

Ellipsis in Enhanced Dependencies: A Case Study on Latin

Lisa Sophie Albertelli^{1,*}, Lorenzo Augello¹, Giulia Calvi¹, Annachiara Clementelli², Federica Iurescia³ and Claudia Corbetta^{4,2}

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

²Università di Pavia, corso Strada Nuova 65, 27100 Pavia, Italy

³Università Cattolica del Sacro Cuore, CIRCSE Research Centre, Largo Gemelli, 1, 20123 Milan, Italy

⁴Università degli studi di Bergamo, via Salvecchio 19, 24129 Bergamo, Italy.

Abstract

This paper presents the manual annotation of ellipsis phenomena in a Latin treebank syntactically annotated following Universal Dependencies framework. Building on ongoing research in the field, it provides an overview of syntactic constructions that pose particular challenges for the annotation and reconstruction of ellipsis. By providing Latin examples, the work contributes to cross-linguistic comparisons and the broader understanding of ellipsis across languages. As a preliminary contribution, the paper offers insights into aspects that remain underspecified in current UD guidelines, suggesting directions for future refinement of annotation standards.

Keywords

Universal Dependencies, Ellipsis, Latin, Enhanced Dependencies, Treebank

1. Introduction

This work describes manual annotation of ellipsis in treebanks, which are syntactically annotated texts. The source data is a portion of UD_Latin-CIRCSE Treebank, a treebank manually annotated following Universal Dependencies (UD) framework.¹ At the time of writing, the treebank consists of three tragedies authored by Seneca (1st CE) – *Hercules Furens*, *Agamemnon*, *Oedipus* – and a treatise authored by Tacitus (1st-2nd CE) – *Germania*. Tacitus' *Germania*, being a prose work, was chosen for the present study to circumvent the stylistic challenges associated with tragic texts, particularly those arising from their diverse metrical structures. We relied on gold data as in the 15th UD release.² Nevertheless, the annotation did present several challenges and provided valuable insights to support ongoing research in the field of ellip-

sis in UD. Section 2 describes the state of the art in this field. Section 3 illustrates some examples of ellipsis and outlines the challenges encountered during the annotation. Section 4 concludes the paper and outlines potential avenues for future research.

2. State of the Art

Before delving into the core of the present work, this section provides a general introduction to UD (subsection 2.1), ellipsis (subsection 2.2), and the annotation of ellipsis in UD (subsection 2.3).

2.1. Universal Dependencies

Universal Dependencies is a framework for morphosyntactic annotation of different human languages [1]. The aim is to provide support for Natural Language Processing (NLP) researches and typologically oriented linguistic studies. In its most recent release, it includes 319 treebanks covering 179 languages.³ UD offers two layers of annotation: basic syntactic annotation and enhanced syntactic annotation.⁴ How they differ in the strategies adopted to annotate ellipsis, is the topic of subsection 2.3.

CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 – 26, 2025, Cagliari, Italy

*Corresponding author.

[†]This paper is the result of the collaboration between the authors. For the specific concerns of the Italian academic attribution system, Federica Iurescia is responsible for sections 1 and 3; Claudia Corbetta for sections 2; 2.1; 2.2; Lisa Sophie Albertelli for section 2.3; 3.1; section 3.2 was collaboratively written by Lisa Sophie Albertelli and Giulia Calvi. Giulia Calvi is responsible for section 3.3; Lorenzo Augello for sections 3.4; 3.5; Annachiara Clementelli for section 3.6. Section 4 was collaboratively written by Federica Iurescia and Claudia Corbetta.

✉ lisasophie.albertelli01@icatt.it (L. S. Albertelli);
lorenzo.augello01@icatt.it (L. Augello); giulia.calvi02@icatt.it
(G. Calvi); annachiara.clementelli01@universitadipavia.it
(A. Clementelli); federica.iurescia@unicatt.it (F. Iurescia);
claudia.corbetta@unibg.it (C. Corbetta)

0000-0001-5100-5539 (F. Iurescia)

© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹https://github.com/UniversalDependencies/UD_Latin-CIRCSE

²The portion of UD_Latin-CIRCSE corresponding to Tacitus' *Germania* is available at https://github.com/CIRCSE/UD_Latin-CIRCSE/blob/main/conllu/03_Tacitus_Germania.conllu

³Details for 2.16 release are available at <https://universaldependencies.org>

⁴<https://universaldependencies.org/u/overview/enhanced-syntax.html>

2.2. Ellipsis

Ellipsis refers to the omission of part of a sentence, indicating an asymmetry between a missing form and its meaning, which remains present even if not overtly expressed [2]. The meaning behind the omission is recoverable through an antecedent [3, p. 14],⁵ which may be either explicitly attested in the text or inferred from world knowledge.

