



UNIVERSITÀ  
DEGLI STUDI  
DI BERGAMO



UNIVERSITÀ DEGLI STUDI DI BERGAMO  
UNIVERSITÀ DEGLI STUDI DI PAVIA

Scuola di Alta Formazione Dottorale  
Dipartimento di Lingue, Letterature e Culture Straniere  
Corso di Dottorato in Scienze Linguistiche  
Ciclo XXXVIII

**THE SOUND OF SILENCE IN THE *DIVINE COMEDY*:**  
**A DATA-DRIVEN STUDY OF PREDICATE ELLIPSIS AND  
OBJECT OMISSION**

Supervisore:

Chiar.ma Prof.ssa Silvia Luraghi

Co-Supervisore:

Chiar.mo Prof. Marco Passarotti

Tesi di dottorato

Claudia CORBETTA

Matricola n. 1056923

ANNO ACCADEMICO 2024-2025

## Table of contents

<b>Introduction .....</b>	<b>7</b>
<b>1. General overview .....</b>	<b>9</b>
<b>1.1 The <i>Divine Comedy</i> .....</b>	<b>9</b>
1.1.1 DanteSearch .....	11
<b>1.2 Old Italian and its digital resources .....</b>	<b>12</b>
<b>1.3 Dependency grammar .....</b>	<b>14</b>
1.3.1 Dependency syntax .....	15
1.3.2 Dependency versus Constituency Grammar and the notion of treebanks .....	16
1.3.3 The impact of dependency grammar in NLP: Universal Dependencies .....	18
<b>1.4 Universal Dependencies .....</b>	<b>18</b>
1.4.1 History of UD .....	19
1.4.2 Introduction to UD .....	19
1.4.3 Dependency relations in UD .....	22
1.4.4 Enhanced Dependencies .....	26
1.4.5 Italian treebanks in UD .....	27
1.4.6 The CoNLL-U format .....	28
<b>2. Building Italian-Old treebank .....</b>	<b>31</b>
<b>2.1 Annotating the <i>Divine Comedy</i> .....</b>	<b>31</b>
2.1.1 Old Italian in Universal Dependencies and the ISO code .....	33
2.1.2 Conversion .....	34
2.1.2.1 DanteSearch tagset .....	35
2.1.2.2 Conversion process .....	36
2.1.2.2.1 Tokenization, sentence splitting and lemmatization .....	36
2.1.2.2.2 Morphological annotation .....	37
2.1.2.2.3 Syntactic annotation .....	39
2.1.2.2.4 Miscellaneum .....	40
2.1.3 Syntactic annotation process .....	40
2.1.3.1 Testing inside the Inferno: road to Purgatorio .....	41
2.1.3.2 Testing with different models: road to Paradiso .....	42
<b>2.2 Annotation issues .....</b>	<b>45</b>
2.2.1 Punctuation marks: the semicolon .....	46
2.2.2 Comparative clauses .....	49
<b>2.3 Possible future works .....</b>	<b>52</b>
<b>3) Ellipsis .....</b>	<b>55</b>
<b>3.1 Key concepts .....</b>	<b>57</b>
3.1.1 Ellipsis site .....	57
3.1.2 Antecedent .....	58
3.1.3 Remnant .....	59
3.1.4 Identity condition .....	60
3.1.5 Structure and (non-)structural approaches .....	62
3.1.5.1 Structural approaches .....	63
3.1.5.2 Non-structural approaches .....	64
3.1.6 Discourse approach .....	64
<b>3.2 Classification of ellipsis .....</b>	<b>65</b>
3.2.1 Predicate ellipsis .....	66
3.2.1.1 Gapping .....	67
3.2.1.2 Pseudo-Gapping .....	68
3.2.1.3 VP-ellipsis .....	68

3.2.1.4 Other predicate ellipsis .....	69
3.2.2 Nominal ellipsis .....	70
3.2.2.1 Subject ellipsis .....	71
3.2.2.2 Object ellipsis.....	71
3.2.3 Clausal ellipsis .....	76
3.2.3.1 Stripping/bare argument ellipsis.....	76
3.2.3.2 Fragments.....	77
3.2.3.3 Sluicing .....	77
3.2.3.4 Ellipsis in comparatives and Comparative deletion.....	78
3.2.3.4.1 Comparative deletion.....	79
<b>3.3 Ellipsis in Diachrony .....</b>	<b>81</b>
3.3.1 Ellipsis in Old Italian .....	82
3.3.1.1 Verbal ellipsis in Old Italian .....	85
3.3.1.2 Object ellipsis in Old Italian.....	87
<b>3.4 Ellipsis and Syntax .....</b>	<b>88</b>
3.4.1 Ellipsis and Dependency Syntax .....	88
3.4.2 Ellipsis in Treebanks.....	90
3.4.3 Ellipsis in Universal Dependencies.....	92
3.4.3.1 Promotion.....	93
3.4.3.2 Orphan.....	95
3.4.2.3 Enhanced Dependency and empty nodes .....	96
<b>3.5 Summary .....</b>	<b>97</b>
<b>4) Ghostbusting Ellipsis.....</b>	<b>99</b>
<b>4.1 How to ghostbust ellipsis: a theoretical method .....</b>	<b>100</b>
4.1.1 Detection.....	101
4.1.2 Identification.....	102
4.1.3 Reconstruction .....	103
4.1.5 Summary.....	106
<b>4.2 How to ghostbust ellipsis: a practical guide .....</b>	<b>107</b>
4.2.1 Detection.....	109
4.2.1.1 Extraction of predicate ellipsis .....	109
4.2.1.1.1 Extraction of orphan .....	112
4.2.1.1.2 Extraction of promoted elements.....	114
4.2.1.1.2.1 Theoretical workarounds for promotion .....	116
4.2.1.1.2.2 Queries for promotion.....	119
4.2.1.2 Extraction of object ellipsis.....	123
4.2.2 Identification.....	125
4.2.2.1 Predicate ellipsis.....	126
4.2.2.2 Object ellipsis.....	129
4.2.3 Reconstruction .....	131
4.2.3.1 Verbal predicate ellipsis .....	134
<b>4.3 Summary .....</b>	<b>141</b>
<b>5) Analysis: predicate ellipses and object omission.....</b>	<b>142</b>
<b>5.1 Predicate ellipses .....</b>	<b>143</b>
5.1.1 General overview .....	144
5.1.2 Syntactic context of the predicate ellipsis .....	147
5.1.2.1 Quantifying the Impact of Predicate Ellipsis in and across the Comedy.....	150
5.1.3 Classification of predicate ellipsis type.....	155
5.1.3.1 Gapping.....	155
5.1.3.1.1 Forward Gapping: .....	158
5.1.3.1.2 Backward Gapping .....	161
5.1.3.1.3 Bi-directional gapping and nested gapping.....	164
5.1.3.2 Elliptical narrative framing of reported speech .....	165

5.1.3.3 Exclamation.....	167
5.1.3.4 Foreign language.....	169
5.1.3.5 Presentative and locative clauses.....	170
5.1.3.6 Listing and stylistic repetition.....	171
5.1.3.7 Truncations e fragments.....	173
5.1.3.8 Sluicing.....	175
5.1.3.9 Stripping/bare argument ellipsis.....	175
5.1.3.9.1 (Positive) Stripping.....	176
5.1.3.9.2 Not-Stripping.....	177
5.1.3.9.2.1 Not-Stripping with antecedent in affirmative clause.....	177
5.1.3.9.2.2 Not-Stripping with antecedent in negative clause.....	178
5.1.3.9.2.3 Not-Stripping with postcedent.....	179
5.1.3.9.3 Nested (Not-) Stripping.....	179
5.1.3.10 Ellipsis of the copular essere (to be).....	181
5.1.3.11 Comparative clause.....	182
5.1.3.11.1 Comparative deletion.....	183
5.1.3.11.2 Comparative ellipsis.....	183
5.1.3.11.2.1 Comparative ellipsis with more than one remnant.....	184
5.1.3.11.2.2 Comparative ellipsis with a single remnant.....	186
5.1.4 Merging qualitative and quantitative analysis: contextualizing predicate ellipses.....	188
5.1.5 What should be considered as predicate ellipsis?.....	194
<b>5.2 Object omission.....</b>	<b>196</b>
5.2.1 Generic omitted objects.....	198
5.2.2 Specific omitted objects without textual referent.....	199
5.2.3 Specific omitted objects with referent.....	204
5.2.3.1 The deictic “ <i>così</i> ”.....	204
5.2.3.2 Comparative structures.....	206
5.2.3.3 Reduplication.....	208
5.2.3.4 Shared dependent.....	211
5.2.3.4.1 Shared dependent with postcedent.....	212
5.2.3.4.2 Shared dependent with antecedent.....	214
5.2.3.4.3 Shared dependent in Modern Italian.....	216
5.2.3.5 Lexical properties.....	219
5.2.3.5.1 Narrative/diegetic function.....	222
5.2.3.7 Non-sentential antecedent.....	224
5.2.4 Non canonical pattern.....	225
5.2.4.1 Indirect object.....	226
5.2.4.2 Ambiguous valency.....	228
5.2.4.3 Annotation issues.....	229
5.2.4.4 Contextual meaning and intransitive use.....	230
5.2.4.5 Divergences in Textual Reconstruction.....	233
5.2.4.6 Use of the object pronoun as expletive:.....	234
5.2.5 Omission of the reflexive clitic: neuter and uso assoluto.....	235
5.2.6 Summary.....	238
<b>6) Summary and conclusion.....</b>	<b>241</b>
<b>References.....</b>	<b>246</b>
<b>Appendix.....</b>	<b>265</b>



*ché nullo effetto mai ragionabile,  
per lo piacere uman che rinovella  
seguendo il cielo, sempre fu durabile.*

*Opera naturale è ch'uom favella;  
ma così o così, natura lascia  
poi fare a voi secondo che v'abbella.*

‘for never  
has any thing produced by human reason  
been everlasting—following the heavens,  
men seek the new, they shift their predilections.

That man should speak at all is nature’s act,  
but how you speak—in this tongue or in that—  
she leaves to you and to your preference.’

*Par. XXVI, vv. 127-132*

## Introduction

*Ma chi pensasse il ponderoso tema  
e l'omero mortal che se ne carica,  
nol biasmerebbe se sott'esso trema:*

‘But he who thinks upon the weighty theme,  
and on the mortal shoulder bearing it,  
will lay no blame if, burdened so, I tremble:’

*Par. XXIII, vv. 64-69*

This study represents a three-year journey which, much like the text at its core, has sought to navigate the vast waters of three kingdoms whose boundaries are less clear-cut than they might initially appear. The present work aspires to have captured their essential aspects, while remaining aware of the breadth and complexity inherent to each of them. The first kingdom is that of Italian Studies, and more specifically the language of Dante Alighieri’s *Divina Commedia* (Section 1.2). Its exploration has led to a detailed investigation of syntax through the construction of a dependency treebank (Chapter 2) in accordance with the Universal Dependencies formalism (Section 1.4), currently regarded as the state of the art in syntactically annotated dependency resources. The second kingdom is that of Computational Linguistics and Natural Language Processing (henceforth NLP), into which Dante’s text has been carried and introduced, with the hope that many other authors and texts may follow the same path. The third kingdom is that of Theoretical and Applied Linguistics, which focuses on the analysis of a specific linguistic phenomenon, ellipsis, with the endeavor to record the ‘sound of silence’ in the *Divine Comedy*. Particular attention is devoted to predicate ellipsis and object omission, for which a data-driven classification and analysis have been provided, together with an annotation proposal aligned with practices in the NLP domain.

The cross-fertilization among these domains arises from a convergence of research interests and opportunities that have shaped this study. Although apparently multi-thematic, the work aims to foster dialogue across three approaches and to offer perspectives that go beyond those of a strictly sector-specific investigation. A central objective is to build bridges between disciplines that are more closely related than might initially be assumed. Computational Linguistics and NLP tools have become indispensable for linguistic and textual analysis, supporting rather than replacing the work of the human researcher. The development of open and accessible resources, in line with the FAIR principles, enables both the texts under examination and their analyses to interact more effectively

with the scientific community, promoting the establishment of a shared and verifiable framework. This dialogue also requires confronting challenges and limitations, which, as with any attempt at text interpretation and analysis, inevitably arise. At the same time, theoretical and applied linguistics provide essential analytical perspectives, deepening linguistic awareness and enriching interpretation by uncovering the mechanisms that sustain textual meaning. Finally, the decision to focus on an historical text, applying linguistic methods and analyses to a historical language and a canonical literary work, stems not only from personal research interest but also from the potential to introduce a diachronic depth that brings the past into productive dialogue with the present.

The study is organized as follows. Chapter 1 briefly introduces the text under examination, the *Divine Comedy*, situating it within the panorama of Old Italian<sup>1</sup> and of the existing resources for Old Italian, and provides a short overview of dependency syntax and the framework adopted. Chapter 2 describes the construction of the treebank *Italian-Old* and the experiments conducted both on the text itself and on related authors. Chapter 3 presents a literature review on ellipsis, with an attempt to bring together studies that address the phenomenon from different theoretical perspectives. Chapter 4 focuses on the treatment of ellipsis in NLP, discussing the challenges of extracting and analyzing a phenomenon that is transparent and absent from the initial data. Chapter 5 is devoted to the analysis, in which classifications are proposed for predicate ellipsis and for object omission. Chapter 6 draws the conclusions. The Appendix and References follow.

---

<sup>1</sup> A definition of what is meant by *Old Italian* is given in Section 1.2.

## 1. General overview

This chapter serves as an introduction to the topics addressed throughout the present study. The various sections do not aim to be exhaustive, given the breadth of the subjects and the richness of the related literature, for which references are provided in the respective sections, but rather seek to distill their essence and highlight the underlying thread that connects them.

Specifically, Section 1.1 introduces the reference text under study, the *Divine Comedy*; Section 1.2 discusses Old Italian more generally; and Section 1.3 offers an overview of dependency syntax and describes the framework within which the present work is situated, namely Universal Dependencies.

### 1.1 The *Divine Comedy*

It would be both unnecessary, given the vast amount of descriptive scholarship on this text<sup>2</sup>, and inconsistent with the aims of the present work to attempt either a detailed account or even an exhaustive bibliography of Dante's *Divine Comedy*. It will suffice here to recall that the work, a poem in Old Italian, more specifically Old Florentine, written around the 1306 and 1321 by Dante Alighieri, is regarded as one of the milestones of Italian literature and language. The *Divine Comedy* is a rhymed poetic text, an allegorical–didactic poem composed in chained tercets of hendecasyllabic verse. However, despite its well-defined metrical structure, it should be emphasized that such formal organization does not limit the author's creativity or poetic skill<sup>3</sup>. The poem is structured into three *Cantiche* narrating Dante's journey through the afterlife: the *Inferno*, consisting of 34 *Canti*; the *Purgatorio*, with 33 *Canti*; and the *Paradiso*, also with 33 *Canti*, for a total of 100 *Canti*. Dante's journey takes shape as a cathartic voyage in which readers may identify themselves, establishing a parallel between Dante as a character<sup>4</sup> and the reader, both the contemporary reader of Dante's time and the modern one, who is directly engaged from the very opening of the poem using the term *nostra vita* 'our life' (*Inf.* I, v. 1). The ability to narrate the human condition and the relevance of human events across time ensures that, despite the chronological distance, this text remains strikingly

---

<sup>2</sup>Among the introductory works on the *Divine Comedy*, the following may be cited as illustrative examples: Inglese (2012); Bellomo (2020: 213–315). For critical interpretation, see Auerbach (2005), Contini (2001), Nardi (1983). Reference is also made to the bibliography included in the recent edition of the *Inferno* edited by Malato (2022).

<sup>3</sup>“Ammettiamo, come per Dante è giocoforza, che il poeta non fosse passivamente soggetto alle necessità metriche, ma capace di risolverle nel senso delle sue preferenze tra forme varie per età e qualità, a seconda della particolare situazione poetica” (Nencioni, 1989a: 27). ‘Let us assume, as it is necessary to do in Dante's case, that the poet was not passively subject to metrical necessities, but rather capable of resolving them in accordance with his preferences among forms of differing age and quality, depending on the specific poetic situation.’

<sup>4</sup>For an analysis of Dante as both character and poet of the poem, see Contini (2001: 33–62).

contemporary and continues to be the subject of study not only among specialists but also by a broader audience of interested readers.

As briefly noted above, the significance of the *Divine Comedy* derives not only from its literary prestige but also from its linguistic dimension. Dante's decision to compose in the *locutio vulgaris* 'the vulgar tongue', based on the Florentine vernacular<sup>5</sup>, and to employ a language *in qua et muliercule cominiant* (*Epistola XIII*, 31; Baglio 2016: 364-367) 'in which also women communicate'<sup>6</sup>, thus emphasizing its status as a language of everyday use, while at the same time breaking with the tradition of Latin as the language of epic poetry (Manni 2013: 85-94), is of fundamental importance. The marked turn toward this *locutio vulgaris* and the expressive power it displays through Dante's mastery highlight a decisive shift toward the use of a language, which over the centuries, also thanks to the contributions of other prominent figures<sup>7</sup>, gradually established itself as the dominant language, Italian. Consequently, in the history of the Italian language, the importance attained by the *Comedy*, which from the fourteenth century onwards became a foundational text and one of the most extensively commented works (Migliorini 2001: 180), is so decisive that Dante has often been regarded as the "father of the Italian language" (Migliorini 2001: 167).

Given the importance of the poem and the wide readership it enjoyed from the very beginning, it is not surprising that a considerable number of copies<sup>8</sup> of the work, both manuscripts and later printed editions, together with commentaries on the text, have been transmitted. It should be recalled, however, that no autograph of the *Commedia* has been preserved, that is, no manuscript in the author's own hand (Bellomo 2020: 289), and the text as we read it today is therefore the result of a collation of the various witnesses attesting to the work. Nevertheless, despite the multiplicity of readings, a *vulgata* emerged, allowing scholars to identify common features across the different witnesses.

