Formae reformandae

For a reorganisation of verb form annotation in Universal Dependencies illustrated by the specific case of Latin

Flavio Massimiliano Cecchini
flavio.cecchini@unicatt.it

Syntaxfest 2021 | Sofia, Bulgaria & rest of the world | March 23rd 2022

▶ **Universal:** "theory"
   ⟶ critical discussion of the *verbal inflectional feature* `VerbForm` in Universal Dependencies (UD)

▶ **Language-specific:** "practice"
   ⟶ demonstration of the results applying them to the **Latin** verbal system

Universal

How many forms can a verb take?

In UD, `VerbForm` is a *universal*, mostly verbal feature with the following *universal* **values**:

- ▶ `Fin` **finite** ⤷ employed by 74 languages
- ▶ `Conv` **converb** ⤷ employed by 36 languages
- ▶ `Gdv` **gerundive** ⤷ employed by 4 languages
- ▶ `Ger` **gerund** ⤷ employed by 21 languages
- ▶ `Inf` **infinitive** ⤷ employed by 75 languages
- ▶ `Part` **participle** ⤷ employed by 75 languages
- ▶ `Sup` **supine** ⤷ employed by 9 languages
- ▶ `Vnoun` **verbal noun** ⤷ employed by 15 languages

Other **non** universal values (not discussed): `Cop`, `FinRel`, `PartFut`, `PartPast`, `PartPres`, `PartRes`, `Post`, `Prov`, `Ser`, `Stat`, `Stem`

No explicit definition, but (from the guidelines):

*Some verb forms in some languages actually form a **gray zone** between verbs and **other parts of speech** (nouns, adjectives and adverbs)*

https://universaldependencies.org/u/feat/VerbForm.html

Supine: what is it?

Latin: a 4th-declension **verbal noun** occurring in accusative after "directional" verbs or in ablative after adjectives of evaluation

- ★ ***potum** dedi* (< *poto* 'to drink' accusative, "active")
- → 'I have given (it for you) **to drink**'

- ★ *difficile **factu*** (< *facio* 'to do'; ablative, "passive")
- → '(something) difficult **to do**'

Swedish: invariable participial form used in **composite tenses** (Faroese and Icelandic too?)

- ★ Jag har **ätit** maten (< *äta* 'to eat')
- → 'I have **eaten** dinner'

- ! Maten är **äten** (< *äta* 'to eat'; common gender in *-en* agreeing with *mat* 'dinner')
- → 'Dinner is **eaten**'

These two verb forms have **completely different** distributions and properties!

Universal verb forms?
Absent and conflicting standards

From the guidelines…

Inf **no definition**: a primitive notion?
- ! **Irish** documentation: *"[t]he infinitive verb form is the same as the verbal noun."*
  → `https://universaldependencies.org/ga/feat/VerbForm.html#inf-infinitive`
- ! Traditional **Latin** grammars: *"a neuter singular verbal noun"*, *"properly a noun"*

Vnoun *"verbal nouns other than infinitives"*: recursive **non-definition**
- ! **Turkish** documentation: *"Note that the forms with -mAK are often called the infinitive (Göksel and Kerslake, p. 412) and used as the citation form of the verb. Nevertheless, we annotate them as `VerbForm=Vnoun`, not `Inf`."*
- ⇒ The `Inf` label does **not** appear in the Turkish documentation
  → `https://universaldependencies.org/tr/feat/VerbForm.html#vnoun-verbal-noun`

Only 9 languages use them **both**, most prefer `Inf`: **complementary** labels?

⇒ Are `Inf` and `Vnoun` not **the same**?

Issue

Most of these terms are **transfered directly** from language-specific, often **Latin-influenced** grammatical traditions to UD's labels!