Being a phenomenon that operates at the intersection of different linguistic domains [4, p. 341], ellipsis has been the subject of numerous studies from various perspectives and theoretical frameworks. Concerning syntactic analysis, a substantial body of research has addressed the topic within a constituency-based approach ([5, 6, 7, 8], among others). In contrast, within the dependency framework, the amount of work on ellipsis is considerably smaller ([9, 10, 11], among others). This is mainly due to a fundamental difference: while constituency-based approaches allow for the existence of empty nodes in the syntactic structure, making it feasible to account for ellipsis, dependency-based approaches are less inclined to do so, and therefore tend to dismiss the treatment of ellipsis. This theoretical divergence is also reflected in the representation of ellipsis in treebanks. Ellipsis is explicitly addressed in the Penn Treebank (PTB) [12], which follows a constituency-based approach, as well as in the BulTreeBank [13], which is based on the Head-Driven Phrase Structure Grammar formalism [14]. As for dependency treebanks, the Prague Dependency Treebank (PDT) [15] handles ellipsis in a separate annotation layer,⁶ whereas the Universal Dependencies framework accounts for ellipsis only marginally (see section 2.3 for further insights on ellipsis in UD). In the field of NLP and Large Language Models (LLMs), the challenges LLMs face in processing ellipsis reflect its inherent complexity [17], thereby underscoring the importance of gold-standard data in ellipsis research [18].

2.3. Ellipsis in UD

As mentioned in section 2.2, dependency-based treebanks are generally not inclined to introduce empty nodes into the syntactic structure. Currently, the most widely adopted and state-of-the-art framework for dependency treebanks is UD (see section 2.1), which handles ellipsis differently depending on the level of annotation—basic or enhanced.

Concerning the basic annotation, ellipsis is annotated using two different strategies:⁷ (i) promotion and (ii) the orphan dependency relation (deprel). More specifically, promotion consists in elevating a dependent of the elided element to take on its syntactic role, effectively replacing the omitted node and assuming its function. This strategy follows a defined hierarchy.⁸ By contrast, when promotion—and thus the preservation of the original syntactic function of the omitted element—would result in an unnatural syntactic structure, the orphan dependency relation is used instead. However, this dependency relation inevitably obscures the underlying syntactic structure, thereby entailing a loss of syntactic information.

It is therefore evident that UD does not directly address ellipsis in its basic annotation. Rather, ellipsis is concealed through the use of promotion, which—without targeted analysis—does not explicitly mark its presence, and is further obscured by the application of the orphan relation. While this relation enables ellipsis to be identified explicitly, it nonetheless obscures the syntactic representation of the sentence.⁹

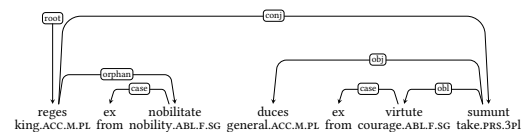
In this work, we focus on examples of ellipsis annotated with the orphan deprel. Example 1 illustrates the basic annotation of such a sentence:

Example 1 – Basic Annotation

Tac., *Germ.* 7,1

reges ex nobilitate duces ex uirtute sumunt

They take their kings on the ground of birth, their generals on the basis of courage¹⁰



In this sentence, predicate ellipsis results in the promotion of the accusative *reges* (“kings”) to the root position, leaving the dependent *ex nobilitate* (“on the ground of birth”) orphaned.

This structure is illustrated in the basic dependency tree, where *reges* governs both *nobilitate* as orphan and *sumunt* (“they take”) as a conjunct (conj).

Basic dependencies fall short of adequately representing implicit structures such as ellipsis. Instead, to thoroughly annotate elliptical constructions, a more suitable strategy within the UD framework is offered by enhanced

⁵In the literature, the term antecedent is typically used in a broad sense, not necessarily referring to an element that precedes the ellipsis. In fact, the element supplying the missing content may also follow the ellipsis, in which case the term postcedent would be more accurate. However, to remain consistent with established usage, we use the term antecedent in both cases.

⁶Refer to Mikulová [16] for further discussion of ellipsis in the PDT.