Among the critical editions of Dante's poem, the Petrocchi edition, also the one used as a reference in this study, remains one of the standard points of reference, thoroughly documented in four volumes (Petrocchi 1966)<sup>9</sup>. Petrocchi's edition, however, is neither the only one available nor the most recent, as later editions have been produced, such as those by Lanza (1996), Sanguineti

---

<sup>5</sup> The term Florentine vernacular, even if not incorrect, is nevertheless restrictive with respect to the richness and complexity of the language of the *Divine Comedy*. For an in-depth discussion of the topic, see, among others, Nencioni (1989a: 22-28), Baldelli (1996), Manni (2013), Bellomo (2020: 267-288) e Frosini (2020: 113-132).

<sup>6</sup> <https://faculty.georgetown.edu/jod/cangrande.english.html>

<sup>7</sup> With regard to Dante's importance for the Italian language and to the subsequent figures who contributed to its normalization, numerous texts and studies have been devoted to the subject. Among the major surveys that aim to provide a comprehensive diachronic perspective are Castellani (2000), Migliorini (2001), Marazzini (2015), Rohlf (2021), among others.

<sup>8</sup> "Si sono conservati circa una ottantina di manoscritti stesi entro la metà del Trecento e ben quattrocento nel XV secolo" (Bellomo, 2020: 290) 'Approximately eighty manuscripts dating to the mid-fourteenth century have been preserved, along with as many as four hundred from the fifteenth century'.

<sup>9</sup> It should be noted that Petrocchi's edition is not the first critical edition of the text; earlier editions include those of Witte (1862), Moore (1900), Vandelli (1928), and Casella (1923).

(2001), and Inglese (2021), Tonelli-Trovato (2022). Despite not being the most recent, the decision to adopt Petrocchi's edition is motivated by the need to maintain consistency with a previously existing resource, namely DanteSearch (Tavoni 2022), which is described in the following section (1.1.1).

The importance of the *Divine Comedy* and the extensive scholarship and readership it has attracted have more recently led to the development of online resources documenting the poem (Tomazzoli 2024). Among these, the abovementioned DanteSearch platform (Tavoni 2022) has proved particularly useful for the present study.

### 1.1.1 DanteSearch

DanteSearch is an online resource<sup>10</sup> dedicated to the works of Dante Alighieri, both in Latin and in the vernacular. The project (Tavoni 2012; 2015) is the result of the work of various scholars and students, carried out under the supervision of Prof. Mirko Tavoni at the University of Pisa. The resource is freely accessible online and easily searchable through a graphical interface that provides access to Dante's works together with their lemmatization and morphological annotation. The morphological annotation consists of the word-level analysis of the part of speech and the grammatical features of each word under examination. Specifically, for each word the interface provides the lemma, its numerical position in the text, and its morphological features. For example, considering the final word of the poem, *stelle* 'stars', the interface displays its lemma (*stella* 'star'), its position (it is the 1005th words of the last *Canto* of *Paradiso*), and its morphological information through a specific alphanumeric code (s f 1 f p), which stands for noun, gender lemma feminine, first class, gender form feminine, number plural. In the case of the *Divine Comedy*, a syntactic annotation is also available, based on an analysis of clause structure types (e.g., concessive clauses, temporal clauses, among others). This work was carried out by Sara Gigli (2004) and later employed by Marta D'Amico (2014) in her study of complex sentence syntax in the poem.

The graphical interface also provides a query system (Tavoni 2022) that allows users to search for specific lemmas, word forms, or morphological tags, as well as syntactic information specifically available for the *Commedia*. In addition to the graphical interface, XML files containing the annotations are freely available<sup>11</sup>. Since the *Commedia* includes both syntactic and morphological annotation, these two layers have been encoded in separate XML files.

---

<sup>10</sup> <https://dantesearch.dantenetwork.it>

<sup>11</sup> <https://dantesearch.dantenetwork.it/DanteSearchXML.zip>

While representing a major step forward in making Dante’s texts more accessible and searchable, the extensive and valuable work carried out for the DanteSearch resource is limited by the adoption of an idiosyncratic annotation formalism. This lack of standardization prevents interoperability with other linguistic resources, thereby rendering DanteSearch something of an isolated case within the broader landscape of annotated corpora.

It is from this observation, and from the intrinsic value of the resource itself, that the idea and decision arose to integrate the *Divine Comedy*, building precisely on the annotation carried out in DanteSearch, into a broader framework capable of interacting with other online resources and adhering to a shared annotation formalism. Such an approach would situate Dante’s work in relation to other texts and languages, making possible comparative analyses, while also enabling the use of research tools and data-production methods that can foster both the dissemination of the material and scholarly discussion around it. The annotation formalism adopted, which belongs to the Universal Dependencies framework (De Marneffe et al. 2021), henceforth UD, and the rationale for adhering to it will be described in Section 1.4.4.

## 1.2 Old Italian and its digital resources

In the present study, the term Old Italian is used to refer to the language of Dante’s *Divina Commedia*. At first glance, this usage may appear to flatten the richness encompassed by the notion of Old Italian, which will be briefly discussed in this section; however, it should be emphasized that this choice responds to a practical need to distinguish between Modern Italian data, which are well attested in UD (Section 1.4.5.1), and the language of the *Divine Comedy*, which constitutes the first Old Italian text to be introduced in the context of UD. For this reason, the latter is referred to here as Old Italian in order to mark the diachronic boundary. The aim of this section, in addition to briefly outlining the complex and debated notion of Old Italian, is therefore to stress that the term is consciously employed with due caution.

The term Old Italian (*italiano antico*) is used in the *Grammatica dell’Italiano Antico* (Salvi & Renzi, 2010), where it is explicitly defined as “the Florentine of the thirteenth century, the first documented phase of the Italian language, and of the early fourteenth century” (Salvi & Renzi 2010: 7)<sup>12</sup>. The label was later extended to encompass all Tuscan diachronic varieties (Salvi & Renzi 2010: 7), thereby designating Tuscan specifically as Old Italian<sup>13</sup>. Although the language of the *Divina Commedia* has often been discussed in terms of its plurilingualism (Manni 2013: 85–94), the “solidly

---

<sup>12</sup> Original text: “il fiorentino del Duecento, prima fase documentata della lingua italiana, e dei primi del Trecento”.

<sup>13</sup> It should be specified that, in this way, the other Early Italo-Romance varieties attested on the Italian peninsula (Rohlf 1966–67; Castellani 2000; among others) are excluded from the definition of *Old Italian*.

Florentine structural basis of the poem's language" (Manni 2013: 106) is beyond dispute and thus allows the text to be classified within the domain of Old Italian. Another point of discussion concerns the conservative nature of Modern Italian in comparison with Old Italian, especially when contrasted with languages where the gap is more evident, such as Old French (Dardano 2012: 6). There is, however, a well-documented body of studies that argues for a distinction between Old and Modern Italian, supporting this view also on syntactic grounds (Tesi 2004). The present study adheres to the view that a distinction between Old and Modern Italian does exist<sup>14</sup>. Although the aim here is not to demonstrate such a distinction, a task that would require more heterogeneous data than a single-authored text, the hope is that, over time, broader and more systematic comparisons, supported by quantitative sampling techniques, may shed further light on the issue. It is evident that, in order to undertake such a task, in addition to the linguistic, textual, and historical-cultural expertise of scholars, a further fundamental requirement is the availability of corpora and resources that enable the consultation and use of Old Italian texts.

At present, among the main resources documenting Old Italian<sup>15</sup>, understood here as an umbrella term also encompassing the Early Italo-Romance varieties, is the *Opera del Vocabolario Italiano* (OVI) corpus<sup>16</sup> (Squillacioti 2021), a lexicographic corpus that includes Old Italian texts composed before the fifteenth century and represents one of the largest collections, with 3,840 texts<sup>17</sup>. The texts, however, although accessible through the GattoWeb interface<sup>18</sup>, do not include grammatical annotation tagsets and are not fully available for free download, sometimes due to editorial restrictions, thus making it difficult to carry out targeted linguistic research. Closely connected to the OVI corpus is the *Tesoro della Lingua Italiana delle Origini* (TLIO) corpus<sup>19</sup> (Beltrami 2003), which corresponds to the lemmatized version of a sample of OVI texts, comprising 3,452 texts<sup>20</sup>. In this case, the word forms are lemmatized and provided with a morphological tagset (which is, however, corpus-specific) that facilitates consultation. However, the corpus does not include any syntactic annotation.

Along with the OVI and TLIO corpora, other resources exist which extend over a broader chronological range. Among these is the MIDIA corpus (*Morfologia dell'Italiano in Diacronia*,

---

<sup>14</sup> Although the position of distinguishing between Old and Modern Italian is adopted, it has been necessary to classify Old Italian data together with Modern Italian one, due to the absence of a specific ISO language code for Old Italian (see Section 2.1.1). Some attempts have been made to request the creation of a dedicated ISO code, but this is a complex process involving institutional bodies and requiring considerable time as well as broad agreement.

<sup>15</sup> The present survey does not claim to be exhaustive with respect to the range of digital resources attesting Italian texts.

<sup>16</sup> [http://gattoweb.ovi.cnr.it/\(S\(vjpbjblh2bcpmtuh0h5hbnh\)\)/CatForm01.aspx](http://gattoweb.ovi.cnr.it/(S(vjpbjblh2bcpmtuh0h5hbnh))/CatForm01.aspx)

<sup>17</sup> [http://gattoweb.ovi.cnr.it/\(S\(4wufpevjxowpgqxt4htcjig\)\)/CatForm01.aspx](http://gattoweb.ovi.cnr.it/(S(4wufpevjxowpgqxt4htcjig))/CatForm01.aspx)

<sup>18</sup> [http://tlioweb.ovi.cnr.it/\(S\(crl0jhyzcai41dnzk0wjnj5i\)\)/HelpPages/Help01.html](http://tlioweb.ovi.cnr.it/(S(crl0jhyzcai41dnzk0wjnj5i))/HelpPages/Help01.html)

<sup>19</sup> [http://tlioweb.ovi.cnr.it/\(S\(crl0jhyzcai41dnzk0wjnj5i\)\)/CatForm01.aspx](http://tlioweb.ovi.cnr.it/(S(crl0jhyzcai41dnzk0wjnj5i))/CatForm01.aspx)

<sup>20</sup> [http://tlioweb.ovi.cnr.it/\(S\(z3gh53qtl3e3xzaygrpfmap\)\)/CatForm01.aspx](http://tlioweb.ovi.cnr.it/(S(z3gh53qtl3e3xzaygrpfmap))/CatForm01.aspx)

D’Achille & Grossmann 2017), a lemmatized and morphologically annotated collection, tagged for part of speech, covering Italian texts from the 13th century to the early 20th century. It spans five chronological periods and includes seven textual typologies. The corpus contains about 7.8 million tokens drawn from 800 texts<sup>21</sup>. Together with the MIDIA, the CODIT corpus (*Corpus Diacronico dell’Italiano*, Micheli 2022) is also worth mentioning. It is a balanced corpus containing about 29 million tokens. Like MIDIA, it is divided into five periods, but unlike the latter, CODIT presents texts in their entirety, whereas MIDIA selects only portions of them (Micheli 2022: 3). The corpus is intended to be enriched with lemma and part-of-speech annotation. Finally, another corpus worth mentioning for linguistic annotation and for its use in the training and evaluation of NLP models, although limited to part-of-speech tagging and lemmatization, is the GDLI-Quotation Corpus (Favaro et al. 2022). It was created through the digitization of the *Grande Dizionario della Lingua Italiana* (Battaglia 1961-2002), from which a selection of quotations was extracted, covering the entire history of the Italian language and thus ranging from the Middle Ages to the present day (Favaro et al. 2022: 94). The linguistic annotation follows the Universal Dependencies standard but, at the present stage, does not include syntactic annotation.

Although the survey presented here is not intended to be exhaustive with respect to the full range of digital resources for Old Italian, its focus has been to highlight the current lack of *syntactically* annotated resources for Old Italian, particularly within the framework of dependency syntax. This framework, in both its descriptive dimension and its applicability in NLP contexts, will be discussed in detail in the following Section 1.4.

### 1.3 Dependency grammar

As with the other sections of this introductory part, the present section provides a brief and necessarily non-exhaustive overview of dependency grammar. Dependency grammar<sup>22</sup> has a long history, with roots tracing back to classical antiquity (Hudson 2017). It is traditionally contrasted with Phrase Structure grammar, also known as Constituency grammar, a theory initiated by Bloomfield (1933) and later formalized by Noam Chomsky (1957; 1993). Phrase Structure grammar gained particular prominence in the 1950s and 1960s, both within the structuralist school (Bloomfield 1933) and Transformational-Generative framework (Chomsky 1957; 1993), as many linguists regarded it as a more suitable model than dependency grammar. The latter, however, regained visibility thanks to the

---

<sup>21</sup> <https://www.corpusmidia.unito.it>

<sup>22</sup> This work adopts the term dependency grammar as used in Hudson (2017). It should be noted that Mel’čuk (1988) appears to prefer the terms *dependency theory* or *dependency language*, using them interchangeably with dependency grammar (Mel’čuk, 1988: 12).

contributions of Tesnière (1959; 2015), and Mel'čuk (1988), among others. More recently, the rise of computation linguistics, which primarily relies on dependency syntax, has further expanded the use and application of dependency grammar, as will be discussed in this section.

This Section, after briefly providing an overview of dependency syntax (Section 1.3.1) and discussing the differences between dependency and constituency frameworks (Section 1.3.2), provides an overview of dependency syntax in natural language processing (NLP) and its application within the Universal Dependency framework (1.3.3), the environment in which the annotation of the *Divine Comedy* is situated.

### 1.3.1 Dependency syntax

It should be noted that the notion of dependency can involve different types, such as semantic, syntactic, and morphological dependencies. For the purposes of the present work, however, the focus will be on syntactic dependency, with particular attention to dependency syntax as theorized by Mel'čuk (1988). When discussing dependency syntax, Mel'čuk (1988) addresses two main issues: the nature of syntactic relations and the nature of syntactic structure.

**1. Syntactic relations:** Starting from the basics, Mel'čuk (1988) distinguishes two types of expressive means that a natural language uses to encode the information conveyed by a sentence: (i) lexical means, which correspond to words, and (ii) non-lexical means, which correspond to linear order of word forms (syntax), prosody and inflections. Among the latter, syntax is considered the most important, as it is universal, that is, present in every language and in all sentences. Syntax, or syntactic means, conveys information through a specific formal device, the syntactic structure. A syntactic structure consists of: (i) all the word forms appearing in the sentence (or lexemes), and (ii) a set of binary n-relations defined over the word forms, the syntactic relations. A syntactic relation, also known as syntactic dependency, is defined as a binary relation between word forms, that must be (i) anti-symmetric, meaning that dependency relations are unidirectional, (ii) antireflexive, as a word form cannot be linearly ordered or inflected with respect to itself, and (iii) anti-transitive. Moreover, syntactic relations are labeled, which raises the issue of establishing a universal inventory of dependency relations that is valid across all languages.

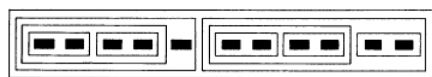
**2. Syntactic structure:** A syntactic structure is a set of word forms linked by syntactic dependency relations. This structure is represented by connected, directed, labeled graph, commonly referred to

as a syntactic tree<sup>23</sup>, where the vertices or nodes represent word forms (usually one node per word forms, though exceptions exist in the case of multiword tokens), and the arcs or branches that stands for syntactic dependency relations. A syntactic structure must contain a top node or root, which is obligatory and serves as the absolute governor of the structure. Furthermore, no node may simultaneously depend on two or more nodes, thereby ensuring the uniqueness of the syntactic governor.

### 1.3.2 Dependency versus Constituency Grammar and the notion of treebanks

To draw a brief contrast between dependency and constituency grammar, Mel'čuk (1988) highlighted five main differences between the syntactic structure of natural sentences when represented through dependency versus constituency:

1. **Configuration vs relation:** Constituency is based on constituents (or phrases<sup>24</sup>) and establishes *inclusion* (i.e., belonging to a specific category) among them, whereas dependency, whose primary unit is the word, sets *relations* between words. Images 1 and 2 provide graphical representations of constituency inclusion (Image 1) and dependency relations (Image 2) as represented by Mel'čuk (1988: 13-14):



**Image 1.** constituent inclusion



**Image 2.** word relation

This distinction does not imply that dependency lacks grouping, understood as hierarchical links between elements, but rather that the nature of such grouping is different. In dependency grammar, a grouping, that is, a word together with its associated phrase, corresponds to a word with its dependent words (a subtree).

In constituency grammar, the representation of wordforms is *configuration* and *distributional*, meaning that the syntactic function of a word is defined by its position (or distribution/configuration) in the syntactic structure. In contrast, in dependency grammar, the representation of wordforms is *relational*, meaning that wordforms participate in relations that link them to other wordforms, determining their syntactic roles in the sentence.

<sup>23</sup> A syntactic tree is an acyclic, directed, rooted graph with a linear ordering of their nodes (Havelka 2007).

<sup>24</sup> Although the term *phrase* is often associated with constituency grammar, it is also used in a more neutral sense. From this point forward, I employ the term *phrase* neutrally, making it applicable in both dependency and constituency contexts.

2. **Syntax vs words:** In constituency grammar, categorization (i.e., the syntactic class membership) is an integral part of the syntactic representation, and no additional information beyond syntax is required (thus posing several challenges to the role of morphology). In contrast, in dependency grammar, syntactic class membership does not depend on syntax, relying on lexemes, which are dictionary entries. This means that a morphological structure is involved and it stands as a distinct level of analysis with respect to the syntactic level. In dependency grammar, words are the units of syntax, and together with the lexicon, they play a crucial role. This leads to the lexical integrity principle, which asserts that syntax does not create words, as words are atomic units (Hudson 2017)
3. **Non-terminal vs terminal:** In constituency grammar, nodes (i.e., the graphic representation in the syntactic structure) may be non-terminal, which means that one-to-one-or-more relations (elements to nodes) is feasible (Tesnière 2015). In contrast, dependency grammar only contains terminal nodes, thus implying a one-to-one relation, resulting in syntactic headed structures. But Mel'čuk (1988) considers also cases of nodes corresponding to zero wordforms and instances where amalgamated wordforms can be represented by two nodes.
4. **Linear word order vs no linear word order:** The existence of a basic linear word order is considered in constituency grammar, whereas dependency grammar is two-dimensional and does not consider a linear order of words that should be reconstructed (i.e., does not consider movements).
5. **Syntactic relations:** Dependency grammar specifies the type of syntactic relations that link two items, giving importance to grammatical functions that are not involved in constituency grammar. This fact poses a challenge to dependency syntax, that is to create inventories of universal (deep-)syntactic relations and language specific (surface-)syntactic relations.