Too many labels, too **inconsistent**
⇒ They do **not** allow for **meaningful** typological comparisons

Another example:

Italian use of `Ger` (from it. *gerundio*) for a clear **converb**:

- ★ ***Arrivando** tardi si perde il treno* (< arrivare 'to come'; uninflectable form)
- → '**Arriving** late you miss the train'
- ! Only **diachronically** from Latin **gerund(ive)** (neuter ablative form, e.g. *veniendo* ∼ 'by the coming', from *veniendum*, from *veniendus/a/um*, from *venio* 'to come'); **synchronically completely different** properties

So, if (from the guidelines)…

> *Some verb forms in some languages actually form a **gray zone** between verbs and **other parts of speech** (nouns, adjectives and adverbs)*

…then as `VerbForm`**s** we may want to consider **only** correspondences with…

ADJ	ADV	NOUN
adjectives	**adverbs**	**nouns**
"participles"	*converbs*	*"infinitives"/"masdars"/verbal nouns*

⇒ **lexical**, **nominal** parts of speech in UD

Universal verb forms

The **new** typological framework for `VerbForm`**s**:

▶ (`Fin` **finite** $\overset{?}{\sim}$ `VERB`, i.e. **never acting** as another part of speech)

▶ `Conv` **converb** $\sim$ `ADV`

▶ ~~`Adv` **gerundive**~~ – **too language-specific** definition

▶ ~~`Ger` **gerund**~~ – **too language-specific** (and unclear) definition

▶ ~~`Inf` **infinitive**~~ – a **better label** already exists for the same form

▶ `Part` **participle** $\sim$ `ADJ` – **NB: renaming** might be desirable

▶ ~~`Sup` **supine**~~ – **too language-specific** definition

▶ `Vnoun` **verbal noun** $\sim$ `NOUN`

No other labels are needed!
↪ Traditional labels might be kept for ease of reference as a special "`Traditional`" feature

Other labels can be subsumed by **other features**, e.g. `Aspect` for Hungarian `PartFut`, `PartPast` and `PartPres`;
or simply by **syntactic dependencies** (e.g. Mbya Guarani `Post`, `Prev`?); or be simply replaced.

Latin

Latin corpora in universal dependencies

Latin is represented by five corpora in Universal Dependencies:

LLCT **Early Medieval** bureaucratic Latin from Tuscia

ITTB **Late Medieval** theological Latin by Thomas Aquinas

UDante **Late Medieval** Latin of poetry, treatises and letters by Dante Alighieri

PROIEL **Classic, Late and Christian** Latin from various sources

Perseus **Classic, Late and Christian** Latin from various sources

LLCT, **ITTB** and **UDante** are **developed** and **maintained** by the **CIRCSE** research centre
→ **improvement** and convergence of annotation
→ development of **guidelines** for Latin (work in progress!)

PROIEL and **Perseus** have "**neglected**" status

`http://quest.ms.mff.cuni.cz/udvalidator/cgi-bin/unidep/validation-report.pl`

Traditional grammars (usually) distinguish:

▶ **Finite forms:** those displaying **person** and **tense**
  → `Fin:` *amo*, *amas*, *amat*…

▶ **Non finite forms:**
  `Part` **Perfect participle:** "action of the verb in the form of an adjective; completed"
    → *amatus*, *amata*, *amatum*…
  `Part` **Present participle:** "action of the verb in the form of an adjective; in progress"
    → *amans*, *amantis*, *amantem*…
  `Part` **Future participle:** "action of the verb in the form of an adjective; still to take place"
    → *amaturus*, *amatura*, *amaturum*…
  `Inf` **Present infinitive:** "properly a noun denoting the action of the verb abstractly; in progress"
    → *amare / amari*
  `Inf` **Perfect infinitive:** "properly a noun denoting the action of the verb abstractly; completed"
    → *amavisse*
  `Ger` **Gerund:** "a verbal noun"
    → *amandum*, *amando*, *amandi*
  `Gdv` **Gerundive:** "adjective implying obligation, necessity, or propriety"
    → *amandus*, *amanda*, *amandum*…
  `Sup` **Supine:** "4th-declension noun used after verbs or adjectives"
    → *amatum*, *amatu*
      (example *amo* 'to love'; definitions from the *Latin Grammar* by Allen & Greenough, 1903)

Both **adjectival** forms ⇒ `Part` also for gerundives

Same **morphology**:

- ⋆ **complete inflection** per `Case`, `Gender`, `Number`
- ⋆ same **inflectional classes** as adjectives
- ⋆ can take `Degree`

Same **functions**:

