in PROIEL.

²⁶While these “schemata” are quite uncontroversial, the identification of a prospective aspect for some forms probably does not represent a common opinion; however, it is to be seen as the natural aspectual counterpart to the traditionally claimed (but inapplicable, see Table 1) future tense.

²⁷That is, determined by agreement with another element, see relations in Table 2.

²⁸They can also take an adverbial form, but are then regularly annotated as `ADVs`, not `Convs`, by all UD Latin treebanks; see §4, `Conv`.

(Barbieri, 1995, §164) ; gerund “is the neuter of the gerundive” (Greenough et al., 2014, §501); infinitive is “properly a noun” that “often admits the distinction of tense” (Greenough et al., 2014, §451), “a neuter singular verbal noun” (Barbieri, 1995, §151). However, the extreme defectiveness of the supine, which, as a predicate (“active” supine; the “passive” supine is to be treated as a simple NOUN, which is the standard for Latin deverbative nouns in Classical literature), appears only in the accusative case (cf. §3.1), sets it apart from more regular verbal nouns and instead supports a reading as a different VerbForm, namely a converb Conv: this analysis is further corroborated by the distribution of its syntactic relations.²⁹ In the same way, the uses of the Gdv identified as Ger can be interpreted as veering towards a less relational VerbForm than Part, and thus VNoun; more under **Syntax**.

VerbForm current	VerbForm proposed	Denominations and respective frequencies in the data	Dependency relations with respective frequencies				
		Perfect participle (37.88%)	finite	acl	advcl		
			33.67%	33.38%	18.52%		
Part	Part	Present participle (18.96%)	advcl	acl	finite	root	
			36.39%	31.58%	9.86%	6.02%	
		Future participle (0.59%)	finite	ccomp	root	advcl	
			43.07%	35.10%	9.14%	5.01%	
Inf	VNoun	Present infinitive (30.58%)	xcomp	ccomp	csubj		
			60.78%	18.35%	14.10%		
		Perfect infinitive (1.14%)	ccomp	xcomp	root	csubj	
			50.61%	21.93%	10.58%	8.59%	
Ger	Part or VNoun	Gerund (6.73%)	advcl	acl			
			49.01%	45.46%			
Gdv	Part	Gerundive (4.08%)	finite	root	advcl	acl	ccomp
			37.31%	19.17%	14.55%	10.23%	10.01%
Sup	Conv (or NOUN)	Supine (0.03%)	advcl	finite			
			94.12%	5.88%			

Table 2: Distribution of nonfinite verb forms and their most frequent ($\geq 5\%$) dependency relations in UD Latin treebanks, broken down by traditional denominations. Only tokens with UPOS VERB, i. e. deemed to have the same argument structure as the predicate of a main, independent clause, and with VerbForm different than Fin, have been taken into consideration, for a total of 57 411 tokens. Relation subtypes (e. g. `advcl:pred` w. r. t. `advcl`) have been neutralised to compensate for different annotation styles. The underspecified relation `conj` (6439 occurrences) has been traced back and substituted with the relation of the respective co-ordination “head”. The label **finite** comprises all nonfinite forms which have a dependent node labelled with `aux` or `cop` (9510 occurrences): for all purposes, these combinations are or derive from finite, albeit periphrastic, predicates, and so their exact syntactic relations are no longer relevant in this context. Annotation errors and inconsistencies, together with elliptic clauses, produce noise in the figures: e. g., `root` and `ccomp` labels are in many cases clues for elliptic, periphrastic, finite predicates, as e. g. in the formula *dicendum est quod...* ‘it is to / will be said that...’, where *est* is the auxiliary ‘to be’ (IT-TB) and *dicendum* a gerundive.

Syntax Table 2 summarises the distribution of syntactic roles, as per UD dependency relations, in all available Latin treebanks.³⁰ As expected, nominal relations are negligible, and syntactic data appear to

²⁹We would like to thank an anonymous reviewer for pointing out this fact, which is *a posteriori* self-evident: another example of how traditional, entrenched points of view (“Latin has no converbs”) often stand in the way of typological awareness.