As the divergences highlighted above make clear, the theoretical framework adopted has a direct impact on the representation of syntactic structure, with consequences that extend to linguistic (syntactic) annotation.

Before moving to the next paragraph, which introduces the connection between dependency grammar and NLP, it is important to define the concept of a treebank. A treebank is, by definition, a collection

of syntactic<sup>25</sup> trees annotated according to a specific formalism, which may vary depending on the theoretical framework adopted. One of the most widely known constituency-based treebank for English is the Penn Treebank (Taylor, Marcus, & Santorini 2003). By contrast, the Prague Dependency Treebank (Hajič 1998) follows a dependency-based approach, while the BulTreeBank (Simov et al. 2002) is based on the Head-driven Phrase Structure Grammar (HPSG) formalism (Pollard & Sag 1994). Treebanks are essential for the study of syntax, as they provide samples of annotated syntactic trees that, however, are necessarily shaped by the underlying theory.

### 1.3.3 The impact of dependency grammar in NLP: Universal Dependencies

As de Marneffe and Nivre (2019) clearly state, the primacy of dependency grammar can be argued for three reasons: (i) its capacity for cross-linguistic generalization, which makes it particularly useful for typological comparisons; (ii) its applicability to studies of processing and sentence comprehension, which benefit from its ability to “operationalize measures that can be linked to language ordering preferences and to syntactic configurations” (de Marneffe & Nivre 2019: 207); and (iii) its transparency and simplicity of representation, which make such structures preferable, though not exclusively, for computational linguistics tasks. The primacy of dependency annotation, which has led to the creation of several annotated corpora, namely treebanks<sup>26</sup>, following this theoretical approach, has also brought about the need for a common dependency annotation, one that follows a unified formalism in order to make data comparable. In this respect, the Universal Dependencies (UD) framework, described in Section 1.4.3.1, was created with the aim of establishing a common universal annotation formalism and is today regarded as the *de facto* standard for dependency annotation.

## 1.4 Universal Dependencies

Universal Dependencies (henceforth UD)<sup>27</sup> is a collaborative framework for cross-linguistic morphosyntactic annotation of human languages (De Marneffe et al. 2021). Through its open community, which is constantly growing, it provides corpora with consistent morphosyntactic annotations across a wide range of languages. The project, initiated in 2014<sup>28</sup>, addresses the issue of annotation schemes varying significantly across languages, which poses challenges for conducting

---

<sup>25</sup> Treebanks can also be enriched with semantic annotation, as in PropBank (Palmer et al., 2005; Bonial et al., 2014).

<sup>26</sup> By way of example, one can mention the Dutch Treebank (Bouma et al. 2001), the Czech Treebank (Hajič et al. 2001), the Danish Dependency Treebank (Kromann 2003), the Turkish Treebank (Oflazer et al. 2003), the Stanford Typed Dependencies (de Marneffe et al. 2006), and the Swedish Treebank (Nivre et al. 2007).

<sup>27</sup> <https://universaldependencies.org>

<sup>28</sup> The initial version of the annotation guidelines was released in October 2014 (Nivre et al., 2016: 1660).

comparative linguistic evaluations and cross-lingual experiments (Nivre et al. 2016). The proposal of a shared annotation scheme aims to facilitate cross-linguistic comparison by capturing similarities among languages while preserving the idiosyncrasies of different typological languages (De Marneffe et al. 2021).

#### 1.4.1 History of UD

The need for a consistent and shared annotation scheme arises from the necessity to avoid divergent analyses of parallel structures across languages. This objective has been pursued by several initiatives (Zeman 2008; McDonald et al. 2013; Tsarfaty 2013; Rosa et al. 2014; de Marneffe et al. 2014), which ultimately converged in the UD framework (Nivre 2016:1659). UD brings together multiple annotation layers developed by earlier projects, all sharing the goal of establishing cross-linguistic consistency.

At the morphological level, UD integrates both the Google Universal tagset (Petrov et al. 2012) and Intersect (Zeman 2008). The Google Universal tagset was developed to reduce divergence in part-of-speech (henceforth PoS) annotation across languages. It defines 12 universal PoS categories covering the most frequent word classes (Petrov 2012: 2090): NOUN (noun), VERB (verb), ADJ (adjective), ADV (adverb), PRON (pronoun), DET (determiner and article), ADP (preposition and postposition), NUM (numeral), CONJ (conjunction), PRT (particle), ‘.’ (punctuation mark) and X (other categories such as abbreviations or foreign words). Intersect, in contrast, was originally designed as a conversion tool, providing a universal inventory for mapping and translating features between different tagsets (Zeman 2008: 214). At the syntactic level, UD incorporates the Stanford Dependencies (de Marneffe et al. 2006; de Marneffe and Manning 2008; de Marneffe et al. 2014), a dependency-based scheme inspired by Lexical-Function Grammar (Bresnan, 2001) and designed as a practical model for sentence representation (de Marneffe et al. 2008: 2). Originally developed for English, the Stanford Dependencies framework has since been adapted to various languages, including Chinese (Chang et al. 2009), Italian (Bosco et al. 2013), Finnish (Haverinen et al. 2014), Persian (Seraji et al. 2014) and Russian (Lipenkova & Souček 2014).

These efforts toward universality and consistency ultimately converged in UD, which builds on an extended version of the abovementioned Google Universal tagset, a refined subset of the Intersect feature inventory, and universal Stanford Dependencies.

#### 1.4.2 Introduction to UD

In its current release (version 2.16)<sup>29</sup>, UD comprises more than 200 treebanks across over 150 languages (specifically, 319 treebanks and 179 languages), covering a variety of language families and genera, including Indo-European, Niger-Congo, Afro-Asiatic and Austronesian. Within Indo-European, for example, the framework covers Romance, Germanic, and Slavic languages, among others. In addition to modern languages, UD also includes historical languages such as Ancient Greek, Latin, Old French, Akkadian and Classical Chinese, thereby enabling comparative analyses from a diachronic perspective. As the name suggests, UD is a framework based on *dependency* syntax, a scheme widely adopted in NLP as an alternative to phrase-structure representations (see Section 1.3.2)<sup>30</sup>.

Dependency relations are established among syntactic words, which constitute the basic annotation units in UD (Nivre 2008: 1660). A syntactic word is defined on syntactic grounds, that is, as any word participating in a syntactic relation. This definition contrasts with the more traditional notion of an orthographic word (or token), understood as an orthographic unit of segmentation. For instance, in UD the Italian word *dimmi* ('tell me', lit. 'tell.meDAT'), although a single orthographic word, is analyzed as two syntactic words, *di* 'tell' and *mi* 'to me', since each participates in a distinct syntactic relation. Punctuation is likewise treated as a syntactic word, as every punctuation mark is assigned a dependency relation. It is worth noting that UD terminology differs slightly from that of Tesnière (2015 [1959]). Tesnière uses the term *word* to denote a simple segment of the spoken chain, defining it as the linear unit of the sentence without syntactic status (Tesnière (2015:41 [1959])). In his framework, syntactic relations are instead established by *nuclei*, conceived as the elementary syntactic entities and “the constitutive cell of the living organism that is the sentence” (Tesnière 2015:38-39 [1959]).

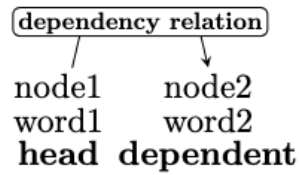
Dependency relations among syntactic words are represented by syntactic trees (see Image 3). In such trees, the nodes<sup>31</sup> correspond to syntactic words, while the directed edges between two nodes (a head and its dependent) encode the syntactic relation between them. This relation is binary and asymmetrical, and is conventionally represented by an arrow.

---

<sup>29</sup> See: <http://hdl.handle.net/11234/1-5901>

<sup>30</sup> The widespread use of dependency syntax in NLP compared to phrase-structure grammar is supported by both theoretical and practical motivations, see (Passarotti 2018).

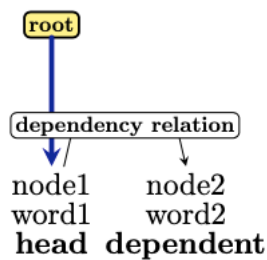
<sup>31</sup> Tesnière makes a more fine-grained distinction between ‘node’ and ‘nuclei’, identifying the node as “nothing more than a geometrical point” (Tesnière 2015:39 [1959]), i.e., “a vertex in the (graphic representation of) sentence structure.” (Tesnière 2015:39 [1959], footnote 43 by translators).



**Image 3.** The basic schema of a tree

As illustrated in Image 3, word1, represented as node1, functions as the head of word2, represented as node2. The dependency relation between the two is graphically encoded by an arrow extending from the head (word1) to its dependent (word2).

Accordingly, sentences are represented as dependency trees in which each word is connected to others through syntactic relations, with the exception of a notional root. The root serves as the origin of the tree and has no head. It is typically indicated by an arrow originating from a dummy head (Image 4).

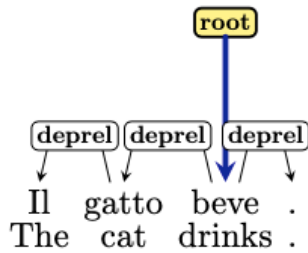


**Image 4.** Root in a dependency tree

One of the core principles of UD is the primacy of content words<sup>32</sup>, whereby lexical items function as heads while function words depend on a content head (Nivre 2016; De Marneffe 2021)<sup>33</sup>. In addition, UD is predicate-centric: the predicate, whether a verb or the nominal/adjectival element in copular constructions, functions as the root of the syntactic tree. Image 5 illustrates this with a basic syntactic tree for the Italian sentence *Il gatto beve*. ‘The cat drinks.’

<sup>32</sup> <https://universaldependencies.org/u/overview/syntax.html#the-primacy-of-content-words>

<sup>33</sup> This principle facilitates the identification of parallel structures across languages, since function words in one language may correspond to morphological inflections in another, or may be altogether absent (Nivre 2016: 1662).



**Image 5.** Syntactic tree of the Italian sentence

The root of the tree is the verb *beve* ‘drinks’, with the noun *gatto* ‘cat’ directly dependent on it via its dependency relation (henceforth *deprel*). The article *il* ‘the’, as a function word, depends on the noun *gatto* ‘cat’. As noted above, punctuation marks (the full stop in Image 5), are likewise treated as syntactic words and thus participate in the tree, receiving a dependency relation directly from the root.

### 1.4.3 Dependency relations in UD

Concerning the syntactic taxonomy of dependency relations<sup>34</sup>, UD draws a key distinction between core arguments (e.g., subjects, objects and clausal complements) and non-core arguments, while not differentiating between adjuncts (modifiers) and oblique arguments (arguments selected by a head but not classified as core). The notion of core argument is closely tied to the semantic roles of participants, as core arguments typically encode the “agent and patient or theme roles of predicates in their unmarked valence”<sup>35</sup>.

Relations are organized by distinguishing between three types of structures (Nivre 2016: 1661; De Marneffe 2021: 257-259):

- nominals: structures that typically refer to entities;
- clauses: structures that typically refer to events;
- modifier words: structures that describe an attribute of entities or events.

The grammatical relations in UD (v.2) are drawn from a universal typology of 37 relations, which represent the surface semantic predicate-argument structure of sentences, facilitating crosslinguistic comparison (De Marneffe, 2021: 265).

Table 1 lists the dependency relations (*deprels*) defined in UD (v.2)<sup>36</sup>:

	Nominals	Clauses	Modifier words	Function words
--	----------	---------	----------------	----------------

<sup>34</sup> <https://universaldependencies.org/u/overview/syntax.html#the-taxonomy-of-typed-dependencies>

<sup>35</sup> <https://universaldependencies.org/u/overview/syntax.html#the-taxonomy-of-typed-dependencies>, access 16.10.2024

<sup>36</sup> <https://universaldependencies.org/u/dep/> reports a revisited version of the relations described in (De Marneffe, 2014).

<b>Core arguments</b>	nsubj obj iobj	csubj ccomp xcomp		
<b>Non-core dependents</b>	obl vocative expl dislocated	advcl	advmod discourse	aux cop mark
<b>Nominal dependents</b>	nmod appos nummod	acl	amod	det clf case
<b>Coordination</b>	<b>Headless</b>	<b>Loose</b>	<b>Special</b>	<b>Other</b>
conj cc	fixed flat	list parataxis	compound orphan goeswith reparandum	punct root dep

**Table 1.** list of deprels

As shown in Table 1<sup>37</sup>, deprels are classified according to the structural categories of the dependent (columns) and the functional categories with respect to the head (rows)<sup>38</sup>. Structural categories include nominals, clauses, modifiers, and function words, while functional categories distinguish between core arguments of clausal predicates, non-core dependents of clausal predicates, and nominal dependents. The lower part of Table 1 lists exceptional relations, which extend the notion of dependency to broader cases. These include coordination, multi-word expressions<sup>39</sup>, loose relations, ellipsis, disfluencies, orthographic errors, as well as relations for punctuation, clausal heads, and other specific phenomena.

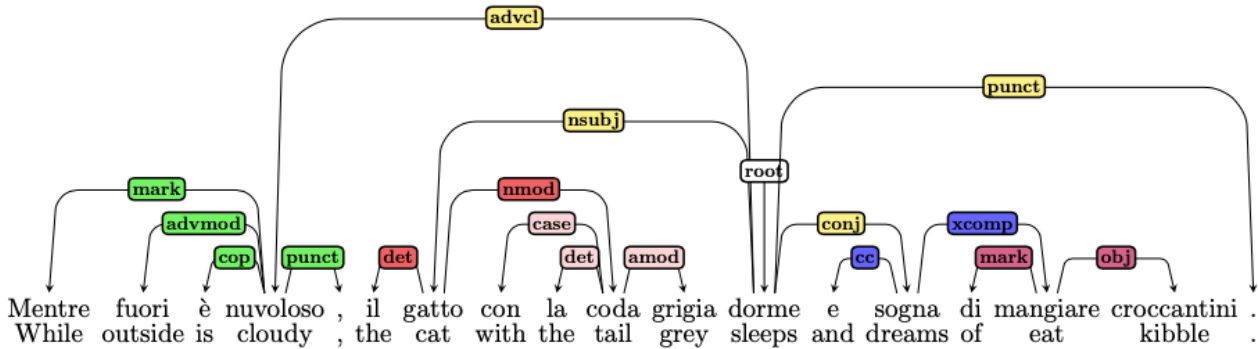
Image 6 shows a syntactic tree with deprels for an Italian sentence<sup>40</sup>: *Mentre fuori è nuvoloso, il gatto con la coda grigia dorme e sogna di mangiare croccantini.*, ‘While it is cloudy outside, the cat with the grey tail sleeps and dreams of eating kibble.’.

<sup>37</sup> List of the abbreviations used in Table 1: *nsubj* ‘nominal subject’, *obj* ‘object’, *iobj* ‘indirect object’, *csubj* ‘clausal subject’, *ccomp* ‘clausal complement’, *xcomp* ‘open clausal complement’, *obl* ‘oblique nominal’, *vocative*, *expl* ‘expletive’, *dislocated*, *advcl* ‘adverbial clause modifier’, *advmod* ‘adverbial modifier’, *discourse*, *aux* ‘auxiliary’, *cop* ‘copula’, *mark* ‘marker’, *nmod* ‘nominal modifier’, *appos* ‘appositional modifier’, *nummod* ‘numeric modifier’, *acl* ‘adnominal clause’ (clausal modifier of noun), *amod* ‘adjectival modifier’, *det* ‘determiner’, *clf* ‘classifier’, *case* ‘case marking’, *conj* ‘conjunct’, *cc* ‘coordinating conjunction’, *fixed* ‘fixed multiword expression’, *flat* ‘flat expression’, *list*, *parataxis*, *compound*, *orphan*, *goeswith*, *reparandum* ‘overridden disfluency’, *punct* ‘punctuation’, *root*, *dep* ‘unspecified dependency’.

<sup>38</sup> For further insights and examples regarding deprels, refer to: <https://universaldependencies.org/u/dep/all.html> and De Marneffe (2021: 266).

<sup>39</sup> Multi-word expressions are sequences of words that function as a single lexical unit.

<sup>40</sup> Although the use of artificially constructed sentences not attested in corpora is generally discouraged, in this case the sentence was created *ad hoc* for illustrative purposes, in order to demonstrate the functioning of UD dependency relations.



**Image 6.** UD deprels in an annotated Italian sentence

The syntactic tree is explained by beginning with its root, proceeding through its direct dependents, and then analyzing each subtree (i.e., portions of the tree) down to the leaves.

Since UD is predicate-centric, the `root` of the tree corresponds to the main predicate of the sentence, in this case the verb *dorme* ‘sleeps’. Attached to this `root` via the `conj` (conjunct) deprel is the coordinated verb *sogna* ‘dreams’. Although coordination is generally considered a symmetrical relation (Goodall, 2017), UD treats the first conjunct as the parent (or “technical head”) of all subsequent conjuncts, which are linked through the `conj` relation. This approach ensures compliance with the structural constraints of dependency trees and applies universally, including to coordination between nominals and adjectives<sup>41</sup>. Coordination conjunctions are attached to their conjunct with the `cc` ‘coordinating conjunction’ deprel. In Image 6, this is exemplified by *e* ‘and’, which depends on the head of the conjunct, *sogna* ‘dreams’. Alongside the `conj` deprel, directly dependent on the `root` and placed to its right, is the full stop, annotated with the `punct` ‘punctuation’ deprel. On the left side of the root lies its nominal subject, annotated with the `nsubj` deprel, namely *gatto* ‘cat’, together with a subordinate clause whose head, *nuvoloso* ‘cloudy’, is marked with the `advcl` deprel ‘adverbial clause’. The `advcl` relation is employed for all clauses that modify the predicate without functioning as a core complement (e.g., temporal, result, purpose, or final clauses, among others).

The subtrees directly attached to the root include the one headed by the adverbial clause (green labels), the one headed by the nominal subject (red labels), and the one headed by the conjunct (blue labels).