⁷<https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>

⁸<https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis-in-nominals>

⁹For a proposal on explicitly marking ellipsis with a dedicated subtype, see https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general_meetings:3rd_unidive_general_meeting_23_how_to_ellipsis_a_proposal.pdf

¹⁰Latin translations are drawn by [19].

dependencies relations. At the current state of the art, enhanced dependencies for ellipsis involve the reconstruction of an empty node in the syntactic structure, along with the recovery of the relevant syntactic relations.¹¹ However, since enhanced annotation is "an optional addition to the basic representation",¹² the guidelines for annotating such cases remain underspecified.

3. Ellipsis in Tacitus' *Germania*

Given the current state of enhanced dependency annotation and the importance of gold-standard annotated data, we adopt the approach proposed by [20], which offers a consistent annotation scheme explicitly designed to address predicate ellipsis. The syntactic annotation of ellipsis in Tacitus' *Germania* was performed manually by a team of four annotators equipped with a background knowledge of Latin language and literature and expertise on syntactic annotation in the UD framework.¹³

More specifically, in this work we focus on predicate ellipsis, presenting cases beyond this scope when they are relevant to illustrate annotation-related issues and decisions. We extend the enhanced annotation to cover primarily cases in which the basic annotation layer displays the orphan dependency relation, as well as, where applicable, cases in which predicate ellipsis is conveyed without being explicitly marked by means of orphan. There are 71 tokens out of 5,674 annotated with the *deprel* orphan in the basic layer.¹⁴ They are distributed across 49 sentences out of a total of 299 sentences.¹⁵ No

¹¹For enhanced guidelines see: <https://universaldependencies.org/v2/enhanced.html>

¹²<https://universaldependencies.org/v2/enhanced.html>

¹³The annotators—three students from the graduate program in Linguistic Computing and one in Theoretical and Applied Linguistics—developed their expertise in syntactic annotation in the context of a dedicated training workgroup (approximately 30 hours) held at the CIRCSE research centre at the Università Cattolica del Sacro Cuore, Milano. In a first stage, the annotators practised on basic annotation layer by reannotating portions of gold standard texts from the CIRCSE treebank, both from poetry and prose. In this stage each dataset was annotated by all students. Their Inter Annotator Agreement (IAA), measured using Fleiss' Kappa coefficient, yielded an average score of 0.86 for edges and 0.85 for *deprel*s. (Our script for calculating IAA on CoNLL-U files is available at https://github.com/johnnymoretti/CoNLL-U_Fleiss_Kappa). In a second stage, the annotators focussed on the advanced layer, by annotating ellipsis in Tacitus' *Germania*. Due to time constraints, in this stage each annotator was responsible for the annotation of a single portion of the text. The annotation of the single subsets were subsequently proof-checked by two expert annotators with advanced competence in Latin language and literature and expertise in the syntactic annotation of ellipsis. This process involved collaborative discussions during which the annotators presented the issues they encountered.

¹⁴Two tokens out of 71 are annotated with *orphan:missing* due to a lacuna in the *Opera Latina* corpus [21], which is the source text of the treebank.

¹⁵Two sentences (Tac., *Germ.* 2,3 and 37,4) stand out, as they present

specific pattern emerges regarding the distribution of elliptical constructions across the text.

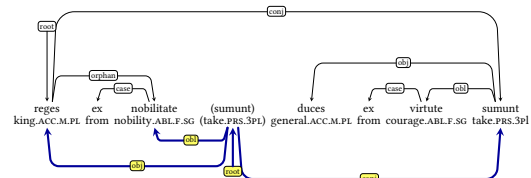
In what follows, we report a few examples of reconstruction of predicate ellipsis (subsection 3.1), of the treatment of orphan *deprel* in cases of nominal ellipsis (subsection 3.2) and of ellipsis in copular constructions (subsection 3.3). We then illustrate cases of ellipsis in predicative (subsection 3.4) and comparative structures (subsection 3.5), providing examples of how we decide whether or not to intervene. Finally, we present an example of nested ellipsis (subsection 3.6).

3.1. Gapping

The main instance of predicate ellipsis found in the *Germania* is gapping. Gapping is a type of ellipsis in which a repeated verb is omitted from a coordinated clause, leaving behind only the contrasting elements [22]. Example 1a illustrates the enhanced annotation of the sentence presented in example 1:¹⁶

Example 1a – Enhanced Annotation

reges ex nobilitate (sumunt) duces ex uirtute sumunt



In the enhanced UD annotation, the elided verb *sumunt* is introduced as an empty node, explicitly marking the omitted predicate. Consequently, in the enhanced tree graph representation, the empty node becomes the root, governing *reges* as its direct object (*obj*) and *nobilitate* as an oblique dependent (*obl*). Finally, the empty node governs *sumunt* as *conj*.