³⁰We note that this overview necessarily represents a mean of diachronic, diastatic and diaphasic varieties (cf. §1), but because of the increasing status, since late antiquity (ca 4th-5th c. CE), of Latin as an international and prestigious *lingua franca* rather than a living and native language, see (Clackson and Horrocks, 2007, §8), also (Wright, 1998; Leonhardt, 2009), we

be in nearly perfect agreement with morphology in Latin. So we observe that, on the one hand, infinitive forms are specialised as heads of clauses that fulfil core arguments, which are prototypically occupied by NOUNs (given the parallels *xcomp/ccomp~obj*, *csubj~nsubj*), and so *VNoun* becomes the natural choice here, as discussed in §4 and §5. On the other hand, the gerundive has the same profile as participial forms: it resembles the most the future participle (also for being rather infrequent), which might not be a coincidence considering the common prospective aspect, which appears to have been marginalised in favour of the main imperfective/perfective opposition.³¹ If the latter is a *Part*, then so is the gerundive; the different functional distributions of participles might be possibly explained in semantic terms tied to aspect and other features, and could mirror preferences with regard to which elements can appear as so-called secondary predicates (Pinkster, 1990, §8.3), but overall they are seen to fulfil attributive/predicative roles (cf. “dominant participles” Pinkster (1990, §7.4.7)). On the contrary, the supine, appearing nearly exclusively in an adverbial function, cannot itself agree with any subject as participles do (cf. Nikitina and Haug (2016)) and e. g. cannot appear in absolute constructions,³² confirming its status as *Conv*³³ (or *NOUN*), as seen for its effectively absent inflectional morphology (compared e. g. to the complete paradigm of a participle). Finally, the syntax of the gerund is more difficult to assess on a generic level: diachronic and diaphasic distinctions are needed. In fact, while the gerund is clearly an inflected form of the gerundive (Haspelmath, 1987; Miller, 2000; Jasanoff, 2006) and so supposedly passive, it is considered distinct from it on the syntactic ground that it can govern an object instead of a subject (without agreeing with it). For example, we find the adnominal (*acl*) [*necessitas*] *plura nomina deo dandi* [necessity-NOM.SG[.F]] more-ACC.PL.N noun-ACC.PL[.N] god-DAT.SG[.M] give-PROSP.PASS.PTCP-GEN.SG.N ‘[the need] of giving God more nouns’ (IT-TB, *train-s1483*), with direct object in the accusative case and uncontroversially transitive reading, in place of a Classically expected passive construction *plurium nominum dandorum* more-GEN.PL.N noun-GEN.PL[.N] give-PROSP.PASS.PTCP-GEN.PL.N, lit. ‘for more names to be given’, in the genitive (notice that this is a not head-coreferent clause embedded as an adnominal modifier, so there is no agreement with *necessitas*). In the same text (the *Summa contra gentiles*), we also find [*necessitatem*] *sustentandi corporis* [necessity-ACC.SG[.F]] sustain-PROSP.PASS.PTCP-GEN.SG.N body-GEN.SG[.N], lit. ‘[the need] of the body being sustained’, i. e. ‘to sustain the body’ (IT-TB, *train-s22169*). Thus, we agree with Haspelmath (1987, §5.2) that, especially when such an alternation can still be found in a significant ratio as in the IT-TB (565 vs. 509 occurrences respectively), the gerund can be simply explained in terms of a gerundive with impersonal value and deponent³⁴ (Greenough et al., 2014, §190) behaviour, and so, in the annotation, we can trace it back to *Part*. Further, of the 1459 identical³⁵ nonfinite clauses headed by a *Ger* in the UD Latin treebanks, at most 635 have a clear direct object: this means that for more than half of *Ger*-clause types a plainer interpretation as gerundives (*Part*) is also possible, and so preferable in general. But in some contexts, the at first only occasional (Miller, 2000) reanalysis from a passive to

can regard linguistic change in written sources as extremely moderate, “frozen” by the adherence to the prestigious Classical standards, in comparison to the contemporary processes that lead to modern Romance languages (Väänänen, 1981; Palmer, 1988; Ledgeway, 2012), which were gradual anyway, cf. Wüest (1998). So, an aggregated picture keeps its significance here.