The adverbial clause subtree exhibits a copular structure, *è nuvoloso* ‘it is cloudy’, where the adjective *nuvoloso* ‘cloudy’ is the head of the subtree, directly attached to the `root`, with all other elements (in green) depend on it. The copula *è* ‘is’ is linked to the head of the `advcl` through the `cop` ‘copula’

<sup>41</sup> <https://universaldependencies.org/u/overview/complex-syntax.html>

deprel, while the adverb *fuori* ‘outside’ is linked via the `advmod` deprel, used for adverbial modifiers, i.e., adverbs or adverbial phrases that modify a predicate or another modifier. Also dependent on the head of the `advcl` is the word *mentre* ‘while’, annotated with the `mark` ‘marker’ deprel, which signals the subordination of a clause. Finally, the comma is attached to the head of the `advcl` with the `punct` relation.

The nominal subject subtree is headed by the content word *gatto* ‘cat’, which depends directly on the root. This head has two immediate dependents (in red): the masculine article *il* ‘the’, marked with the `deprel det` ‘determiner’, and the noun *coda* ‘tail’, marked with the `nmod` relation ‘nominal modifier’, used for nominal modifiers of another noun or noun phrase.

The nominal modifier *coda* ‘tail’ has three dependents (in pink): the preposition *con* ‘with’ annotated with the `case` ‘case marking’ deprel, used to signal case-marking elements; the feminine article *la* ‘the’, attached via `det` deprel; and the adjective *grigia* ‘grey’, which depends on the noun it modifies (*coda*, ‘tail’) through the `amod` relation, ‘adjectival modifier’.

Finally, the conjunct subtree, headed by the verb *sogna* ‘dreams’, has two dependents (in blue): the coordinating conjunction *e* ‘and’, attached with `cc` deprel, and the dependent clause *di mangiare croccantini* ‘to eat kibble’. This clause is headed by the verb *mangiare* ‘to eat’, annotated with the `xcomp` deprel ‘open clausal complement’. The `xcomp` deprel is used for clauses that are core arguments of the verb but lack their own subject, with a referent determined by an argument external to the `xcomp` (in this case, the subject *gatto* ‘cat’ is external to the `xcomp`). The verbal head of the `xcomp` has two dependents (in purple): the marker *di* ‘of’ attached with `mark` deprel, and the object *croccantini* ‘kibble’, attached with the `obj` deprel.

In addition to deprels, UD also provides a set of subtypes<sup>42</sup> that may be specific to individual languages or groups of languages. These subtypes are optional in the annotation process and are intended to capture language- or group-specific peculiarities.

These subtypes are optional in the annotation process and are intended to capture language- or group-specific peculiarities. In the annotation, subtypes follow the deprel, separated by a colon (e.g. `aux:pass`, for passive auxiliaries, where the subtype `:pass` specifies the passive voice), according to the format `universal:extension`. Among the most widely shared subtypes are `relcl` ‘relative’, `pass` ‘passive’, `agent` ‘agent’, `poss` ‘possessive’, among others. Some subtypes,

---

<sup>42</sup> <https://universaldependencies.org/ext-dep-index.html>

however, are language-specific; for example, the subtype `:purp` ‘purposive’ is used exclusively in the Abkhaz<sup>43</sup> language.

#### 1.4.4 Enhanced Dependencies

UD also provides an additional layer of annotation to account for relations not represented in its basic scheme, resulting in the creation of a graph that may include cycles. This extension, known as Enhanced Dependencies<sup>44</sup>, introduces several types of enhancements (Nivre 2020: 4039-4040), which are not captured by a standard-basic UD annotation. This paragraph briefly outlines the implementations provided by Enhanced UD<sup>45</sup>:

- null nodes for elided predicate: captures elided predicates by introducing an empty node that represents the omitted predicate, thereby restoring the dependency relations that are lost in basic UD annotation (or obscured by the `orphan deprel`)<sup>46</sup>.
- propagation of conjuncts: makes explicit the shared dependents that, in basic UD annotation, are attached only to one conjunct.
- controlled and raised subjects: similar to the propagation of conjuncts, this enhancement makes explicit the shared arguments between the matrix predicate and the embedded one, which in basic UD annotation are only attached to the matrix predicate.
- relative pronouns: signals the coreferential status of relative pronouns and makes explicit the relation between the verb of the relative clause and its antecedent, which is not annotated in basic UD.
- case information: augments modifier relations by specifying the adposition or the case information in the dependency relation.

The enhancement of annotation covers the aforementioned types, although not all of them need to be applied simultaneously; the use of a single type is sufficient to render a treebank enhanced. This optionality, both in the adoption of enhancement and in the selection of which enhancements to include, tends to discourage the systematic use of Enhanced Dependencies and, at the same time, complicates the tracking of treebanks<sup>47</sup> that implement them. It should also be noted that some of the

---

<sup>43</sup> The Abkhaz language is a Northwest Caucasian language. Refer to [https://wals.info/languoid/lect/wals\\_code\\_abk](https://wals.info/languoid/lect/wals_code_abk) for more details.

<sup>44</sup> <https://universaldependencies.org/v2/enhanced.html>

<sup>45</sup> For a more detailed description, see: <https://universaldependencies.org/v2/enhanced.html>

<sup>46</sup> A detailed discussion of the treatment of elided elements in UD is provided in Section 3.4.2.

<sup>47</sup> In version 2.15, only 19 treebanks included all types of enhancement, while 39 treebanks contained at least one type. In total, 24 languages featured at least one enhancement type in at least one treebank. These data, which are not easily computed automatically, were kindly provided by Daniel Zeman.

enhanced treebanks were automatically converted using systems specifically designed to predict enhanced dependencies from basic dependencies (Nivre 2020), such as the rule-based English system adapted from Schuster and Manning (2016) and the data-driven Finnish approach based on the supervised method by Nyblom et al. (2013). Specifically for Italian treebanks (further details on Italian treebanks in UD will be provided in the next paragraph 1.4.5), this task was carried out through a semi-automatic rule-based process (Simi et al. 2018), in which nodes subject to enhancement were identified using linguistic heuristics (e.g., propagating heads and dependents in coordination structures). The resulting output was subsequently reviewed by a human annotator and employed to automatically enhance the basic UD tree, or to reshape it, if necessary (Simi et al. 2018). Among the difficulties highlighted in the process of enhancement, the creation of empty nodes was identified as the most problematic to handle among UD enhancement (Simi 2018: 3). The complexity of this task<sup>48</sup>, along with the stated rarity of such phenomena in treebanks (Simi et al. 2018: 3), led to the decision to prioritize the elision of subject over predicates, as the former is considered by the authors to be more meaningful for the Italian language. The issue of enhancing treebanks with empty nodes for ellipsis, and in particular with predicate ellipsis, will be further discussed and commented on in Chapter 4.

#### 1.4.5 Italian treebanks in UD

Among the languages represented in UD is Italian. In version 2.16, UD comprises nine<sup>49</sup> treebanks, spanning different genres and containing a substantial amount of data (approximately 942K token<sup>50</sup>). With regard to their temporal coverage, all of them contain modern Italian data and do not extend to diachronic analyses of Old Italian (in the broad sense). Table 2 lists the treebanks, together with the number of syntactic words and the genres they represent.

Trebank	Syntactic words	Genre
ISDT	298K	legal, news, wiki
VIT	280K	news, non fiction
ParTUT	55K	legal, news, wiki
ParlaMint	20K	government legal
TWITTIRO	29K	social
Valico	6K	learner-essays
PoSTWITA	124K	social
MarkIT	40K	grammar-examples
PUD	23K	news, wiki

**Table 2.** Italian treebanks (excluding *Italian-Old*)

<sup>48</sup> The complexity is due to several factors, including the correct insertion point of the empty node, the presence of clitics referring to the verb, and phraseological verbs, that require copying more material than just the verb itself (Simi, 2018: 3).

<sup>49</sup> The number rises to ten if Italian-Old is included, which is introduced in Chapter 2.

<sup>50</sup> <https://universaldependencies.org>

The largest treebank is ISDT (Bosco et al., 2013; Bosco et al., 2014), Italian Stanford Dependency Treebank, 298K syntactic words, which covers legal texts, news, and wiki articles. This resource is available with both basic UD annotation and Enhanced Dependencies, and it is the only Italian treebank currently providing enhancement. In terms of size, ISDT is followed by the Venice Italian Treebank (VIT), which contains news and non-fictional texts (Delmonte et al. 2007; Delmonte 2009). The third largest treebank is PoSTWITA (Sanguinetti et al. 2018), counting 124K syntactic words, which, together with TWITTIRO (Cignarella et al. 2019), represents Italian social media language. ParTUT (Sanguinetti et al. 2014a,b) is a treebank including legal, news, and wiki data, while PUD (McDonald et al. 2013), which is part of the Parallel Universal Dependencies treebanks<sup>51</sup>, contains news and wiki texts. The domain of government and legal language is instead documented by ParlaMint (Agnoloni et al. 2022), counting 20K syntactic words. Finally, the VALICO-UD (Di Nuovo et al., 2022) treebank collects learner texts, whereas MarkIT (Paccosi et al. 2022) constitutes a corpus of grammar examples. Among the Italian treebanks, ISDT, VIT, ParTUT, PoSTWITA, and PUD originate from a conversion of earlier annotation schemes, whereas ParlaMint, TWITTIRO, VALICO-UD, and MarkIT were automatically annotated from pre-existing texts or corpora and subsequently manually revised.

#### 1.4.6 The CoNLL-U format

As mentioned in Section 1.4, one of the aims of UD is to provide a common and shared formalism to enable comparison. To this end, UD adopts a shared format: the CoNLL-U file<sup>52</sup>.

CoNLL-U encodes tokens using a tab-delimited structure, where each line specifies their annotation through ten fields. In cases where a field has no applicable value or has not been annotated, the symbol '\_' is inserted. The organization of these ten fields is as follows:

- ID: an integer that assigns a sequential index to each word. For multiword tokens, the index is given as a range.
- FORM: the word form.
- LEMMA: the lemma of the word.
- UPOS: the Universal part-of-speech tag<sup>53</sup>, which marks the core part-of-speech categories.
- XPOS: the language-specific part-of-speech tag, which may also encode morphological information.

---

<sup>51</sup> <https://universaldependencies.org/conll17/>

<sup>52</sup> <https://universaldependencies.org/format.html>

<sup>53</sup> <https://universaldependencies.org/u/pos/index.html>

- FEATS: the set of morphological features<sup>54</sup>, drawn either from the universal feature inventory or defined in a language-specific way. Features are represented in a 'Name=Value' format, with multiple features separated by a '|'.
- HEAD: the ID of the syntactic head of the current word. If the word is the root of the syntactic tree (i.e., it has no head), the value is 0.
- DEPREL: the UD dependency relation linking the word to its head. If the word has no head, the relation is `root`.
- DEPS: an optional field that encodes the enhanced dependency graph (see Section 1.4.4). If no enhanced representation is provided, the field is filled with the symbol `'_'`.
- MISC: a field reserved for any additional annotation and comment.

An example of CoNLL-U file with an Old Italian sentence is provided in Image 7. The example corresponds to the final sentence of the *Inferno*: *E quindi uscimmo a riveder le stelle*. ‘Thence we came forth to rebehold the stars.’<sup>55</sup> (*Inf.* XXXIV, v. 139).

```
# sent_id = OldItalian_Dante_Inferno-1228
# text = E quindi uscimmo a riveder le stelle.
1      E      e      CCONJ  cscc  _      3      cc      3:cc      Canto=34|Verso=139
2      quindi quindi ADV    b      _      3      advmod 3:advmod      Canto=34|Verso=139
3      uscimmo uscire VERB   vi3irp1
      Aspect=Perf|Mood=Ind|Number=Plur|Person=1|Tense=Past|VerbForm=Fin 0      root      0:root
      Canto=34|Verso=139
4      a      a      ADP    epsf  _      5      mark    5:mark      Canto=34|Verso=139
5      riveder rivedere VERB   vta2fp Tense=Pres|VerbForm=Inf|Voice=Act 3      advcl
      3:advcl Canto=34|Verso=139
6      le      il      DET    rdffp Definite=Def|Gender=Fem|Number=Plur|PronType=Art 7      det
      7:det Canto=34|Verso=139
7      stelle stella NOUN   sf1fp  Gender=Fem|Number=Plur5      obj      5:obj
      Canto=34|SpaceAfter=No|Verso=139
8      .      .      PUNCT  sf1fp  _      3      punct   3:punct      Canto=34|Verso=139
```

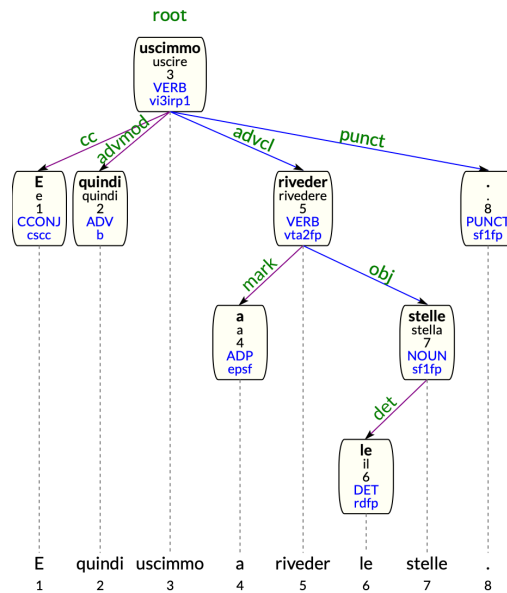
**Image 7.** CoNLL-U file of a sentence

Details about the specific annotation system will be provided in Chapter 2. It is important to note here that the first two lines in Image 7, introduced by the symbol `#`, are comment lines that record information about the sentence ID (`sent_id`) and the text. These comments are followed by the linguistic annotation encoded in the CoNLL-U format, which is organized into ten tab-separated fields containing the information described above.

<sup>54</sup> <https://universaldependencies.org/u/feat/index.html>

<sup>55</sup> The translations of the *Divina Commedia*, edited by Allen Mandelbaum and Henry Wadsworth Longfellow, cited in this work are taken from *Digital Dante*: (<https://digitaldante.columbia.edu/dante/divine-comedy/>).

Since this format is machine-readable, several tools are available to generate a more human-readable representation. Among these, CoNLL-U Editor<sup>56</sup> (Heinecke 2019) provides an efficient local solution. For remote visualization, Arborator-Grew<sup>57</sup> (Guibon et al. 2020) requires file upload but offers a wide range of functionalities. Image 8 shows the same sentence as in Image 7, visualized with CoNLL-U Editor in its tree structure.



**Image 8.** CoNLL-U file in a human readable format: CoNLL-U Editor

<sup>56</sup> <https://github.com/Orange-OpenSource/conllueditor>

<sup>57</sup> <https://arborator.grew.fr/#/>

## 2. Building *Italian-Old* treebank

The construction of a treebank for an Old Italian text, specifically the *Divine Comedy*, was undertaken with the aim of creating a syntactic resource capable of interacting with state-of-the-art initiatives such as UD<sup>58</sup>. These initiatives not only foster the dissemination and accessibility of the resource, encouraging its use and creating opportunities for its presentation, but also enhance its scholarly value through collective research activities that promote dialogue within the community, as well as through the development of tools that facilitate study and enable both intra- and cross-linguistic comparison. One of the underlying goals of publishing such a resource within the domains of historical linguistics and the history of the Italian language is to bring historical-linguistic research closer to the computational field, through a methodology that is reproducible, open, and shareable. The use of computational tools, despite their initial complexity, facilitates the dissemination and sharing of data, thereby promoting transparency and advancing research. These tools, however, cannot replace human expertise, which remains essential for guiding the analysis and critically interpreting the results. The synergy between computational skills, which enable extensive and systematic analyses and comparisons, and the critical acumen and humanistic knowledge of the scholar, who is responsible for overseeing and ensuring the appropriate use of analytical tools, offers a new perspective and an enriching methodology for the study of ancient texts through modern approaches.

The *Italian-Old* resource, whose construction is described in this Chapter, is freely available and accessible to users<sup>59</sup>. Revisions and possible corrections are continuously updated in line with UD releases and the ongoing maintenance of the resource.

### 2.1 Annotating the *Divine Comedy*

Before detailing the development of the resource, it is important to justify the choice of the text, Dante Alighieri's *Divine Comedy* (see Section 1.2). Why the *Divine Comedy*?

---

<sup>58</sup> A special mention should also be made of the UniDive COST Action project (UniDive COST-Action: <https://www.cost.eu/actions/CA21167/>), whose activities and dissemination efforts have supported the present work. Any remaining errors are solely the responsibility of the author. The UniDive (Universality, diversity and idiosyncrasy in language technology) COST-Action CA21167 tackles the challenge of combining technological progress in NLP with the preservation of linguistic diversity, understood both across and within languages. It develops universal methodologies while quantifying and promoting diversity, resulting in resources and tools that cover a wide range of linguistic phenomena. This approach advances the understanding of language universals and supports many languages, including low-resourced and endangered ones.

<sup>59</sup> [https://github.com/UniversalDependencies/UD\\_Italian-Old](https://github.com/UniversalDependencies/UD_Italian-Old). The data on which this study was based were released with the UD release on 2025-11-15.

The motivations are twofold: one historical-linguistic and the other practical, as is fitting for a work situated at the intersection of historical and computational linguistics.

From a historical-linguistic perspective, the *Divine Comedy* is universally recognized as one of the cornerstones of Italian literature. Still studied in schools, though differently judged over the course of literary history<sup>60</sup>, it remains an indispensable cultural reference, with an influence that continues to this day, extending from theatrical performances, films, music albums to video games, and other cultural domains. Yet its literary impact is not its sole merit: its significance is equally crucial from a linguistic standpoint (refer to Section 1.1). Dante's decision to compose the *Divine Comedy* in the *locutio vulgaris*, a language grounded in the Florentine vernacular<sup>61</sup>, represents a turning point in the history of Italian literature and language, fostering both the diffusion and the recognition of the vernacular as a legitimate literary medium (Nencioni 1989a; Baldelli 1965; 1996; Manni 2013; Frosini 2020, among others). The undeniable literary and linguistic importance of Dante's work has fostered the proliferation of online resources for its consultation, among which the already mentioned DanteSearch (Section 1.1.1). It is within this context that the second motivation, the practical one, comes into play. The existence of an online resource such as DanteSearch has enabled the development of *Italian-Old*, which directly builds upon it (more details in Section 2.1.2). The possibility, and indeed the necessity, of reusing existing resources and data is fully aligned with the FAIR principles<sup>62</sup> for scientific data. These principles (Findability, Accessibility, Interoperability, and Reuse of digital assets) emphasize the importance of making data machine-actionable, that is, easily discoverable, accessible, and reusable by computational systems, as researchers increasingly rely on automated support to cope with the growing volume, complexity, and speed of data production. Such an approach promotes the sharing, reproducibility, and reuse of scientific data, thereby laying the groundwork for the advancement of knowledge. This is precisely the case with DanteSearch: by publishing XML data, it has made the development of *Italian-Old* possible, contributing both to the accessibility of information and to the expansion of knowledge.