3.2. Nominal Ellipsis

Example 2 illustrates an example of nominal ellipsis marked with the *deprel* orphan in the basic layer. The enhanced tree of this and of the following examples are detailed in the appendix A.

Example 2

Tac., *Germ.* 11,2

nec dierum numerum ut nos sed noctium computant

They count not by days as we do, but by nights

4 and 6 tokens annotated as orphan, respectively, whereas the average distribution of orphan equals to 1.44 tokens per sentence.

¹⁶For gapping in Latin, see, e.g., [23, p. 585-586; 610-611].

This sentence illustrates a simple case of nominal ellipsis,¹⁷ where the elided element is the noun *numerus* (“number”) before *noctium* (“of nights”). In the basic UD dependency tree, the ellipsis is captured by the presence of the orphan relation assigned to the dependent *noctium*, directly governed by the main verb *computant* (“they count”).

In the enhanced dependency graph, the elided noun *numerus* is introduced as an empty node to resolve the syntactic discontinuity. Thus, *noctium* is annotated as a dependent of the reconstructed empty node *numerus* via the *nmod* relation. This structure reproduces the expected syntactic structure and mirrors the dependencies found in the first clause *nec dierum numerum*, where the root *numerus* governs *dierum* as its nominal modifier *nmod*. Example 3 is an instance of nominal ellipsis where a missing antecedent led to creation of an empty node, which could be semantically inferred, but still does not have an actual linguistic counterpart in the phrase:

Example 3

Tac., *Germ.* 15,1

mos est ciuitatibus ultro ac uiritim conferre principibus uel armentorum uel frugum quod pro honore acceptum etiam necessitatibus subuenit

It is the custom in their states to bestow upon the chief unasked and man by man some portion of one's cattle or crops: it is accepted as a compliment, but also serves his needs

In this sentence, the verb *conferre* (“to bestow”) governs an argument in dative (*principibus*, “upon the chiefs”) and semantically requires a direct object representing what is being bestowed. The coordinated genitives *armentorum* (“of cattle”) and *frugum* (“of crops”) function as modifiers of the verb, but not as arguments. They would imply a partitive relationship, and require an implicit head noun (e.g., *pars*, “a portion”) to form a semantically complete direct object depending on *conferre*.

In the basic UD representation, this ellipsis is captured through the presence of an orphan, while *frugum* is a conjunct. The orphan label here marks the gap in the basic annotation: the genitive lacks a head noun, and the sentence lacks a direct object for *conferre*. This missing noun is both semantically necessary and syntactically expected, and this motivates its reconstruction. So, in the enhanced annotation, we reconstruct the argument expectations of *conferre* and introduce: a new empty nominal node functioning as the direct object of *conferre*; *armentorum* as *nmod* of the nominal empty node and *frugum* as its conjunct; the elimination of the orphan relation.

Crucially, we add an empty node without storing any lexical content, as there is no explicit antecedent to make

reference to. Unlike more typical cases of predicate or comparative ellipsis (where the verb or noun is missing but recoverable from a parallel structure) here, no noun referring to e.g., “portion” appears in the clause, distinguishing this case from anaphoric ellipsis. Then, in the absence of an antecedent, the reconstructed node in our enhanced annotation is just empty, and not a copy of something else. So, the new reshaped phrase would appear like this: *conferre principibus uel armentorum uel frugum* _ , with the empty node occupying the last position.¹⁸

Example 4 features an instance of ellipsis of the subject which lacks an overt antecedent, requiring the reconstruction of an empty node not filled with any lexical or morphological information.

Example 4

Tac., *Germ.* 26,1

faenus agitare et in usuras extendere ignotum ideoque magis seruatur quam si uetustum esset

To exploit capital and to increase it by interest are unknown, and the principle is accordingly better observed than if there had been actual prohibition

The first clause describes the unfamiliarity of usury among the Germans. The second clause has the verb *seruatur* (“observed”) as head: its implicit subject is understood as the negation of the previous clause, *faenus non agitare neque in usuras extendere* (“not to exploit capital and not to increase it”), whose negative meaning has to be inferred from *ignotum* (“are unknown”).