³¹In fact, both forms (together with the supine) have disappeared in modern Romance languages, together with a morphologically expressed inchoative aspect, leaving only fossilised lexemes, see (Harris, 1988, §3). An explanation for this might be that prospective adjectival/adverbial forms are less time-stable than prototypical adjectives/adverbs, and so are preferentially expressed by “finite” predicates by languages, cf. (Stassen, 2003, §5), eventually leading, in the case of Latin, to their exclusion.

³²An absolute construction is a nonfinite embedded adverbial clause with a subject different from any actors of its matrix clause, and, at least for Latin, headed by a participial form which agrees in gender and the number with its subject, both in the ablative case.

³³Specifically, of purpose, with same or main subject than its matrix clause.

³⁴Deponency can be seen, in general terms, as a mismatch between canonical morphological and syntactical behaviours (Baerman, 2007): in Latin, this happens for verbs displaying a passive morphology, but a transitive/active syntactical behaviour, e. g. *sequor* ‘to follow (someone)’, receiving a direct object argument in the accusative case. For a brief sketch of the problem posed by Latin deponent verbs, cf. discussion at <https://github.com/UniversalDependencies/docs/issues/713>.

³⁵With the same forms in the same order, considering only the head and all possible core or functional dependent nodes (*obj*, *ccomp*, *xcomp*, *nsubj*, *csubj*, *mark*). So, *ad censum uobis perexoluendum*, *ad censum nobis perexoluendum* and *ad censum perexoluendum* ‘for tributes to be quitted [by you/us]’ (LLCT) are all considered the same clause. This equivalence is needed to deal with formulaic repetitions especially in LLCT, where e. g. *ad censum perexoluendum* alone is repeated 68 times, or (*ad legem et iustitia[m] faciendum* ‘to carry out law and justice’ 134 times.

an active construction of the spoken language may appear stabilised also in written documents: so, in the LLCT treebank, representing a Latin heavily influenced by early medieval Romance varieties, active gerunds like [*potestas*] *remittendi peccata* [power-NOM.SG[.F]] put.back-PROSP.PASS.PTCP-GEN.SG.N sin-ACC.PL[.N] ‘[the power] to forgive sins’ (LLCT, `train-s21623`) (instead of an expected *remittendorum peccatorum* put.back-PROSP.PASS.PTCP-GEN.PL.N sin-GEN.PL[.N], in the genitive) represent 649 (97.89%) occurrences among a total of 663 Gdv/GER constructions with core arguments, whereas only 14 clear cases of Classical gerundival constructions like in *ad dedicandam ipsam basilicam* to dedicate-PROSP.PASS.PTCP-ACC.SG.F same-ACC.SG.F basilica-ACC.SG[.F] ‘for this basilica to be consecrated’ (LLCT, `train-s835`) can be found. Of the former, 63 occurrences, like the 40 variations of *non pondum leuandum* not weight-ACC/NOM.SG[.N] lift-PROSP.PASS.PTCP-ACC/NOM.SG.N ‘(the) taking away not a (single) pound’ have an ambiguous reading, since the alleged object is a neuter singular like the gerund. The situation is reversed in PROIEL, more skewed towards Classical Latin: 49 (13.46%) gerunds with object and 315 (86.54%) gerundives with passive subject (out of 364 Gdv/GER with core arguments).