Finally, it is worth mentioning that Universal Dependencies already includes a treebank containing part of Dante's Latin works, namely UDante<sup>63</sup> (Cecchini et al., 2020; Passarotti et al., 2021). The inclusion in UD of a treebank by the same author based on a different language, namely the Florentine

---

<sup>60</sup> See, for instance, the opinion of Croce; Contini (1970).

<sup>61</sup> It is worth noting that the *locutio vulgaris* of the *Divine Comedy*, grounded in Florentine, also draws on other dialectal varieties from the Tuscan area and beyond, thereby contributing to its plurilinguism (Serianni, 2021: 119-131). The richness of the language of the *Divine Comedy* has been extensively discussed in the scholarly literature, which is only briefly and non-exhaustively referred to here (Nencioni, 1989b; Parodi, 1896; Baldelli, 1996; Serianni, 2007, 2021; Tomasin, 2013; Frosini, 2017, among others).

<sup>62</sup> <https://www.go-fair.org/fair-principles/>

<sup>63</sup> [https://github.com/UniversalDependencies/UD\\_Latin-UDante](https://github.com/UniversalDependencies/UD_Latin-UDante)

vernacular (see footnote 60), would ensure continuity and further facilitate comparative studies within a single authorial corpus.

For the reasons outlined above, the *Divine Comedy* was selected and holds the distinction of being the first Old Italian text to enter the UD community. The prospect, and indeed the aspiration, is that, over time, with the growing involvement of scholars eager to contribute to such a project, coverage will be extended to other authors and texts, both contemporary with Dante and from later periods. In this way, the corpus of available material will be broadened, enabling more accurate linguistic investigations and further research.

### 2.1.1 Old Italian in Universal Dependencies and the ISO code

As already noted, the *Italian-Old* treebank is the first to include a text in Old Italian within UD, in contrast to the other Italian treebanks, which cover Modern Italian (Section 1.4.5). Despite the diachronic difference, *Italian-Old* is still listed among the Modern Italian treebanks, as shown on the UD homepage and illustrated in Image 7.

Italian		10	1,001K	IE, Romance
<b>Italian treebanks</b>				
ISDT	298K	UD@	W	★★★★★
VIT	280K	UD		★★★★★
Old	122K	UD		★★★★★
TWITTIRO	29K	UD		★★★★★
Valico	6K	UD		★★★★★
PoSTWITA	124K	UD		★★★★★
ParlaMint	20K	UD		★★★★★
ParTUT	55K	UD	W	★★★★★
MarkIT	40K	UD		★★★★★
PUD	23K	UD	W	★★★★★

See [here](#) for comparative statistics of Italian treebanks.

**Image 7.** Italian treebanks in UD

The decision to classify *Italian-Old* among the Modern Italian treebanks stems from the absence of a specific ISO code<sup>64</sup> designating Old Italian as an independent language, and thus not distinguishing it from Modern Italian. This lack of a dedicated code has led to the inclusion of the treebank containing the *Divine Comedy* among Modern Italian treebanks, while at the same time making it necessary to emphasize the distinction through the chosen name, *Italian-Old*, in order to clearly mark the difference from the others.

<sup>64</sup> The *International Organization for Standardization* (ISO) is the world’s leading body for the development of technical standards. Its role is to harmonize national standards in areas such as technical procedures, producing internationally recognized ISO standards. ISO codes also exist for languages, supporting their classification through the ISO 639 standard and creating a universally recognized catalogue of conventional codes that can be unambiguously associated with individual languages (<https://www.iso.org/iso-639-language-code>).

On the one hand, the need for a specific ISO code for Old Italian, so that it can be clearly distinguished from data documenting Modern Italian, may appear self-evident for practical purposes of classification and separation of linguistic data. Indeed, other languages already possess ISO codes that differentiates the currently spoken variety from its historical counterpart, such as Old French, Old English, and Old High German, among others<sup>65</sup>. On the other hand, however, this issue raises complex questions concerning the very nature of the Italian language and its historical development. Whether Old Italian should be considered distinct from Modern Italian, or whether it is preferable to preserve cohesion by signaling continuity without modification, remains an open question, on which contrasting views have been expressed<sup>66</sup>, and one to which this work cannot, of course, provide a definitive answer (this topic has also been addressed in Section 1.2). It is nevertheless worth noting that among the texts and arguments in favor of distinguishing Old Italian from Modern Italian (Dardano and Frenguelli 2004), syntax plays a central role (Tesi 2004). More fine-grained studies in this direction, though not limited to it, may eventually help provide a systematic answer to this question, which lies at the heart of a debate that is not only practical but also conceptual. Adding further complexity to the picture outlined so far, it should be noted that the question of ISO code also extends to the broader field of contemporary Italian, including the rich domain of Italian dialectology (Rohlf 1972; Pellegrini 1964). This, in turn, raises reflections on the possibility of establishing specific ISO codes for the various regional dialects, which are both highly diverse and extensively studied<sup>67</sup>.

In conclusion, while NLP may require distinctions to account for both contemporary and historical linguistic varieties, the evolution of Italian, shaped on the one hand by its much-debated diachronic development (see Section 1.2) and on the other enriched by its contemporary diversity, renders such a decision particularly challenging, calling for careful consideration at the community level.

### 2.1.2 Conversion

As mentioned in Chapter 1, *Italian-Old* builds upon the pre-existing DanteSearch corpus (Tavoni 2011), which provides grammatical annotation of various works by Dante Alighieri, both in Latin

---

<sup>65</sup> [https://www.loc.gov/standards/iso639-2/php/code\\_list.php](https://www.loc.gov/standards/iso639-2/php/code_list.php)

<sup>66</sup> Ascoli (1882–1885, p. 124; cf. Tomasin 2019, ch. VI) argued that Old and Modern Italian display considerable similarities, likely highlighted by comparison with the evolution of other Romance languages, such as French, where the gap between the old and modern varieties is much more evident (Dardano 2012). Yet many studies now stress the distinction between Old and Modern Italian (Dardano and Frenguelli 2004), most notably in syntax (Tesi 2004).

<sup>67</sup> It should be specified that certain varieties already have a dedicated ISO code, such as Neapolitan, Sicilian, and Sardinian ([https://www.loc.gov/standards/iso639-2/php/code\\_list.php](https://www.loc.gov/standards/iso639-2/php/code_list.php)).

and in the vernacular, made available in XML-TEI format<sup>68</sup>. With regard to the *Divine Comedy*, the lemmatization and the morphological annotation were carried out by Enrica Calchini, Sara Gigli, Clara Corallo, and Carla Moretti (2000)<sup>69</sup>, while the syntactic annotation was produced by Sara Gigli (2004). The reference edition of the *Divine Comedy* used is Petrocchi (1994). It is important to note that, during the conversion, the original DanteSearch tagset was preserved whenever possible and encoded in the CoNLL-U file within the `xpos` field (Section 1.4.6).

The following paragraphs describe the DanteSearch tagset (Section 2.1.2.1) and the conversion process (Section 2.1.2.2).

### 2.1.2.1 DanteSearch tagset

The annotation in DanteSearch includes tokenization, lemmatization, and morphological analysis of each token in Dante's works, with the numerical position of each word in the text also recorded. Tokenization additionally involves the splitting of multi-words, such as articulated prepositions or univerbations. For example, in the case of *col* 'with the', a single word is represented by two tokens (*con* 'with' and *il* 'the') each carrying its own morphological annotation but sharing the same positional index. The grammatical annotation in DanteSearch includes word-by-word tagging that indicates the part of speech and the specific morphological features of each token, encoded as an alphanumeric string. Each morphological feature is represented by a specific number or letter, and the morphological information of a token is derived from the combination of these alphanumeric values. For instance, the noun *luna* 'moon' is annotated with the tag `sflfs`, where `s` = *sostantivo* (noun); `f` = feminine gender of the lemma, `l` = first declension, `f` = feminine gender of the form, and `s` = singular number, yielding `sflfs`.

For the *Divine Comedy*, an additional layer of syntactic annotation is available, produced by Sara Gigli (2004), which is based on clause-level analysis. In this annotation, clauses are labeled with tags such as `causal`, `relative`, `temporal`, `consecutive`, among others. This annotation, however, is not compatible with the UD framework, which operates at the word level and assigns the syntactic functions of each token (e.g., `obj` 'object', `nsubj` 'subject', `obl` 'oblique').

As mentioned also in 1.1.1, the resource and its annotations can be consulted online through a graphical interface<sup>70</sup> that allows for specific queries and provides access to the XML files<sup>71</sup>. In the case of the *Divine Comedy*, two XML files are available: one containing the morphological annotation

---

<sup>68</sup> The XML-TEI format used by DanteSearch follows the guidelines of the Text Encoding Initiative (TEI), an international consortium responsible for developing and maintaining standards for the digital representation of texts. For further information, see <https://tei-c.org>.

<sup>69</sup> <https://dantesearch.dantenetwork.it>

<sup>70</sup> <https://dantesearch.dantenetwork.it>

<sup>71</sup> The XML files are available in the download section.

and another containing the syntactic structure of the text at the clause level. However, since the clause-level syntactic information is not compatible with UD, only the XML file with the grammatical annotation was used in the conversion process.

### 2.1.2.2 Conversion process

The XML grammatical file of the *Divine Comedy* served as the starting point for the conversion process, which involved reusing its tokenization, lemmatization, and morphological annotation, while adapting and converting the DanteSearch formalism to the UD standard. The conversion required not only an adaptation of the tagset but also a change of format, from the original XML-TEI of DanteSearch to the CoNLL-U format of Universal Dependencies (Section 4.1.6).

#### 2.1.2.2.1 Tokenization, sentence splitting and lemmatization

Concerning tokenization, no major changes were made, since articulated prepositions and multi-words were already treated as two tokens in DanteSearch, and this division was therefore maintained in the CoNLL-U format, corresponding to a split as shown in Image 8.

```

1-2   Nel           _           _           _           _           _           _           Canto=1|Comment=ne-1|Verso=1
1     in           in           _           _           _           _           _           Canto=1|Comment=ne-1|Verso=1
2     il           il           _           _           Definite=Def|Gender=Masc|Number=Sing|PronType=Art _           _
      _           Canto=1|Comment=ne-1|Verso=1

```

**Image 8.** articulated prepositions in CoNLL-U format

The range (1-2) preserves the surface form of articulated prepositions (e.g., *nel* ‘in the’) as they appear in the text, while in the split tokens the forms of two components are reconstructed at the lemma level (e.g., *in* ‘in’ and *il* ‘the’). In the `Misc` field, a comment is provided with the artificial segmentation of the form, which, however, remains purely notional (e.g., `Comment=ne-1`).

An exception was made for cases of epithesis, namely the addition of non-etymological elements introduced for phonetic reasons (Garavelli 2003: 122). In these instances, the grammatical annotation of DanteSearch does not follow the Petrocchi edition<sup>72</sup>, treating the epithesis as two separate words (*più* ‘more’ and *e*) and consequently as two distinct tokens. In the CoNLL-U format, however, they were unified, in accordance with Petrocchi (1994), and treated as a single token, since the vowel *e* carries no independent semantic value. To keep track of this phenomenon, the feature `Epithesis=Yes` was added in the `Misc` field (Image 9).

```

9     piùe      più     ADV     b           _           _           _           _           Canto=8|Epithesis=Yes|Verso=46

```

<sup>72</sup> By contrast, the syntactic XML file reproduces the text according to its correct reading with epithesis.

### Image 9. epithesis in CoNLL-U format

Another exception concerns the treatment of proper names composed of more than one token (such as Vanni Fucci), which in DanteSearch were annotated as single token but, in accordance with UD guidelines, have been split. For instance, what appears as a single token in DanteSearch, lemmatized as vanni fucci, is represented in CoNLL-U as two distinct token (Image 10):

```
18      Vanni  Vanni  PROPN  np      _      _      _      _      Canto=24|Verso=125
19      Fucci  Fucci  PROPN  np      _      _      _      _      Canto=24|Verso=125
```

### Image 10. composed proper noun in CoNLL-U format

Similarly to proper noun, locutions, namely sequences of two or more words occurring in a fixed order (Serianni & Castelvechi 1991: 491), were not separated but treated as single tokens in DanteSearch. In the CoNLL-U format, by contrast, they were split and handled independently as distinct token, with each assigned its own token-specific analysis (Image 11).

```
8      prima  prima  ADV    _      _      _      _      _      Canto=8|Verso=1
9      che   che   SCONJ  _      _      _      _      _      Canto=8|Verso=2
```

### Image 11. locutions in CoNLL-U format

Regarding sentence splitting, it was applied only in the presence of a full stop (.), in order to respect Petrocchi's editorial choices without further intervention in this direction. Consequently, even cases of direct speech containing a period were split.

Finally, lemmatization was preserved in line with DanteSearch to ensure consistency with the original resource. Adjustments were made only where the tokenization differed, as in the case of proper names and locutions, where the lemma was assigned to each individual token (see Images 10 and 11).

#### 2.1.2.2.2 Morphological annotation

The alphanumeric morphological tagset of DanteSearch was converted into the UD formalism, which in some cases required specific adaptations. To this end, conversion tables were created to ensure that each alphanumeric value of the DanteSearch tagset corresponded, where necessary, to one or more UD values. Dedicated conversion tables were developed for each part of speech. An example of such a conversion table for nouns is provided in Image 12.

A	B	C	D
DanteSearch			UD
Parte del discorso:	sostantivo volgare	s	NOUN
Genere lemma:			
	Maschile	m	
	Femminile	f	
Declinazione:			
	1ª declinazione	1	
	2ª declinazione	2	
	3ª declinazione	3	
Genere forma:			
	Maschile	m	Gender= Masc
	Femminile	f	Gender= Fem
Numero:			
	Singolare	s	Number=Sing
	Plurale	p	Number=Plur
Other:	In locuzione	l	vedi lista locuzioni

**Image 12.** Conversion table for nouns

As shown in Image 12, the tables were structured with two main entries, DanteSearch (henceforth DS) and UD, each containing the morphological information encoded in its respective terminology. DS values are distributed across three columns, labeled A, B and C (for illustrative purposes), while Column D contains the output consistent with the UD formalism, which constitutes the information encoded in the final CoNLL-U file. Columns A and B of DS contain explicit grammatical information, which is rendered in alphanumeric form in Column C, where the DS code is split vertically, with each symbol placed on a separate row so that it can be converted individually. More specifically, Column A indicates a general grammatical property (e.g., *parte del discorso* ‘part of speech’, *genere del lemma* ‘lemma gender’, *declinazione* ‘declension’, and so on), while Column B lists the possible values for that property. For example, for gender the possible values are masculine and feminine, while for declension the values are first, second, and third. Since each part of speech has its own dedicated table, the property part of speech is assigned only one fixed value. The values listed in Column B and associated with a single property in Column A, referred to as a block (e.g., value ‘feminine’ and ‘masculine’ in B are associate with the property ‘lemma gender’ in Column A, forming one block) are mutually exclusive, meaning that they cannot occur simultaneously: only one value can apply (either feminine or masculine). Each block corresponds to a fixed position in the DS alphanumeric code. More specifically, the ‘part of speech’ occupies the first position, the ‘lemma gender’ the second, and so on.

To provide a concrete example, the noun *luna* ‘moon’, annotated with the DS tag *s f 1 f s* (see Section 2.1.2.1), yields the scheme shown in Image 13, with the corresponding tagset values highlighted in yellow.

A	B	C	D
DanteSearch			UD
Parte del discorso:	sostantivo volgare	s	NOUN
Genere lemma:			
	Maschile	m	
	Femminile	f	
Declinazione:			
	1ª declinazione	1	
	2ª declinazione	2	
	3ª declinazione	3	
Genere forma:			
	Maschile	m	Gender= Masc
	Femminile	f	Gender= Fem
Numero:			
	Singolare	s	Number=Sing
	Plurale	p	Number=Plur
Other:	In locuzione	l	vedi lista locuzioni

**Image 13.** Example of conversion table for the noun *luna* ‘moon’

The first position of the DS tagset, defined as *parte del discorso* ‘part of speech’, corresponds to what in UD is referred to as the universal part of speech (UPOS) and is therefore entered in the fourth field of the CoNLL-U file (Section 1.4.6). Accordingly, a DS tagset classified as *sostantivo volgare* ‘vernacular noun’ and marked with the code *s* in first position (as the case of the noun *luna* ‘moon’) is rendered in UD as *NOUN* in the fourth field of the CoNLL-U. The subsequent alphanumeric values of the DS tagset, starting from the second position, correspond to what UD defines as morphological features, which are classified under the *Feats* field and recorded in the sixth field.

As shown in Images 12 and 13, not all blocks have a corresponding feature in UD. For example, information concerning lemma gender (*genere lemma*) and declension (*declinazione*) is not represented in the UD formalism. In such cases, the corresponding cells in the UD column are left blank. When the morphological information is represented in UD, however, it is converted into the UD formalism. For instance, the feminine gender of the form, expressed in the fourth position as *f* in the DS tagset (e.g., *sf1fs*), is rendered as *Gender=Fem*, in accordance with the UD encoding conventions.

It should also be noted that the original information is not lost, as the DanteSearch tagset is preserved in its entirety in the fifth field of the CoNLL-U format, namely the *xpos* field.