Following the interpretation adopted by [19] in their translation, we reconstruct a single empty node for the omitted subject, representing this implied negative concept that may be paraphrased as a generic “principle”. Hence, the reconstructed node is assigned the dependency relation *nsubj* : *pass*, under a nominal reading of the elided material. For reasons of accuracy, no lexical or morphological features are encoded, as no explicit antecedent is present in the text.

Example 3 and 4 serve to exemplify the procedure adopted in analogous cases: when no explicit antecedent is present in the text, the reconstructed node is left lexically and morphologically underspecified, and only the appropriate dependency relation is assigned. The absence of a textual antecedent is indicated in the MISC field of the CoNLL-U file by marking the source of interpretation as world knowledge (wk).

3.3. Copular Constructions

Example 5 illustrates a case of ellipsis of a copular construction. Such cases concern the omission of a predicate

¹⁷For nominal ellipsis in Latin, see, e.g., [24, p. 962].

¹⁸For a tentative reconstruction of the elided part, see, e.g., [25, p. 41], who posits ellipsis of an indefinite pronoun, as head of *uel armentorum uel frugum*.

formed by the nominal component and a form of the verb *sum*.

Example 5

Tac., *Germ.* 27,2

feminis lugere honestum est uiris meminisse

Lamentation becomes women: men must remember

This sentence illustrates a case of nominal predicate ellipsis, involving the omission of both the copula *est* and the nominal component *honestum* (“convenient”). The structure is parallel, contrasting actions appropriate for women and men in mourning. In the first clause, the argument *feminis* (“for women”) precedes the clausal subject *lugere* (“to mourn”) of the nominal predicate; in the second, *uiris* (“for men”) is followed by the infinitive *meminisse* (“to remember”), on which it depends, in basic annotation, via an orphan relation that marks the ellipsis of the predicate.

Hence, in the enhanced annotation, the ellipsis is resolved through the insertion of an empty node at the end of the second clause, mirroring the structure and the dependency relations of the first. In accordance with the proposal in [20], we reconstruct only the nominal part of the predicate, the head *honestum*, which carries the semantic content of the predicate and ensures cross-linguistic consistency.

3.4. Predicative Constructions

As mentioned in section 2.3, ellipsis is annotated only when it creates syntactic discontinuities, specifically, when the absence of a word leaves dependents orphaned. This is most visible in constructions like gapping (section 3.1), while in other structures, an elided predicate does not result in unattached dependents or broken syntactic structure, and therefore there are cases of elliptical constructions which are not annotated as orphan in UD (section 2.3). Example 6 is a case of predicative constructions involving coordinated or juxtaposed clauses:

Example 6

Tac., *Germ.* 13,1

... scuto frameaque iuuenem ornant ... ante hoc domus pars uidentur, mox rei publicae

... [they] equip the young man with shield and spear ... up to this point they seem a part of the household, next a part of the state

We focus on the final segment of the sentence: *ante hoc domus pars uidentur, mox rei publicae* (“up to this point they seem a part of the household, next a part of the state”). The first clause (*domus pars uidentur*) is a standard predicative construction consisting of a subject (understood as *illi* and referring to *iuuenem*), the verb *uidentur* (“they seem”), and the predicative nominal *domus pars*. In the second clause (*mox rei publicae*), both

the verb *uidentur* and the predicative noun *pars* are absent. Nevertheless, they can be clearly inferred and the intended structure formed by the two clauses is parallel: *domus pars uidentur* (“they seem a part of the household”) *rei publicae pars uidentur* (“they seem a part of the state”).

However, despite the interpretative clarity, as both the verbal head and its nominal predicate are missing, this is a clear case of predicate ellipsis that is not marked in the basic UD annotation, since no dependent is left syntactically orphaned.

So, in the enhanced UD annotation, we insert two reconstructed nodes: the verb *uidentur* as the verbal head and the predicative nominal *pars*, which has *rei publicae* as a dependent. Therefore, in the enhanced tree graph representation the structure would include: a new verbal node *uidentur* connected to the first verb with the *deprel conj*; a new nominal node *pars* connected as *xcomp* to the reconstructed *uidentur*; the nominal phrase *rei publicae* dependent of *pars* as *nmod*.