It so appears that the Latin nonfinite verbal system is naturally and, in a crosslinguistic prospective, effectively explained and annotated in terms of the only three labels Part, VNoun and Conv, avoiding the too language-specific and idiosyncratic values Gdv, GER and Sup, and substituting Inf with a more universal label. Even when using identical VerbForm values, all forms identified by traditional grammars are kept distinct by virtue of morphological features or syntactic dependencies; conversely, were two forms not to be distinguished at any level (like the object-, subjectless gerund), the reasons for keeping them distinct would become questionable.³⁶ Finally, the Latin system is, despite what could show through traditional grammar, seen to possess a Conv form, whose presence is however marginal already in Classical times and completely outshined by the use of so-called “conjoined participles” (*participia coniuncta*) and/or eminently participial/adjectival absolute constructions (*ablativus absolutus* and secondary predications) with similar adverbial functions. Only much later, in Romance varieties, a crystallised gerund takes on the form of a converb.

7 Conclusion and last remarks

This paper proposes a reorganisation of the annotation of nonfinite Latin verb forms in the UD formalism (§3.1, §6), accomplished with respect to the morpholexical feature VerbForm, situating it in the wider perspective of achieving a simpler and “more universal”, crosslinguistically valid system than the current one (§2, §5), highlighting the inconsistencies in its implementation across UD treebanks, also by Latin treebanks themselves (§3.2, §4). Latin has been chosen as a testbed, beyond showing extensive nonfinite verb formations and falling into the competences of the author (contributor to the IT-TB, LLCT and UDante Latin treebanks), especially because of its particular position at the origin, more or less foundedly, of a large part of (traditional) grammatical terminology, notably of European languages, which is encountered again in UD (e.g. Inf, Part, etc. for the feature VerbForm). Contributions from the work on typologically radically different languages would be a highly valued complement to the survey in §4, and to spark discussion about this topic in the UD community is one of the major goals of this paper.

We can lastly briefly mention two possible additions to the system, left for future examination, as a corollary to the discussion in §5: a) the introduction of a fourth nonfinite VerbForm value for highly specialised, frozen forms like the Swedish supine (§3.2), with a probable orientation towards an AUX-like category; b) the introduction of a “terminological feature”³⁷ that, parallelly to the constellation of UD morphosyntactic features/UPOS/reasons characterising a verb form, would help retrieve it through its traditional denomination, thereby acknowledging historical, common language-specific conventions without however interfering with the universal analysis.

³⁶And this seems to be the case for the Slovene infinitive and supine, who might be seen as the same form in a lexically determined, complementary allomorphic variation: then, the reading of intentionality (§3.2) would actually depend on the predicate rather than on the form itself.

³⁷Cf. discussion at <https://github.com/UniversalDependencies/docs/issues/775>.