Returning to the example of the conversion of *luna* ‘moon’, Image 14 illustrates the final output of the conversion from the DanteSearch tagset (*sf1fs*) into UD formalism:

```
1      luna      luna      NOUN      sf1fs      Gender=Fem|Number=Sing      _ _ _ Canto=1|SpaceAfter=No|Verso=115
```

**Image 14.** The output of conversion into UD formalism of the term *luna* ‘moon’

### 2.1.2.2.3 Syntactic annotation

As specified in Section 2.1.2.1, the syntactic annotation available for the *Divine Comedy* (Gigli 2004) is clause-based, indicating the function of the clause (e.g., causal, consecutive, among others). This level of annotation, however, is not compatible with UD syntax, which requires annotation at the token level (e.g., object, subject, among others). Consequently, unlike lemmatization and morphology, this annotation could not be obtained through conversion but had to be carried out manually, as described in detail in Section 2.1.3.

#### 2.1.2.2.4 Miscellaneum

An addition made with respect to the XML of DanteSearch (although already available through the online interface) is the indication of each token’s position within the work, namely its *Canto* and line number. This information was included in the `MISC` field (Miscellaneum), an optional field for comments, where different types of information can be recorded. In the case of *Italian-Old*, in addition to being used for annotations considered useful, such as instances of epithesis or the reconstruction of artificially split articulated prepositions (see Section 2.1.2.2.1), given the importance of the division into *Canti* and the versification of the work, it also records the corresponding *Canto* and *verso* ‘line’ for each token (see Image 15).

```
3      Beatrice      Beatrice      PROPn  np      _      _      _      _      Canto=2|Verso=70
```

**Image 15.** `MISC` in CoNLL-U format

#### 2.1.3 Syntactic annotation process

Whereas lemmatization and morphological parts of speech and features were derived through semi-automatic conversion, syntactic annotation was carried out manually from scratch. Annotation was performed entirely by hand for the first *Cantica*, the *Inferno*, both to ensure accuracy and to allow the annotator to familiarize themselves with the annotation process without the potential bias of pre-annotated text. It should also be recalled that, since *Italian-Old* is the first treebank containing Old Italian data, no gold-annotated resources were available to serve as training data.

Concerning size, the *Inferno* comprises 41,369 tokens (including punctuation), amounting to a total of 1,228 sentences.

Since the process of manual annotation proved highly time-consuming (approximately one year), for the subsequent *Cantiche*, *Purgatorio* and *Paradiso*, parsers were trained on the gold-annotated portion of the *Commedia* itself in order to accelerate the annotation process (see Sections 2.1.3.1 and 2.1.3.2), with human intervention limited to correction. This approach also provided an opportunity to evaluate parser performance on gold-annotated data, with the expectation that high scores in

syntactic metrics would demonstrate effectiveness and thereby reduce the amount of manual work required. The experiments conducted are presented in the following sections.

### 2.1.3.1 Testing inside the *Inferno*: road to *Purgatorio*

Once the syntactic annotation of *Inferno* was completed, an experiment was conducted to evaluate the performance of a dependency parser in order to facilitate the annotation of the second *Cantica*, *Purgatorio*. The goal was to employ the parser for pre-parsing, thereby reducing the effort required for subsequent manual revision.

As a preliminary experiment (Corbetta et al. 2023), UDPipe1<sup>73</sup> (Straka et al. 2016; Straka et al. 2017), a trainable pipeline for tokenization, tagging, lemmatization and dependency parsing, was applied to conduct in-domain evaluations on a portion of the gold data of *Inferno*. In the experiment, a 10%/90% cross-validation was applied to *Inferno*, testing both random and consecutive-*Canti* partitions, and two scenarios were considered: +MORPH, where lemmas, parts of speech, and morphological features were provided and the parser was required to predict only syntax; and -MORPH, where the model had to predict all information (both lemmatization and morpho-syntactic annotation) from scratch. The first scenario (+MORPH) was designed to obtain results applicable to the other two *Cantiche* (*Purgatorio* and *Paradiso*), for which morphological annotation and lemmatization were already available from the DanteSearch conversion (Section 2.1.2). The second scenario (-MORPH), by contrast, was intended to assess potential performance gaps when applying the model to texts lacking pre-existing morphological and lemmatization layers.

Results presented in Corbetta et al. (2023), and reproduced here in Image 16 for clarity, show that the two main syntactic metrics, Label Attachment Score (LAS) and Unlabeled Attachment Score (UAS)<sup>74</sup>, are overall high, averaging 81.89 UAS and 77.08 LAS in the +MORPH scenario, and 75.11 UAS and 67.84 LAS in the -MORPH scenario. In particular, the +MORPH scenario showed a clear improvement over the -MORPH one, with gains of about 6–9% in both LAS and UAS, confirming that prior morphological annotation significantly enhances parsing accuracy. These results support positive expectations for the parsing of the remaining *Cantiche*, where, as mentioned, morphosyntactic information is already available from the DanteSearch conversion.

---

<sup>73</sup> <https://github.com/ufal/udpipe>

<sup>74</sup> Label Attachment Score (LAS) measures the proportion of tokens for which the system correctly predicts both the head and the dependency label, while Unlabeled Attachment Score (UAS) evaluates only the correctness of the head (Buchholz and Marsi 2006).

Partition	Scenario	Avg. UAS	Avg. LAS
random	+Morph	81,95±0,94%	77,07±1%
consecutive	+Morph	81,79±1,38%	77,09±1,34%
random	-Morph	75,32±0,91%	67,97±0,80%
consecutive	-Morph	74,90±1,37%	67,71±1,17%

**Image 16.** UAS and LAS for the parsing of *Inferno* (internal evaluation)

On the basis of these encouraging results, *Purgatorio* was pre-parsed using UDPipe1 trained on the *Inferno* gold data, after which it was manually reviewed and corrected. While the manual annotation of *Inferno* required an entire year of work, the revision of the *Purgatorio* took approximately seven months. *Purgatorio* comprises a total of 41,270 tokens (including punctuation) and 1,174 sentences.

### 2.1.3.2 Testing with different models: road to *Paradiso*

While manual checking the *Purgatorio*, further experiments (Corbetta et al. 2024a; Corbetta et al. 2024b) were carried out to assess whether the initial model, UDPipe1 (Straka et al. 2017), could be replaced by Stanza<sup>75</sup> (Qi et al. 2020), a neural network pipeline providing tokenization, tagging, lemmatization, and dependency parsing, and to evaluate the potential contribution of Modern Italian data, of which large amounts are available (Section 1.4.5), to parsing performance.

Specifically, Corbetta et al. (2024a) compared UDPipe1 and Stanza, both trained on in-domain data<sup>76</sup> and evaluated on a subset of the *Comedy*, the first three *Canti* of each *Cantica*, previously manually annotated, and found that Stanza outperforms UDPipe1 (Image 17).

	Inf		Purg		Par	
Metr.	UDP	Stan	UDP	Stan	UDP	Stan
UAS	82.65	87.73	77.93	82.67	78.50	84.50
LAS	77.87	84.02	71.42	77.31	71.33	78.85

**Image 17.** In-domain UAS and LAS across sampled *Canti* of the three *Cantiche*

As shown in Image 17, Stanza outperforms UDPipe1 by an average of approximately +5.90 points overall (+5.27 UAS; +6.52 LAS), thereby indicating its suitability for the pre-parsing of *Paradiso*. The same study (Corbetta et al. 2024a) also reports experiments testing whether models trained exclusively on larger datasets of Modern Italian, for which substantially more annotated data are available (Section 2.1.5), could yield better results than parsers trained on smaller but chronologically

<sup>75</sup> <https://stanfordnlp.github.io/stanza/index.html>.

<sup>76</sup> The training data consisted of *Canti* IV–XXXIV of *Inferno*, covering 30% of the entire *Divine Comedy* and 91% of *Inferno* (a total of 1,118 sentences and 37,806 syntactic words).

coeval datasets relative to the test set. For this experiment, Stanza models trained on the training sets of three major Modern Italian treebanks, ISDT (Bosco et al. 2013), VIT (Delmonte et al. 2007; Delmonte 2009; Tonelli et al. 2008), and ParTUT (Bosco et al. 2012), were evaluated on the same test set used for in-domain parsing (the first three *Canti* of each *Cantica*). The results are reported in Image 18.

	Inf	Purg	Par
UAS	69.05	66.28	67.74
LAS	56.14	53.30	54.31

**Image 18.** LAS and UAS scores with Stanza model trained on Modern Italian

As shown by the comparison between tables in Image 17 and 18, in-domain training (shown in Image 17) yields better performances. In this respect, the study confirmed that in-domain models, even when trained on limited data, outperform larger out-of-domain datasets from Modern Italian. This finding is consistent with previous literature (Mambrini and Passarotti 2012; Khan et al. 2013a,b) and further underscores the importance of maintaining in-domain training.

A latter attempt was made to challenge the performance of in-domain parser by creating a *combined* model, namely, a parser trained on Modern Italian data and in-domain data of the *Comedy* itself (Corbetta et al. 2024b). The combined model was obtained by merging portions<sup>77</sup> of two Modern Italian treebanks, VIT (Tonelli et al. 2008) and ISDT (Bosco et al. 2013), respectively, with gold-annotated data of the *Inferno* and *Purgatorio* (named OI in Images 19-21).

The evaluation<sup>78</sup> of these combined models (labeled Combi in Images 19-21) against non-combined ones, and tested on both Old and Modern Italian data, confirmed that the similarity between training and test data is more relevant than the size of the training set, since combined models, despite relying on larger training data, yielded lower performances than in-domain tests (Images 19, 20 and 21).

	VIT1 -> VIT1	OI -> OI	VIT1 -> OI	OI -> VIT1
LAS	71.60	75.86	42.83	68.53
UAS	77.70	82.24	56.13	75.53

**Image 19.** LAS and UAS with VIT1<sup>79</sup> and OI models (where "->" stands for "on" indicating Model on Test)

<sup>77</sup> The Modern Italian datasets were partitioned to match the size of the *Italian-Old* treebank at its current stage, namely 82,644 tokens (including punctuation) and 2,402 sentences, comprising *Inferno* and *Purgatorio*; for further details, see Corbetta et al. (2024b).

<sup>78</sup> The data of each model were partitioned into training, development, and test sets, following a 70%/15%/15% ratio, respectively.

<sup>79</sup> The number "1" after VIT1 refers to the specific partition used among those created during the internal tests on VIT (see *ibid.*).

	ISDT -> ISDT	OI -> OI	ISDT -> OI	OI -> ISDT
LAS	88.55	75.86	51.62	74.83
UAS	91.41	82.24	63.03	80.93

**Image 20.** LAS and UAS with ISDT and OI models

	CombiVIT -> VIT	CombiVIT -> OI	CombiISDT -> ISDT	CombiISDT -> OI
LAS	69.11	55.37	87.76	75.18
UAS	74.96	64.86	90.85	81.98

**Image 21.** LAS and UAS with combined models

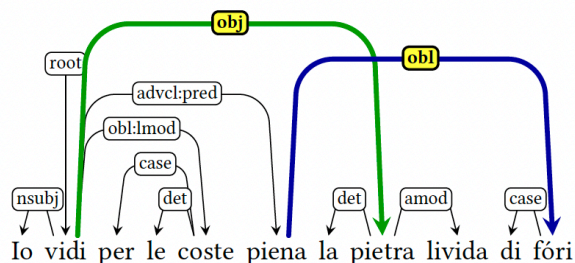
As shown by the comparison of Images 19, 20, and 21, the scores obtained by the combined models (CombiVIT and CombiISDT) are lower than those achieved by in-domain models (namely, VIT1 → VIT1; ISDT → ISDT; OI → OI), both when tested on Old Italian texts and on Modern Italian ones. Another noteworthy outcome of this experiment is that models trained on Old Italian data yielded higher scores on Modern Italian texts (OI → ISDT compared to ISDT → OI in Image 20, and OI → VIT compared to VIT → OI in Image 19) than the reverse.

These results prompted further analyses and comparisons between Modern and Old data, highlighting non-projectivity as the main factor of divergence. Non-projectivity refers to configurations in which a head and its dependents do not form a continuous constituent, but instead exhibit discontinuous relations. An example of non-projectivity taken from Corbetta et al. (2024b) is shown in Example 1 and Image 22:

1) *Inf.* XIX, vv. 13-14:

*Io vidi per le coste (...) piena la pietra livida di fòri*

‘Along the sides (...) I saw that livid rock was perforated.’



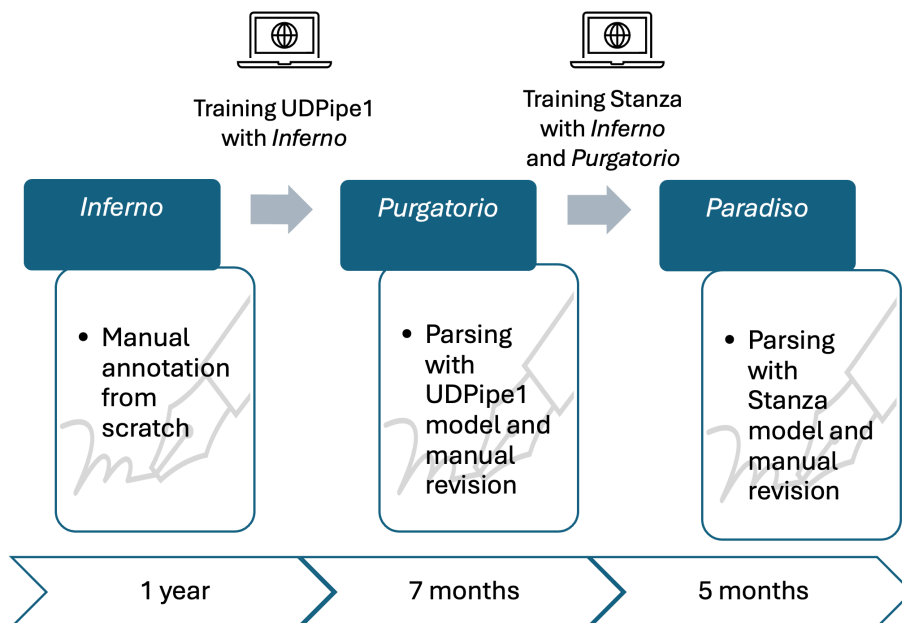
**Image 22.** Non-projectivity in *Inf.* XIX, vv. 13-14

Example 1 presents a sentence from *Inferno* XIX, describing the third *bolgia* where the ground displays holes on the walls of the rock. Non-projectivity occurs between *piena* ‘full’ and *di fòri* ‘of holes’, since the two dependent phrases are separated by *la pietra livida* ‘the livid rock’. In the syntactic tree, this phenomenon is illustrated by the crossing arcs. In this case, non-projectivity

corresponds to the use of a specific rhetorical figure, namely hyperbaton, which operates on the linear order of constituents by altering it and separating elements that would normally be adjacent.

The findings and parsing results displayed in Images 19-21 suggest that a text with higher non-projectivity is more effective as training data for a text with lower non-projectivity, rather than the other way around.

To sum up, for all the reasons mentioned above, both the choice of parser and the nature of the training data led to the decision to pre-annotate *Paradiso* with a Stanza model<sup>80</sup> trained exclusively on in-domain gold data of *Inferno* and *Purgatorio*, in order to maximize parsing accuracy and accelerate the subsequent manual revision of *Paradiso*. The manual check of *Paradiso*, comprising 39,385 tokens (including punctuation) and 1,017 sentences, required approximately five months to complete. Image 23 summarizes the steps involved in the syntactic annotation process.



**Image 23.** Steps of syntactic annotation of the *Divine Comedy*

The following section discusses some challenges encountered during manual annotation.

## 2.2 Annotation issues

This section examines the main issues encountered during the annotation process and the solutions adopted. Particular attention is devoted to elliptical constructions, which, due to their complexity, have been selected as the central focus of this study; these will be discussed in detail in Chapters 3, 4, and 5. The remaining issues are presented by linguistic phenomenon: Section 2.2.1 briefly

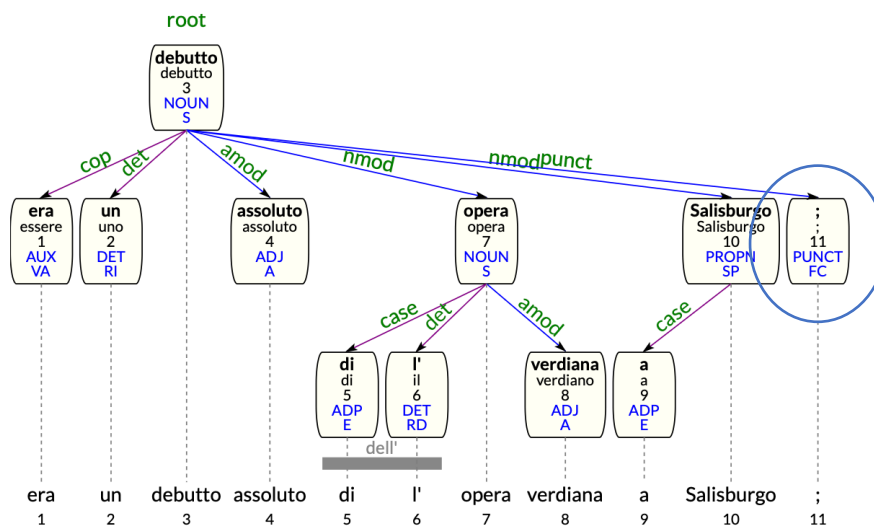
<sup>80</sup> Special thanks are due to Giovanni Moretti, who was responsible for the development of the Stanza model.

addresses punctuation marks, with attention to the treatment of the semicolon; and Section 2.2.2 comments on annotation choices regarding comparative clauses.