3.5. Ellipsis in Comparative Constructions

Another significant case of ellipsis in the *Germania* occurs in comparative clauses.¹⁹ The following case is one of this kind:

Example 7

Tac., *Germ.* 14,3

nec arare terram aut expectare annum tam facile persuaseris quam uocare hostem et uulnera mereri pigrum quin immo et iners uidetur sudore acquirere quod possis sanguine parare

You will not so readily persuade them to plough the land and wait for the year’s returns as to challenge the enemy and earn wounds: besides, it seems limp and slack to get with the sweating of your brow what you can gain with the shedding of your blood

In this sentence, the verb *persuaseris* (“you will persuade”) governs the first clause of the comparative construction (with the two infinitive verbs *arare*, “to plough”, and *expectare*, “to wait”) but is not repeated in the second (verbs *uocare*, “to challenge”, and *mereri*, “to earn”), despite clearly being the intended meaning. This omission might seem to be a candidate for ellipsis annotation. However, UD guidelines do not annotate such cases of comparative ellipsis for several theoretical and practical reasons, and we decide to do the same.

There is no orphan relation in the basic tree, so there is no ellipsis to resolve in the enhanced representation either. We see that there are no orphaned dependents in the second clause, as the infinitive verb *uocare* depends on *persuaseris* as *advcl : cmp* and *mereri* is its conjunct. Since all constituents are attached with appropriate

¹⁹For ellipsis in comparative construction in Latin, see, e.g., [23, p. 721].

heads, no orphan relation is needed. Being the sharing of the predicate a common phenomenon in comparative constructions, especially in Latin, we accept this syntactically and semantically recoverable pattern, and choose not to annotate it. So, even by doing nothing, here the clause remains structurally intact, and no dependents are left without a head, although *persuaseris* is intuitively present in both parts of the comparative construction.

3.6. Nested Ellipsis

Among the knotty sentences to annotate, example 8 stands out: exemplifying Tacitus' concise and condensed prose, it features a case of a nested ellipsis.

Example 8

Tac., *Germ.* 43,2

e quibus Marsigni et Buri sermone cultuque Suebos referunt Cotinos Gallica Osos Pannonica lingua coarguit non esse Germanos et quod tributa patiuntur

Among them the Marsigni and Buri in language and mode of life recall the Suebi: as for the Cotini and Osi, the Gallic tongue of the first and the Pannonian of the second prove them not to be Germans; so does their submission to tribute

The two clauses *Cotinos Gallica* and *Osos Pannonica lingua coarguit* are in asyndeton, with no subordinating connective present: indeed, in accordance with the guidelines, their relationship has been annotated with the label *conj.* The clause *Cotinos Gallica* shows a predicate ellipsis (*coarguit*) and instantiates an example of gapping (section 3.1): the use of a singular verb would otherwise be inexplicable.

What makes the situation more complex is that the reconstructed *coarguit* implies, even in this first clause, *non esse Germanos*. Therefore, another empty node (*Germanos*) has been reconstructed after *Cotinos*. In the enhanced UD annotation, the reconstructed *coarguit* thus takes on the role that *Gallica* played in the previous annotation; *Cotinos*, on the other hand, becomes the *nsubj* of the reconstructed *ccomp Germanos*.²⁰

4. Conclusion

This work describes the challenges encountered during manual annotation of ellipsis in Tacitus' *Germania*. The treebank enhanced with this annotation will be included in the next UD release. Building on the proposal outlined in [20], it provides examples of ellipsis in Latin, thereby

offering material for comparison on the treatment of ellipsis across languages. Unlike the approach taken in [20], where the reconstruction focussed exclusively on predicate ellipsis and did not attempt to recover omitted arguments, thus deliberately excluding cases of nominal ellipsis, the present study has highlighted the need to consider nominal ellipsis as well, as illustrated in Section 3.2. This opens the possibility of expanding the domain of reconstruction to include a broader range of omitted elements. Accordingly, this work offers insights into additional aspects that should be considered in the development of guidelines for ellipsis annotation within the UD framework, which remain currently underspecified.

For this work, we focussed on the description of the challenges and identified some directions for future works. First, a thorough examination of the position of the elided material shall pave the way for a study on the communicative reasons that may have guided the author in choosing an elliptical structure, such as topical focussing, among others. More research is needed to explore patterns of usage of elliptical constructions in the *Germania* and, more in general, in Tacitus' oeuvre. Second, a classification of the types of ellipsis encountered in Tacitus' *Germania* shall contribute to the ongoing discussion on ellipsis. The addition of other (Latin) treebanks enhanced with annotation focussing on ellipsis remains a desideratum, which will foster both research focussing on linguistic aspects of ellipsis and on stylistics. Such additions would, among other benefits, enable the training of NLP tools for ellipsis detection, thereby facilitating large-scale research into its frequency patterns and distribution across treebanks representing different genres and, for literary texts, across different authors.