References

- Daniel Abondolo. 1998. Finnish. In Daniel Abondolo, editor, *The Uralic Languages*, Language family descriptions, pages 115–148, London, UK; New York, NY, USA. Routledge.
- Linda Alfieri and Fabio Tamburini. 2016. (Almost) Automatic Conversion of the Venice Italian Treebank into the Merged Italian Dependency Treebank Format. In *Third Italian Conference on Computational Linguistics - CLiC-IT 2016*, pages 19–23, Naples, Italy, December. CEUR-WS.org.
- Erik Andersson. 1994. Swedish. In Ekkehard König and Johan van der Auwera, editors, *The Germanic Languages*, Language family descriptions, pages 271–312, London, UK; New York, NY, USA. Routledge.
- Matthew Baerman. 2007. Morphological Typology of Deponency. In Matthew Baerman, Greville G. Corbett, Dunstan Brown, and Andrew Hippisley, editors, *Deponency and Morphological Mismatches*, number 145 in Proceedings of the British Academy, Oxford, UK. British Academy (through Oxford University Press).
- David Bamman and Gregory Crane. 2011. The ancient Greek and Latin dependency treebanks. In Caroline Sporleder, Antal van den Bosch, and Kalliopi Zervanou, editors, *Language Technology for Cultural Heritage, Theory and Applications of Natural Language Processing*, pages 79–98, Berlin/Heidelberg, Germany. Springer.
- Giovanna Barbieri. 1995. *Nuovo corso di lingua latina*. Loescher, Turin, Italy.
- Flavio Massimiliano Cecchini, Marco Passarotti, Paola Marongiu, and Daniel Zeman. 2018. Challenges in Converting the Index Thomisticus Treebank into Universal Dependencies. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 27–36, Brussels, Belgium, November. Association for Computational Linguistics.
- Flavio Massimiliano Cecchini, Timo Korhakangas, and Marco Passarotti. 2020a. A New Latin Treebank for Universal Dependencies: Charters between Ancient Latin and Romance Languages. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 933–942, Marseille, France, May. European Language Resources Association.
- Flavio Massimiliano Cecchini, Rachele Sprugnoli, Giovanni Moretti, and Marco Passarotti. 2020b. UDante: First Steps Towards the Universal Dependencies Treebank of Dante’s Latin Works. In *Seventh Italian Conference on Computational Linguistics - CLiC-IT 2020*, pages 1–7, Bologna, Italy, March. CEUR-WS.org.
- James Clackson and Geoffrey Horrocks. 2007. *The Blackwell History of the Latin Language*. Blackwell Publishing, Malden, MA, USA.
- William Croft. 2001. *Radical Construction Grammar*. Oxford University Press, Oxford, UK.
- Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308, 07.
- Hanne Eckhoff, Kristin Bech, Gerlof Bouma, Kristine Eide, Dag Haug, Odd Einar Haugen, and Marius Jøhndal. 2018. The PROIEL treebank family: a standard for early attestations of Indo-European languages. *Language Resources and Evaluation*, 52(1):29–65.
- Marc L. Greenberg. 2006. *A Short Reference Grammar of Standard Slovene*. University of Kansas. Available online at http://www.seelrc.org:8080/grammar/pdf/stand_alone_slovene.pdf.
- J. B. Greenough, G. L. Kittredge, A. A. Howard, and Benjamin L. D’Ooge. 2014. *New Latin Grammar for Schools and Colleges*, volume 1. Dickinson College Commentaries, Carlisle, PA, USA. Available online at <https://dcc.dickinson.edu/grammar/latin/credits-and-reuse>.
- Bruno Guillaume. 2021. Graph Matching and Graph Rewriting: GREW tools for corpus exploration, maintenance and conversion. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 168–175, Online, April. Association for Computational Linguistics.
- Martin Harris. 1988. The Romance Languages. In Martin Harris and Nigel Vincent, editors, *The Romance languages*, Language family descriptions, pages 1–25, London, UK; New York, NY, USA. Routledge.
- Martin Haspelmath. 1987. Verbal noun or verbal adjective? The case of the Latin gerundive and gerund. Arbeitspapier N.F.3, Institut für Sprachwissenschaft, Universität Köln. Available online at <http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/24318>.