- ⋆ attributive
  - → *oppidum* ***amatum*** 'the loved town', *oppidum* ***amandum*** 'the town to be loved'
- ⋆ periphrastic tenses:
  - → *oppidum* ***amatum*** *est* 'the town has been loved', *oppidum* ***amandum*** *est* 'the town is (∼needs) to be loved'
- ⋆ heads of embedded clauses:
  - → ***amato*** *oppido…* 'as the town has/had/will have been loved…' (ablativus absolutus),
    voluntas ***amandi*** *oppidi* 'the desire that the town will be loved' ∼ 'to love the town' (adnominal clause)

All **distinguished** by different combinations of `Aspect` and `Voice` values

gerundive: `Prosp` and `Pass`

Gerunds are **neuter singular forms** of the **gerundive** $\Rightarrow$ `Part` also for gerunds

However, **diachronically** we observe a **shift to transitive constructions** despite its originally **passive** `Voice` and subsequent **loss of agreement** with a subject:

- ▶ *[potestas] **remittendi** peccata* '[the power] to forgive sins' <small>(Early Medieval Latin, LLCT)</small>
- → *remittendi* is genitive <small>(nmod)</small>, *peccata* is accusative <small>(obj)</small>, no agreement in `Case/Number`

instead of **Classically expected**

- ▶ *[potestas] **remittendorum** peccatorum* lit. '[the power] of sins being forgiven'
- → *peccatorum* genitive <small>(nsubj:pass)</small>, agreement in `Case/Gender/Number`

$\Rightarrow$ in some **later varieties** of Latin the gerund can be analysed as a `Vnoun`

"Passive" (ablative) supine **cannot be distinguished** from other deverbal nouns ⇒ NOUN

▶ *difficile **factu*** 'difficult to do', i. e. 'with regard to the doing' (*factus*)

→ **ablative of specification** as in ***virtute*** *praecedunt* 'they excel in courage'

→ **deverbal nominals** are otherwise **never** annotated as VERBs in Latin

      e. g. ***amabilis*** 'lovely' ADJ from *amo* 'to love'; ***visio*** NOUN from *video* 'to see'

"Active" (accusative) supine is a **crystallised case form** always **associated to other verbs**
⇒ Latin's only (marginal) Conv

▶ *Tigranem ires **salutatum*** '(so that) you would go to greet Tigranes'

→ keeps the possibility to have **arguments** (e. g. here *Tigranem* obj)

→ **same or different** subject as main clause (e. g. ***potum*** *dedi* '**I** have given so that **you** drink')

Latin Overview

► **Finite forms** ⟶ `Fin` (…if it still makes sense)
  → *amo*, *amas*, *amat*…

► **Non finite forms**
`ADJ` **Perfect participle** ⟶ `Part` with `Aspect=Perf` and `Voice=Pass`
  → *amatus*, *amata*, *amatum*…
`ADJ` **Present participle** ⟶ `Part` with `Aspect=Imp` and `Voice=Act`
  → *amans*, *amantis*, *amantem*…
`ADJ` **Future participle** ⟶ `Part` with `Aspect=Prosp` and `Voice=Act`
  → *amaturus*, *amatura*, *amaturum*…
`ADJ` **Gerundive** ⟶ `Part` with `Aspect=Prosp` and `Voice=Pass`
  → *amandus*, *amanda*, *amandum*…
`ADJ/NOUN` **Gerund** ⟶ `Part` with `Aspect=Prosp` and `Voice=Pass` **or** (Late Latin) `Vnoun` with `Case`
  → *amandum*, *amando*, *amandi*
`NOUN` **Present infinitive** ⟶ `Vnoun` with `Aspect=Imp` and `Voice=Act|Pass`
  → *amare / amari*
`NOUN` **Perfect infinitive** ⟶ `Vnoun` with `Aspect=Perf` and `Voice=Act`
  → *amavisse*
`ADV` **Supine** ⟶ ("active") `Conv` with `Aspect=Prosp` and `Voice=Act` **or** ("passive") simply a `NOUN`
  → *amatum*, *amatu*

Vobis pro auditu gratiae agendae sunt!

Thank you for your attention!

Flavio Massimiliano **Cecchini**

Università Cattolica del Sacro Cuore di Milano
CIRCSE research centre
Largo Gemelli 1, 20123 Milano

✉ `flavio.cecchini@unicatt.it`

 `https://github.com/CIRCSE`
 `https://centridiricerca.unicatt.it/circse-home`
 `https://itreebank.marginalia.it/`
 `https://lila-erc.eu`