References

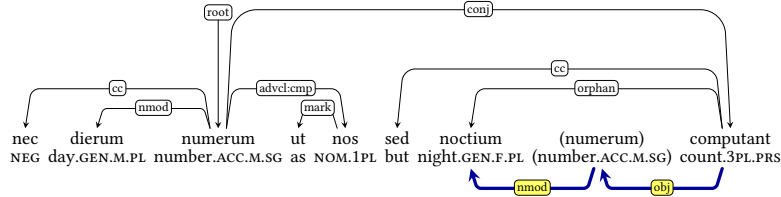
- [1] M.-C. De Marneffe, C. D. Manning, J. Nivre, D. Zeman, Universal dependencies, *Computational linguistics* 47 (2021) 255–308.
- [2] J. V. Craenenbroeck, T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis*, Oxford Handbooks, Oxford University Press, Oxford, 2018.
- [3] M. J. McShane, *A theory of ellipsis*, Oxford University Press, 2005.
- [4] A. Kehler, Ellipsis and discourse, in: J. van Craenenbroeck, T. Temmerman (Eds.), *The Oxford Handbook of Ellipsis*, Oxford Handbooks, Oxford University Press, 2018. doi:10.1093/oxfordhb/9780198712398.013.13, online edition, Oxford Academic, 8 Jan. 2019. Accessed 6 June 2025.
- [5] J. R. Ross, Guess who?, in: *Proceedings from the annual meeting of the Chicago linguistic society*, volume 5, Chicago Linguistic Society, 1969, pp. 252–286.

²⁰It should also be noted that this is a case of promotion: in the first clause *Gallica* is an adjective functioning as the subject, due to the ellipsis of *lingua*. In line with our annotation criteria, we do not reconstruct this ellipsis, since in the basic annotation the relation is not marked as orphan.

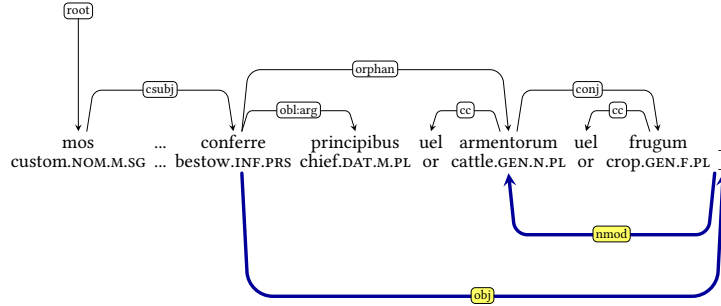
- [6] J. Merchant, *The syntax of silence: Sluicing, islands, and identity in ellipsis*, University of California, Santa Cruz, 1999.
- [7] C. Kennedy, *Ellipsis and syntactic representation*, in: *The interfaces: Deriving and interpreting omitted structures*, John Benjamins Publishing Company, 2003, pp. 29–53.
- [8] I. Ortega-Santos, M. Yoshida, C. Nakao, *On ellipsis structures involving a wh-remnant and a non-wh-remnant simultaneously*, *Lingua* (2014).
- [9] I. A. Mel'čuk, *Dependency Syntax: Theory and Practice*, SUNY Press, Albany, NY, 1988.
- [10] R. Hudson, *An introduction to word grammar*, Cambridge University Press, 2010.
- [11] T. Osborne, *Ellipsis*, in: *A Dependency Grammar of English*, John Benjamins Publishing Company, 2019, pp. 349–378.
- [12] M. Marcus, B. Santorini, M. A. Marcinkiewicz, *Building a large annotated corpus of english: The penn treebank*, *Computational linguistics* 19 (1993) 313–330.
- [13] P. Osenova, K. Simov, *The bulgarian hpsg treebank: Specialization of the annotation scheme*, in: *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*; Växjö, Sweden, 2003.
- [14] C. Pollard, I. A. Sag, *Head-driven phrase structure grammar*, University of Chicago Press, 1994.
- [15] J. Hajič, E. Hajičová, M. Mikulová, J. Mirovský, *Handbook on linguistic annotation, chapter prague dependency treebank*, 2017.
- [16] M. Mikulová, *Semantic representation of ellipsis in the prague dependency treebanks*, in: *Proceedings of the 26th Conference on Computational Linguistics and Speech Processing (ROCLING 2014)*, 2014, pp. 125–138.
- [17] D. Čavar, Z. Tiganj, L. V. Mompelat, B. Dickson, *Computing ellipsis constructions: Comparing classical nlp and llm approaches*, in: *Proceedings of the Society for Computation in Linguistics 2024*, 2024, pp. 217–226.
- [18] D. Čavar, L. Mompelat, M. Abdo, *The typology of ellipsis: a corpus for linguistic analysis and machine learning applications*, in: *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 2024, pp. 46–54.
- [19] M. Hutton, E. H. Warmington, *Agricola ; Germania ; Dialogus*, Loeb classical library 62, rev. ed. ed., Harvard Univ. Press, Cambridge Mass, 1970.
- [20] C. Corbetta, F. Iurescia, M. C. Passarotti, «are you afraid of ghosts?» a proposal for busting predicate ellipsis in Universal Dependencies, in: S. Jablotschkin, S. Kübler, H. Zinsmeister (Eds.), *Proceedings of the 23rd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2025)*, Association for Computational Linguistics, Ljubljana, Slovenia, 2025, pp. 52–63. URL: <https://aclanthology.org/2025.tlt-1.6/>.
- [21] J. Denooz, *Opera Latina : une base de données sur internet*, *Euphrosyne* 32 (2004) 79–88. URL: <https://www.brepolonline.net/doi/10.1484/J.EUPHR.5.125535>. doi:10.1484/J.EUPHR.5.125535.
- [22] K. Johnson, *Gapping*, in: M. Everaert, H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax*, volume II, Blackwell, 2006, pp. 407–435.
- [23] H. Pinkster, *The Oxford Latin Syntax*, volume 2, Oxford University Press, Oxford, UK, 2021.
- [24] H. Pinkster, *The Oxford Latin Syntax*, volume 1, Oxford University Press, Oxford, UK, 2015.
- [25] U. Zernial, *Germania, Sammlung griechischer und lateinischer Schriftsteller mit deutschen Anmerkungen*, 2. verbesserte aufl. ed., Weidmannsche Buchhandlung, Berlin, 1897.