- Martin Haspelmath. 1995. Converbs in cross-linguistic perspective. Number 13 in *Empirical Approaches to Language Typology*, pages 1–55, Berlin, Germany. Mouton de Gruyter.
- Martin Haspelmath. 2009. Terminology of case. In Andrej Malchukov and Andrew Spencer, editors, *Handbook of Case*, pages 505–517, Oxford, UK. Oxford University Press.
- Jay H. Jasanoff. 2006. The Origin of the Latin Gerund and Gerundive: A New Proposal. *Harvard Ukrainian Studies*, 28(1/4):195–208.
- Heinrich Keil. 1855. Prisciani Institutionum grammaticarum libri XVIII (ex recensione Martini Hertzii). In *Grammatici latini*, volume 3, Leipzig, Germany. Teubner.
- Maria Koptjevskaja-Tamm. 1994. Finiteness. In Ronald E. Asher and J. M. Y. Simpson, editors, *The Encyclopedia of Language and Linguistics*, volume 3, pages 1245–1248, Oxford, UK. Pergamon Press. NB: This entry does not appear in the second edition of the *Encyclopedia*.
- Maria Koptjevskaja-Tamm. 2006. Nominalization. In Keith Brown, editor, *Encyclopedia of Language and Linguistics*, volume 8, pages 652–659, Amsterdam, Netherlands. Elsevier. Second Edition.
- Timo Korkiakangas and Marco Passarotti. 2011. Challenges in Annotating Medieval Latin Charters. *Journal for Language Technology and Computational Linguistics*, 26(2):103–114.
- Vivien Law. 2003. *The History of Linguistics in Europe*. Cambridge University Press, Cambridge, UK.
- Adam Ledgeway. 2012. *From Latin to Romance*. Number 1 in *Oxford studies in historical and diachronic linguistics*. Oxford University Press, Oxford, UK.
- Jürgen Leonhardt. 2009. *Latein. Geschichte einer Weltsprache*. Beck, Munich, Germany.
- Eleonora Litta and Marco Passarotti. 2019. (When) inflection needs derivation: a word formation lexicon for Latin. In Nigel Holmes, Marijke Ottink, Josine Schrickx, and Maria Selig, editors, *Words and Sounds*, pages 224–239. De Gruyter, Berlin, Germany; Boston, MA, USA, December. Interrogable online at <http://wfl.marginalia.it/>.
- Graham Mallinson. 1988. Rumanian. In Martin Harris and Nigel Vincent, editors, *The Romance languages, Language family descriptions*, pages 391–419, London, UK; New York, NY, USA. Routledge.
- D. Gary Miller. 2000. Gerund and gerundive in Latin. *Diachronica*, 17(2):293–349.
- Tatiana Nikitina and Dag Trygve Truslew Haug. 2016. Syntactic nominalization in Latin: a case of non-canonical agreement. *Transactions of the Philological Society*, 25–50(1):195–208.
- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France, May. European Language Resources Association.
- Leonard Robert Palmer. 1988. *The Latin Language*. University of Oklahoma Press, Norman, OK, USA.
- Niko Partanen, Rogier Blokland, KyungTae Lim, Thierry Poibeau, and Michael Rießler. 2018. The first Komi-Zyrian Universal Dependencies treebanks. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*, pages 126–132, Brussels, Belgium, November. Association for Computational Linguistics.
- Marco Passarotti. 2019. The Project of the Index Thomisticus Treebank. In Monica Berti, editor, *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, number 10 in *Age of Access? Grundfragen der Informationsgesellschaft*, Berlin, Germany; Boston, MA, USA. De Gruyter Saur.
- Harm Pinkster. 1990. *Latin Syntax and Semantics*. Romance linguistics. Routledge, London, UK; New York, NY, USA.
- T. M. S. Priestly. 1993. Slovene. In Bernard Comrie and Greville G. Corbett, editors, *The Slavonic Languages, Language family descriptions*, pages 388–451, London, UK; New York, NY, USA. Routledge.
- Alexander M. Schenker. 1993. Proto-slavonic. In Bernard Comrie and Greville G. Corbett, editors, *The Slavonic Languages, Language family descriptions*, pages 60–121, London, UK; New York, NY, USA. Routledge.
- Anna Siewierska. 2013. Verbal person marking. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. <https://wals.info/chapter/102>.