### 2.2.1 Punctuation marks: the semicolon

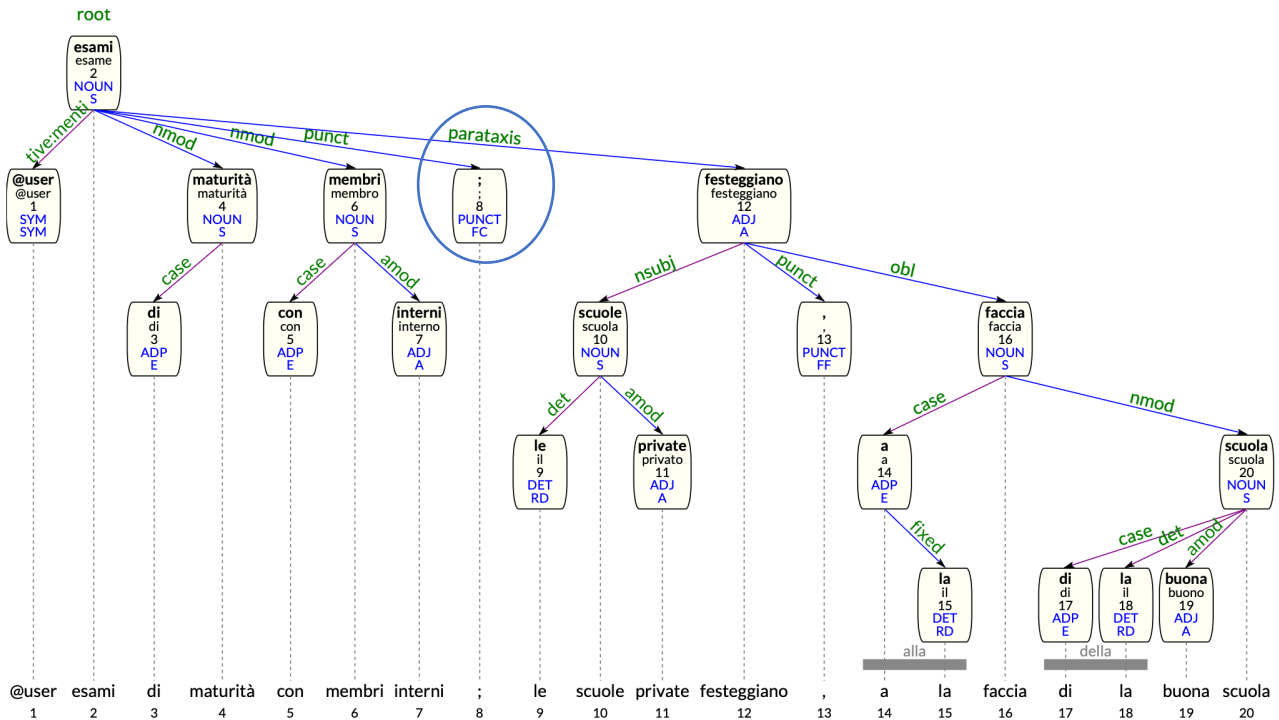
Punctuation marks are an inherent feature of written text. Although their treatment as tokens does not affect the syntactic structure, since they are assigned the dependency relation `punct` (punctuation), the connections they establish between clauses remain a matter of interpretation.

In some Italian treebanks (including ISDT and VIT), strong punctuation marks such as the semicolon are regarded as markers of segmentation and therefore correspond to sentence splitting. Image 24 illustrates a sentence from ISDT in which the semicolon, circled in the syntactic tree to facilitate identification, is treated as a sentence boundary (ISDT, sent: `isst_tanl-1772`: *era un debutto assoluto dell'opera verdiana a Salisburgo*; lit. ‘It was the absolute premiere of the Verdi opera in Salzburg;’).



**Image 24.** Example of semicolon in ISDT (sent: `isst_tanl-1772`)

By contrast, other Italian treebanks, such as PoSTWITA and Valico, do not treat the semicolon as a sentence boundary, but rather interpret the relation that holds between the connected clauses, as shown in Image 25 (TWITTIRO, sent: `train_412`: *@user esami di maturità con membri interni; le scuole private festeggiano, alla faccia della buona scuola* ‘@user final high-school exams with internal members; private schools are celebrating so much for the good school’).



**Image 25.** Example of semicolon in TWITTIRO (sent: train\_412)

Image 24 shows the syntactic tree in which the semicolon is not treated as a sentence boundary. Instead, the two clauses are linked by the *parataxis* relation, used for side-by-side sentences, that is, “employed for a pair of what could have been standalone sentences, but which are treated together as a single sentence.”<sup>81</sup>

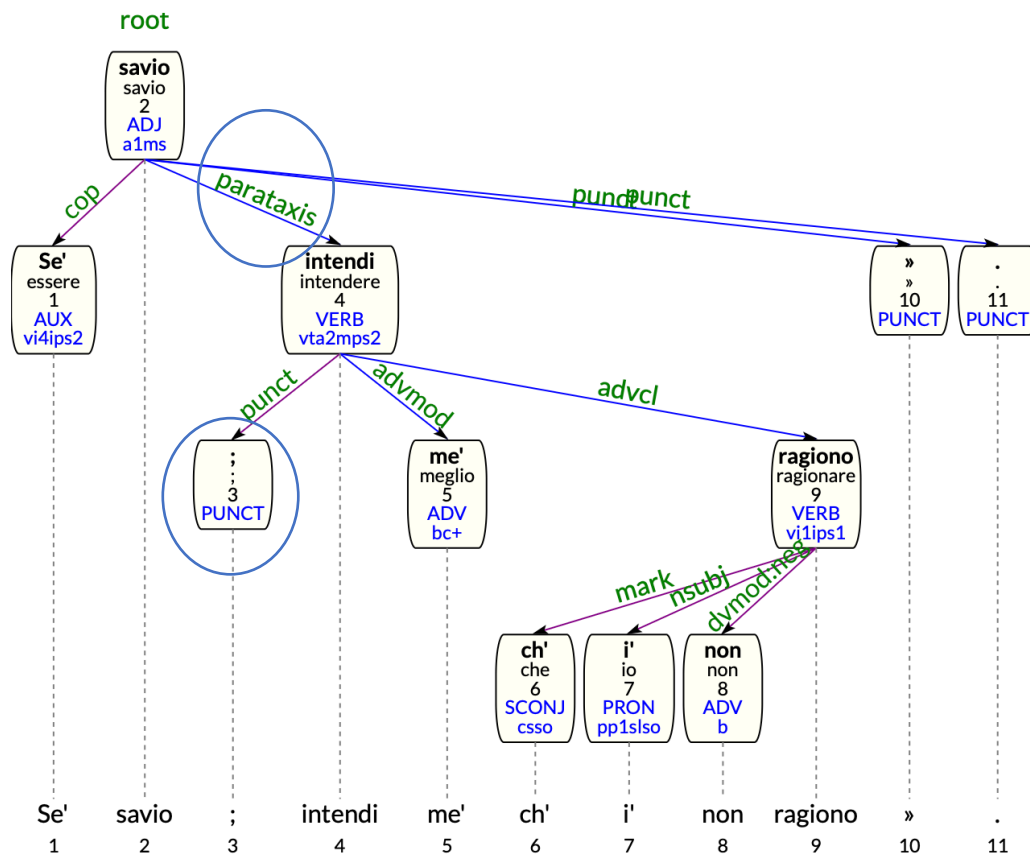
This heterogeneous treatment of punctuation highlights the variability and context-dependent value that such punctuation marks may assume. While there is general awareness of the strong break introduced by the semicolon, its function, whether as a true boundary marker (and thus interpreted as a sentential *caesura*) or merely as a weaker separator (therefore introducing what is commonly described as *parataxis*), is ultimately determined by the context in which it appears.

For *Italian-Old*, the decision was to remain as faithful as possible to Petrocchi’s editorial choices, without in any way altering the work carried out by the editor. This approach, motivated also by an awareness of the critical value of Petrocchi’s edition, directly influenced the decision to preserve his sentence-splitting choices unchanged. It is worth recalling that, since the *Divina Commedia* is a manuscript dating from the late 13th to early 14th century and originally written in *scriptio continua*, punctuation was not part of the text. The punctuation is therefore entirely the work of Petrocchi, who applied his own criteria (Petrocchi 1994: 472). According to Petrocchi, the semicolon “separa sia pur lievemente fatti distinti o susseguenti cronologicamente” (is used to

<sup>81</sup> <https://universaldependencies.org/u/dep/parataxis.html>

separate, albeit slightly, distinct facts or those following chronologically) (Petrocchi 1994: 472), in contrast to the colon, which is employed to connect cases in which “la seconda proposizione nasce come sviluppo o come esplicazione e precisazione della prima” (the second clause arises as a development, explanation, or specification of the first) (ibid.).

For annotation purposes and considering both the decision to adhere to the editorial choices and the need to maintain consistency with (at least some) Italian treebanks, the semicolon has been mapped to the dependency relation *parataxis*, avoiding its interpretation as a *caesura*. Image 26 illustrates a case of a semicolon in *Italian-Old*, taken from *Inferno* II, v. 36 (Italian-Old, sent: Dante\_Inferno-43: *Se' savio; intendi me' ch'i' non ragiono*». ‘You’re wise; you know far more than what I say’.’)



**Image 26.** Example of semicolon in *Italian-Old* (sent: Dante\_Inferno-43)

In general, as mentioned above, the punctuation of the *Divine Comedy* is not authorial but reflects the discretion of the critical editor, and may therefore vary depending on the reference edition<sup>82</sup>. This

<sup>82</sup> A particularly interesting case of how the editorial criteria affect the punctuation of the text is found in *Inferno* XXVIII vv. 14;18, where Petrocchi adopts a semicolon, whereas Inglese (2007) does not employ a semicolon and instead uses only a comma:

(a.) *per contastare a Ruberto Guiscardo; ... dove sanz'arme vinse il vecchio Alardo;*

divergence reinforces the importance of relying on a single reference edition, highlights how punctuation, introduced into a work originally devoid of it, may be subject to personal and potentially problematic interpretations, and opens up opportunities for comparative analyses of punctuation across different editions, which could be explored in future work.

### 2.2.2 Comparative clauses

Another annotation issue arose in the treatment of comparative clauses<sup>83</sup>, an area that has been the subject of long-standing debate within UD<sup>84</sup>. Specifically, comparative constructions in the *Divine Comedy* exhibit a wide variety of forms, ranging from fully developed sentences to extremely reduced ones, consisting of only a few words and often elliptical<sup>85</sup>. In light of this variability, such constructions have been annotated uniformly by labeling their heads with the clausal dependency relation `advcl` ('adverbial clause modifier'), further specified with the subtype `cmp` ('comparative'). The boundary between a comparative elliptical clause and a nominal phrase used comparatively is not always straightforward<sup>86</sup>. In this work, nominal phrases functioning as comparatives are also annotated as `advcl:cmp`. This decision was made to ensure uniformity in the annotation of comparative structures, which are pervasive in a poetic text such as the *Divine Comedy*, and to avoid conflating comparative nominal phrases with other, non-comparative nominal phrases. It should be noted that in UD v2.16 the subtype `:cmp` is formally accepted only for adverbial clauses and not for nominal phrases. To preserve comparative information, such cases are therefore analyzed as adverbial clauses with an elided verb and labeled `advcl:cmp`.

Example 2 illustrates a case of a comparative clause in the *Divine Comedy*, whereas Image 27 depicts its syntactic tree:

2) *Inf. V*, vv. 28-29:

*Io venni in loco d'ogne luce muto, che mugghia come fa mar per tempesta*

'I reached a place where every light is muted, which bellows like the sea beneath a tempest,'

---

(b.) *per contastare a Ruberto Guiscardo, ... dove sanz'arme vinse il vecchio Alardo,*  
'when they fought hard against Robert Guiscard; ... where old Alardo conquered without weapons;'

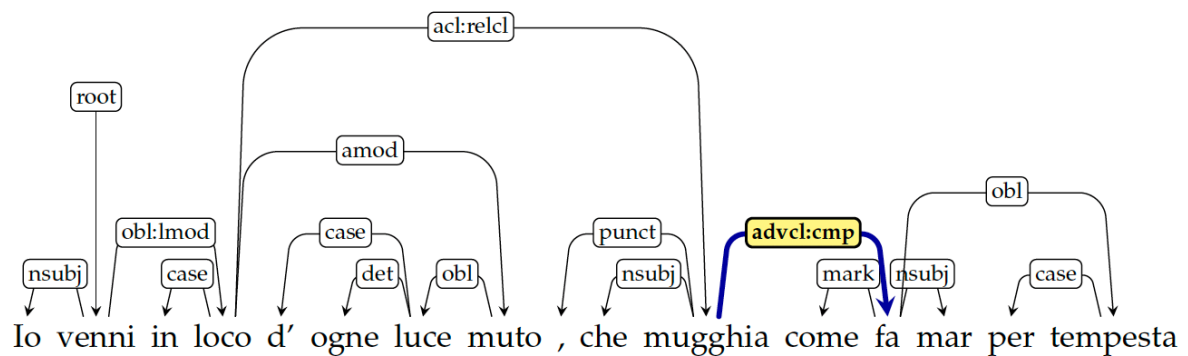
The different punctuation choices also give rise to different textual interpretations: in Petrocchi's edition they result in a more marked pause, whereas in Inglese's edition they suggest a coordinative reading.

<sup>83</sup> This section is an adaptation of the discussion presented in Corbetta et al. (2025 a: 38–41).

<sup>84</sup> <https://universaldependencies.org/u/overview/comparatives.html>

<sup>85</sup> For an analysis of comparative structures in the *Divine Comedy*, see D'Amico (2014: 322–328).

<sup>86</sup> For a discussion of comparative constructions in Old Italian, see Pelo (2012: 441–465). It is worth noting that, in Pelo (2012: 461), the construction *come colui che* 'like the one who' is treated as an elliptical comparative construction.



**Image 27.** Syntactic tree of *Inf.* V, vv. 28-29

The verses are taken from *Canto V*, where the souls of the lustful are punished for their sins. This is the famous *Canto* of the lovers Paolo and Francesca. Dante depicts the setting as devoid of light, *d'ogne luce muto* 'where every light is muted', employing synesthesia by associating vision (*luce*, 'light') with sound (*muto*, 'muted'). This imagery prepares the ground for the subsequent comparative description based on hearing. Indeed, the noise of the place is immediately likened to a stormy sea through the comparative structure *come fa mar per tempesta* 'like the sea beneath a tempest', lit. 'like the sea does in a storm'. The head of the comparative clause *come fa mar per tempesta* 'like the sea beneath a tempest' is the verb *fa* 'it makes', which depends on the verb of the relative clause (*acl:relcl*) *mugghia* 'bellows' as an *advcl*, functioning as an adverbial clause modifier with the specific subtype *:cmp*, thereby marking its comparative function. The remaining words in the clause are attached to the head of the comparative: *come* 'like' as the marker (*mark*), *mar* 'sea' as the subject (*nsubj*), and *tempesta* 'storm' as the oblique (*obl*), with the adposition *per* 'in' attached to *tempesta* as a case marker (*case*).

Alongside explicit cases of comparison, such as the one illustrated in Example 2, there are also instances where comparison is expressed through a nominal phrase with the predicate omitted. An instance of this is given in Example 3 (see Image 28 for the syntactic tree), which is reconstructed for exemplification in Example 4 (see Image 29 for the syntactic tree).

3) *Inf.* VI, v. 19:

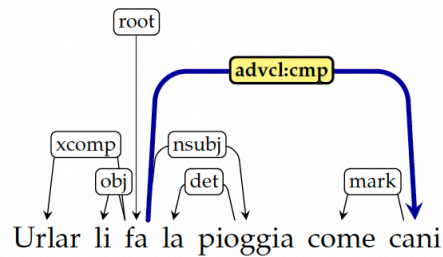
*Urlar li fa la pioggia come cani*

'That downpour makes the sinners howl like dogs'

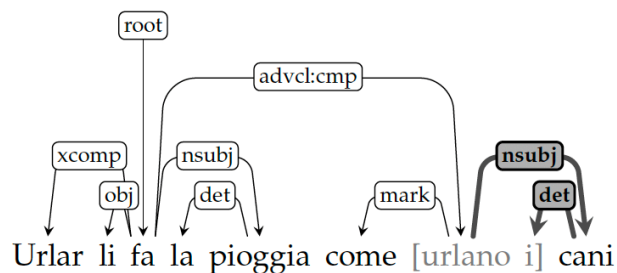
4) *Inf.* VI, v. 19 (reconstructed):

*Urlar li fa la pioggia come [urlano i] cani*

‘That downpour makes the sinners howl like dogs [howl]’



**Image 28.** Syntactic tree of sentence *Inf.* VI, v. 19



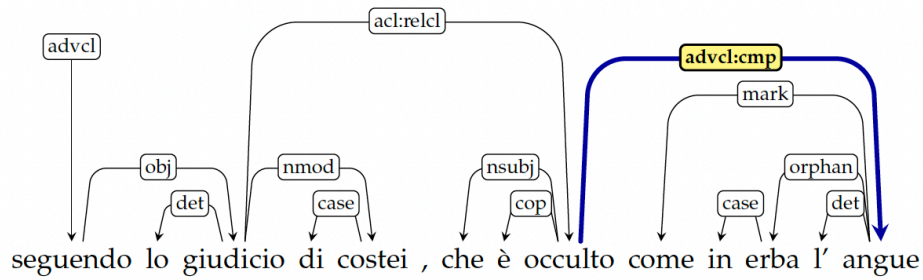
**Image 29.** Syntactic tree of the reconstructed sentence *Inf.* VI, v. 19

As shown in Image 28, the comparative phrase *come cani* has been annotated as an elliptical comparative clause, represented in Image 29. Although the reconstruction in Image 29 may appear debatable, the decision to treat *cani* ‘dogs’ as the head of an elliptical comparative clause was made in order to assimilate such structures to cases where the presence of a comparative predicate is more easily inferred, as in the example 5 and Image 30 reported below.

5) *Inf.* VII, vv. 83–84:

*seguendo lo giudicio di costei, che è occulto come in erba l' angue*

‘obeying the decision she has given, which, like a serpent in the grass, is hidden’



**Image 30.** Syntactic tree of the sentence *Inf.* VII, vv. 83–84

This sentence has been interpreted as containing a locative argument that depends on the elided predicate *è occulto* ‘is hidden’, a reading also reflected in the annotation through the use of the `orphan` relation (see Section 3.4.2.2). The decision to analyze the oblique phrase *in erba* ‘in the grass’ as a direct dependent of an omitted predicate, rather than as a nominal modifier (`nmod`) of the noun *anguè* ‘serpent’, is further supported by the fact that the verse is a Virgilian reference (Eclogues III, 93: *latet anguis in herba* ‘the snake lurks in the grass’), as noted by several commentators, including Chimenz (Alighieri 1962).

### 2.3 Possible future works

To conclude this chapter, which has focused on the construction of *Italian-Old* and the parsing experiments conducted so far to support and facilitate the creation of the resource, other research directions (those already initiated and those that could be undertaken) are briefly discussed in this section.