A. Appendix

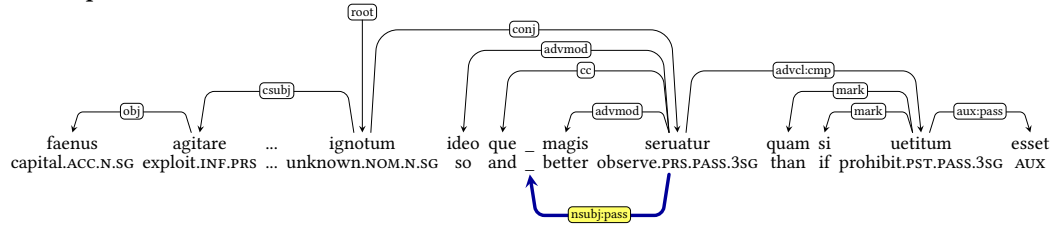
Example 2.1: Enhanced Tree



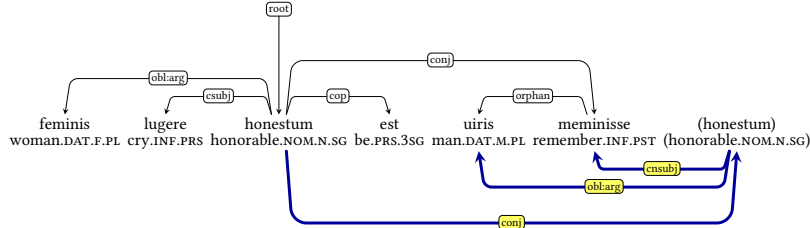
Example 3.1: Enhanced Tree



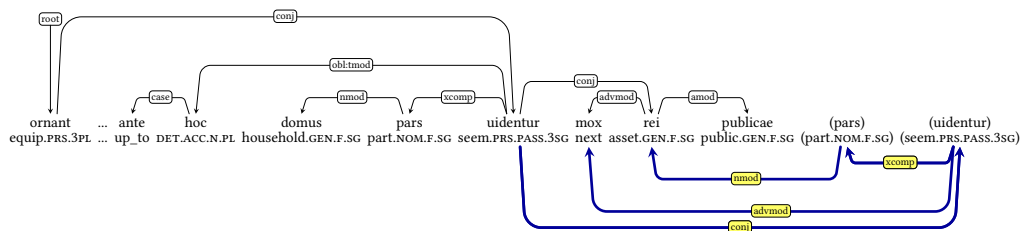
Example 4.1: Enhanced Tree



Example 5.1: Enhanced Tree



Example 6.1: Enhanced Tree



Example 8.1: Enhanced Tree