- Ineke Sluiter. 2000. Seven grammarians on the ablative absolute. *Historiographia Linguistica*, xxvii(2/3):379–414.
- Leon Stassen. 2003. *Intransitive Predication*. Oxford Studies in Typology and Linguistic Theory. Oxford University Press, Oxford, UK.
- Mirko Tavoni. 2011. DanteSearch: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica. In Anna Cerbo, Roberto Mondola, Aleksandra Žabjek, and Ciro Di Fiore, editors, *Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni*, volume 2 (2004–2005), pages 583–608, Naples, Italy. Il Torcoliere – Officine Grafico-Editoriali di Ateneo.
- Veikko Väänänen. 1981. *Introduction au latin vulgaire*. Klincksieck, Paris, France.
- Tiit-Rein Viitso. 1998. Estonian. In Daniel Abondolo, editor, *The Uralic Languages*, Language family descriptions, pages 149–183, London, UK; New York, NY, USA. Routledge.
- Françoise Waquet. 2001. *Latin or the Empire of a Sign*. Verso, Brooklyn, NY, USA; London, UK. Translation of: Françoise Waquet. 1998. *Le latin ou l'empire d'un signe*. L'évolution de l'humanité. Albin Michel, Paris, France.
- Roger Wright. 1998. Il latino: da madrelingua nativa a lingua straniera. In *La transizione dal latino alle lingue romanze. Atti della tavola rotonda di linguistica storica, Università Ca' Foscari di Venezia, 14–15 giugno 1996*, pages 77–85, Tübingen, Germany. Niemayer.
- Jakob Wüest. 1998. Pour une linguistique historique non linéaire: Les formes analytiques du latin. In *La transizione dal latino alle lingue romanze. Atti della tavola rotonda di linguistica storica, Università Ca' Foscari di Venezia, 14–15 giugno 1996*, pages 87–98, Tübingen, Germany. Niemayer.
- Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielè Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Bilge Nas Arıcan, Thórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Deniz Baran Aslan, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Starkadhur Barkarson, Rodolfo Basile, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Anouck Braggaar, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Lauren Cassidy, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Neslihan Cesur, Savas Cetin, Özlem Çetinoğlu, Fabricio Chalub, Shweta Chauhan, Ethan Chi, Taishi Chika, Yongseok Cho, Junho Choi, Jayeol Chun, Juyeon Chung, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Mihaela Cristescu, Philemon Daniel, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Elisa Di Nuovo, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Sandra Eiche, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaz Erjavec, Aline Etienne, Wograin Evelyne, Sidney Facundes, Richárd Farkas, Jannatul Ferdousi, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Gustavo Godoy, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteins-son, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỳ, Na-Rae Han, Muhammad Yulistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinecke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Olájdé Ishola, Kaoru Ito, Siratun Jannat, Tomáš Jelínek, Apoorva Jha, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Neslihan Kara, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Elena Klyachko, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Mehmet Köse, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sandra Kübler, Oğuzhan Kuyrukçu, Aslı Kuzgun, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phu'o'ng Lê H'ông, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Bruna Lima Padovani,

Krister Lindén, Nikola Ljubešić, Olga Loginova, Stefano Lusito, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Menel Mahamdi, Jean Maillard, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Büşra Marşan, Cătălina Mărânduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, Lorena Martín-Rodríguez, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Alessandro Mazzei, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Tatiana Merzhevich, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Judit Molnár, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Giovanni Moretti, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Manuela Nevaci, Lu'o'ng Nguyê'n Thị, Huyê'n Nguyê'n Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adéday`o Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvrelid, Şaziye Betül Özateş, Merve Özçelik, Arzucan Özgür, Balkız Öztürk Başaran, Hyunji Hayley Park, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Lapińska, Siyao Peng, Cenal-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Mizanur Rahoman, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Mathilde Regnault, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Putri Rizqiyah, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roşca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoal Sadde, Pegah Safari, Benoît Sagot, Aleksí Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Ezgi Sanıyar, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Shefali Saxena, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Lane Schwartz, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Syeda Shahzadi, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Yana Shishkina, Muh Shohibus-sirri, Dmitry Sichinava, Janine Siewert, Einar Freyr Sigurdhsson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Shafi Sourov, Carolyn Spadine, Rachele Sprugnoli, Steinthór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Chihiro Taguchi, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Dipta Tanaya, Samson Tella, Isabelle Tellier, Marinella Testori, Guillaume Thomas, Liisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uriá, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Rob van der Goot, Martine Vanhove, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Natalia Vlasova, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Paul Widmer, Sri Hartati Wijono, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Arife Betül Yenice, Olcay Taner Yıldız, Zhuoran Yu, Arlisa Yuliawati, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, He Zhou, Hanzhi Zhu, Anna Zhuravleva, and Rayan Ziane. 2021. Universal dependencies 2.9. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-4611>. Licence Universal Dependencies v2.9.