As mentioned in the Introduction, one of the expected outcomes of this work is the creation of an accessible resource that can serve as a starting point for advancing and disseminating research on Old Italian texts, specifically in dependency syntax. Although this work represents only a small contribution, the development of a parser trained on the gold data of the entire *Divine Comedy* constitutes a first step toward the annotation, still (always) requiring manual verification, of other texts. In Corbetta et al. (2025 c), an experiment conducted on texts by Guido Cavalcanti<sup>87</sup>, a contemporary author of Dante Alighieri best known for his *Rime*, for which a Stanza model was created by training it on all the *Divine Comedy*, yielded promising but also raised important issues,

<sup>87</sup> Guido Cavalcanti (c. 1255–1300) was a major poet of the late 13th century. His work played a crucial role in shaping the early Italian lyric tradition and the development of vernacular literature. For further discussion of Cavalcanti, see Marti (1972).

such as the choice of the critical edition and the need for accessible and shareable data<sup>88</sup>. The same experiment also examined whether neural parsers such as Stanza can still be regarded as more reliable than large language models such as ChatGPT, thereby raising questions about the choice of methodology<sup>89</sup>.

In sum, future research directions include the evaluation of different parser types and the portability of the available data, extending not only to other authors but also to different genres, with the aim of testing performance on prose texts as well.

## 2.4 Toward ellipsis: motivation and contextualization

This paragraph serves as a bridging section between two core components of the present work: the creation of a linguistic resource based on an Old Italian text (Chapter 2) and the analysis of ellipsis within that text (Chapters 3, 4, and 5). Although these two dimensions may initially appear distinct and divergent, a unifying thread is clearly present. This underlying connection, much like the analysis itself, emerged in a data-driven manner, arising from close and sustained engagement with the data.

While annotating the text, the challenge of dealing with a phenomenon such as ellipsis<sup>90</sup>, previously mentioned in Section 2.2, emerged clearly, highlighting an area within the Universal Dependencies framework that remains only partially explored, and thus warranting further consideration. Problematizing a modelling approach, such as the treatment of ellipsis in Universal Dependencies, necessarily requires an in-depth understanding of the phenomenon, achieved through careful examination of the existing literature. Engagement with the literature on ellipsis (Chapter 3) revealed an exceptionally rich and diverse field of analysis, addressed across multiple frameworks, highlighting both the breadth of its applications and the inherent difficulty of capturing a phenomenon that is, by its very nature, absent.

Given the context in which this research question arose, one of the aims of this work is to explore possible approaches to modelling ellipsis in Universal Dependencies, with the goal of facilitating its identification and systematic analysis. To this end, a specific case study is presented, consisting of an analysis of ellipsis in the *Italian-Old* treebank, with a particular focus on instances of predicate and object ellipsis.

---

<sup>88</sup> Recent critical editions are often subject to copyright restrictions that limit their use, forcing researchers to rely on outdated editions that do not reflect the most up-to-date critical and philological scholarship.

<sup>89</sup> The results suggest that Stanza remains preferable (Corbetta et al. 2025 b), although experiments with other tools such as UDify (Kondratyuk et al. 2019) and UDepLLaMA (Hromei et al. 2024), as well as with custom transformer-based pipelines using MaChAmp (van der Goot et al. 2021) and Trankit (Nguyen et al. 2021), are still needed.

<sup>90</sup> An in-depth definition of ellipsis is provided in Chapter 3.

This study aims to stimulate greater awareness of, and scholarly interest in, ellipsis, both within the debate on the modelling of linguistic data and with respect to the linguistic properties that ellipsis exhibits in the contexts in which it occurs. By presenting a specific case study based on annotated data, the study seeks to demonstrate, by way of illustration, that concrete and viable avenues for the investigation of this phenomenon with a data-driven approach do indeed exist.

Finally, before proceeding, it is useful to reiterate that, for the reasons outlined above, the analysis of ellipsis is restricted to a specific text type, namely poetic texts, which are regarded as less binding texts (Sabatini 1990). Such texts are characterized by the fact that the addresser does not require the addressee to adhere to a single, strictly determined interpretation of the intended meaning; as a result, the communicative interaction is less constrained than in other text types (Sabatini 1999:149). This communicative interplay between addresser and addressee relies on a range of linguistic mechanisms that, depending on the communicative purpose, allow the text to display varying degrees of rigidity and explicitness or, conversely, elasticity and implicitness according to the context (Sabatini 1999). Among these linguistic mechanisms, ellipsis can be characterized as increasing the degree of textual elasticity and implicitness by omitting part of the information (Sabatini 1999:155). While the textual function of ellipsis is not the primary focus of the present analysis, the observations reported here offer initial insights that may inform future investigations into this aspect of communicative interplay.

### 3) Ellipsis

The first question to address when discussing ellipsis is *What is ellipsis?*. Although this question may at first appear trivial, it is in fact far from simple, as demonstrated by the numerous studies and extensive debates that have engaged scholars on the topic.

One of the earliest attempts to define ellipsis comes from St. Isidore (Isidoro 2004). In his *Etymologiae* (XXXIV, 10), while describing grammatical faults (*vitia*) to be avoided in speech (*quae in eloquio cavere debemus*), he defines ellipsis as a deficiency in expression characterized by the omission of necessary words: *eclipsis est defectus dictionis, in quo necessaria verba desunt*. Isidore illustrates ellipsis with an example from Virgil's *Aeneid*, in a passage describing Dido wearing a cloak, carrying a golden quiver, and with her hair fastened by a gold pin<sup>91</sup>. The ellipsis occurs in the following verse (IV, 138):

*cui pharetra ex auro*

‘whose quiver (is made) of gold’

As Isidore observes, *deest enim ‘erat’* (‘*erat*’ is missing). Ellipsis is defined as the omission of words, and it follows that identifying and analyzing something that is, by its very nature, absent poses significant challenges. It is no coincidence that Merchant (2012), in one of his studies on ellipsis, compares the phenomenon to a black hole, justifying the analogy by noting that “one can tell it’s there (ellipsis) only by its effects on surrounding material” (Merchant 2012:8). This analogy, proposed by a leading scholar in the field, effectively captures the elusive and complex nature of ellipsis.

In recent studies, ellipsis has been described in terms of an asymmetry between form and meaning (Merchant 2018:19; van Craenenbroeck and Temmerman 2019), implying the presence of meaning without a corresponding form. Consider sentence (1), taken from a song by De André, *Se ti tagliassero a pezzetti* ‘If they cut you into pieces’:<sup>92</sup>

1) *La luna tesserebbe i capelli e il viso, e il polline di Dio, di Dio il sorriso.*

‘The moon would weave the hair and the face, and the pollen of God, God’s smile’

---

<sup>91</sup> The verses: *Sidoniam picto chlamydem circumdata limbo; cui pharetra ex auro, crines nodantur in aurum, aurea purpuream subnectit fibula vestem*. ‘her Tyrian pall is bordered in bright hues, /her quiver, gold; her tresses are confined only with gold;/ her robes of purple rare meet in a golden clasp’ (translation of Theodore C. Williams: <https://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.02.0054%3Abook%3D4%3Acard%3D129>).

<sup>92</sup> [http://www.fabriziodeandre.it/faber/wp-content/uploads/2016/03/Se\\_ti\\_tagliassero\\_a\\_pezzetti.pdf](http://www.fabriziodeandre.it/faber/wp-content/uploads/2016/03/Se_ti_tagliassero_a_pezzetti.pdf)

The underlined portion of the example illustrates the asymmetry between form and meaning. The verse of the song describes a woman, with the author suggesting that the moon would weave the woman's hair and face, while the pollen of God would do the same to her divine smile. The action of weaving in the sentence is expressed only once (*tesserebbe* 'would weave'), in association with the moon and the woman's hair and face. However, the action, and thus the meaning of weaving, remains active in the interpretation of both author and reader/listener for the second clause (*e il polline di Dio, di Dio il sorriso* 'and the pollen of God, God's smile'), even though it is not explicitly expressed in any overt form. This mismatch between what is not explicitly expressed (the form) and what is nevertheless inferred by the reader or listener (the meaning) challenges the theory of compositionality, also known as Frege's principle (Goldberg 2015). According to this theory, the meaning of an expression is determined by the meaning of its individual components and the syntactic rules that govern their combination (Goldberg 2015:419). However, in the context of ellipsis, meaning emerges despite the absence of an explicit linguistic form.

Defining ellipsis is a challenge that involves both the domain of grammar and cognition in general. This has led to the development of various accounts and analyses that span multiple fields of linguistics, including syntax, lexical semantics, semantics, discourse, prosody, stylistics, and psycholinguistics. The complexity of ellipsis arises from its position at the intersection of multiple linguistic layers, what McShane (McShane 2005:3) refers to as "doing linguistics across language modules", thus pushing to view ellipsis as "a multi-modular or interface-based phenomenon, that is: one involving several (and potentially parallel, rather than sequential) levels of grammatical and cognitive/processing representation" (Gergel and Cyrino forth: 5).

To conclude this section, it is worth recalling the observations of Kehler (2018), who effectively captures the complexity and interdisciplinary nature of the phenomenon under analysis:

"Perhaps what is most fascinating about ellipsis is how a process that seems so conceptually simple can give rise to such complexity in the data. This is no doubt due in part to the fact that it interacts with so many other complex things—spread across the domains of syntax, semantics, anaphora, and information structure—and hence why the benefits of studying ellipsis go far beyond gaining an understanding of the phenomenon itself. Indeed, studying ellipsis is so exciting precisely because it gives us a window into big questions across such a wide range of language phenomena." (Kehler 2018:341)

In light of these considerations, the literature on ellipsis is extensive and diverse, reflecting the range of theoretical frameworks through which the phenomenon has been analyzed. This Section offers an overview of ellipsis, aiming to capture the key insights of the major frameworks that have been

applied to its study. The discussion does not engage with the details of individual analyses, for which references are provided<sup>93</sup>.

### 3.1 Key concepts

Before turning to the classification of ellipsis ([Section 3.2](#)), it is essential to outline several key concepts that have been extensively discussed in the literature: i) the ellipsis site ([Section 3.1.1](#)), ii) the antecedent ([Section 3.1.2](#)), iii) the identity constraint ([Section 3.1.3](#)), iv) the notion of structure and the approaches associated with it ([Section 3.1.4](#)), and v) the discourse approach.

The aim is to introduce these concepts, with reference to the relevant literature<sup>94</sup>, in order to provide the reader with a clear understanding of these terms and their application in the present analysis.

#### 3.1.1 Ellipsis site

When processing an ellipsis, which is inherently invisible and defined by its absence, one of the first tasks the reader/listener must accomplish is to recognize it. That is, to understand that there is a gap somewhere in the sentence. Therefore, the initial step in analyzing ellipsis is to detect and locate it within the sentence<sup>95</sup>.

The term *ellipsis site* is used in the literature to refer to the gap where the ellipsis is located. In sentence (1), the ellipsis site is indicated with an underscore:

- 1) *La luna tesserebbe i capelli e il viso, e il polline di Dio \_ , di Dio il sorriso.*  
'The moon would weave the hair and the face, and the pollen of God \_ , God's smile'

The sentence/clause containing the ellipsis site is known as the target, whereas the sentence/clause without it and containing the antecedent (see [Section 3.1.2](#)) is named as the source:

- 1) *La luna tesserebbe i capelli e il viso* > **source**  
'The moon would weave the hair and the face'  
  
*e il polline di Dio \_ , di Dio il sorriso.* > **target**

---

<sup>93</sup> For a comprehensive examination of theories and approaches to ellipsis, refer to *The Oxford Handbook of Ellipsis* (van Craenenbroeck and Temmerman, 2018).

<sup>94</sup> Given the productivity of the studies on ellipsis, the bibliographic references provided should not be considered as exhaustive of the phenomenon.

<sup>95</sup> The tasks of identification and localization serve an explanatory purpose in the grammatical description of ellipsis and do not necessarily correspond to cognitive or neural processes in the mind of the reader or listener, as similar considerations fall outside the scope of the present work.

‘and the pollen of God <sub>(E)</sub>, God’s smile.’

The collocation of the ellipsis site is not always a trivial task, as will be shown in Chapter 4.

### 3.1.2 Antecedent

Another task the reader/listener must accomplish in order to process the ellipsis is what may be termed identification, namely the recovery of the meaning that fills the gap and thereby gives substance to the ellipsis. In this process, a crucial role is played by the antecedent.

The term *antecedent* refers to any linguistic expression that helps identify the meaning of an ellipsis. Ellipsis has been defined as inherently anaphoric (Phillips and Parker 2014:79), meaning that its interpretation and processing depend on linguistic material, namely an antecedent. This relationship enables the reader/listener to recover the meaning of the ellipsis. An example of an antecedent is shown in example (2):

2) *La luna tesserebbe<sub>(A)</sub> i capelli e il viso, e il polline di Dio <sub>(E)</sub>, di Dio il sorriso.*

The moon would weave<sub>(A)</sub> the hair and the face, and the pollen of God <sub>(E)</sub>, God’s smile

Here, the meaning of the ellipsis<sub>(E)</sub> is recovered through its antecedent<sub>(A)</sub>, the verb *tesserebbe* ‘would weave’ expressed in the first clause (*La luna tesserebbe i capelli e il viso* ‘The moon would weave the hair and the face’). The operation of the reader/listener is to identify the antecedent, which fits in the ellipsis site (refer to [Section 3.1.4](#) for details).

While the antecedent has traditionally been defined as anaphoric, i.e., preceding the ellipsis, it can in certain cases also be cataphoric<sup>96</sup> (Miller and Pullum 2014), in which case it is referred to as a *postcedent* (McShane 2005:14)<sup>97</sup>.

Another feature of the antecedent is that it can be overtly expressed within the sentence where the ellipsis occurs, as in example (3), or it can be recovered from a previous sentence<sup>98</sup>. In some cases, the antecedent is not explicitly stated and must be inferred from the context in which the elliptical sentence is uttered. For example, in the context of a bartender selling beers to a group of customers, during the ordering phase, a customer (C) may order the beer by simply saying:

3) *C: Io <sub>(E)</sub> la prima.*

---

<sup>96</sup> Miller and Pullum (2014) define the instances of some cataphorical ellipsis as “rhetorical device aiming to capture attention by puzzling the reader or addressee” (Miller and Pullum 2014: 20).

<sup>97</sup> In line with the majority of the literature, I will exclusively use the term *antecedent* to refer to the contextual element necessary for recovering the ellipsis, thereby also encompassing cases of possible *postcedent*.

<sup>98</sup> In his corpus-based analysis of ellipsis, Hardt (1992) found that 5% of the instances of ellipsis, specifically VP-ellipsis (see Section 3.2.1.3), in his corpus displays antecedents located at least two sentences back in the discourse.

C: I the first.’

In general, the analysis of antecedents has been extensively explored in the literature, with research focusing primarily on the type of antecedent required by ellipsis and on the nature of the relationship that must hold between the two (Merchant 2012: 27)<sup>99</sup>. This issue will be briefly addressed in [Sections 3.1.4](#) and [3.1.5](#).

### 3.1.3 Remnant

The term *remnant* generally refers to elements that are not elided in an elliptical clause. They are also described as “ellipsis survivors” that “escape the ellipsis site” (Ortega-Santos 2014:55). This term has been subject to different applications depending on the approach used to analyze ellipsis. In transformational frameworks, it is strictly associated with the notion of movement ([Section 1.3.2](#)), as a remnant is “a constituent from which material has been extracted” (Stabler 1999: 299), inherently implying movement. This term is typically associated with the *correlate*, which refers to the antecedent of the remnant. It is specifically used to describe a particular type of ellipsis in *wh*-questions, namely *sluicing* (see [Section 3.2.3.2](#)). In such contexts, the *wh*-phrase that is not elided is referred to as the *remnant*, while its antecedent is called the *correlate* (Barros et al. 2016:57). For example, in sentence (4):<sup>100</sup>

(4) Jack saw someone, but we don’t know who ~~Jack saw~~.

The remnant is “who”, and its correlate is “someone”, while the crossed-out part in the sentence represents, according to the authors, the ellipsis.

Outside of transformational accounts, the term *remnant* was also adopted in the first version of Universal Dependencies<sup>101</sup> (now no longer in use) to label a dependency relation when ellipsis occurs. The *remnant* relation is established between the correlate, selected as the head, and its remnant, which depends on the correlate as a child, as illustrated in the following sentence (5) and image 31, taken from UD guidelines<sup>102</sup>:

---

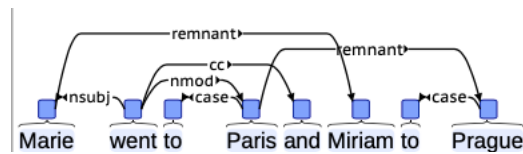
<sup>99</sup> With respect to the position of the antecedent relative to the ellipsis, it is worth noting (by way of illustration, although this study does not approach the issue from this perspective) that psycholinguistic research has examined the matter in terms of processing timing (Murphy, 1985; Martin and McElree, 2009, 2011, among others).

<sup>100</sup> The sentence is extracted from Barros et al. (2016:57).

<sup>101</sup> <https://universaldependencies.org/docs/u/dep/remnant.html>

<sup>102</sup> Both the sentence and the dependency tree are taken from the UD webpage: <https://universaldependencies.org/docs/u/dep/remnant.html>

(5) Marie went to Paris and Miriam to Prague.



**Image 31.** Use of remnant in UD v.1

As shown in the dependency tree, the *remnant* relation occurs respectively between “Marie”, as the head, and “Miriam”, as its child, and similarly between “Paris”, as the head, and “Prague”, as its child. Both pairs of words display the same syntactic function: in the first case, both words (“Marie” and “Miriam”) are subjects (*nsubj*): “Marie” is the subject of the first clause (“Marie went to Paris”), whereas “Miriam” is the subject of the omitted verb in the second clause (“and Miriam [went] to Prague”); in the second case, both words (“Paris” and “Prague”) serve the same syntactic function in their respective clause, here marked as nominal modifiers (*nmod*). Since no empty nodes are realized in basic UD (Section 1.4), the elliptical verb is not present in the syntactic tree, making it impossible to annotate the syntactic information of the two words “Miriam” and “Prague”, respectively as the subject and as a nominal modifier of the omitted verb. However, the use of the *remnant* relation and its dependency structure allow for the inference of the unexpressed syntactic relations through parallelism.

As can be understood from the examples already provided, the two approaches (transformational and dependency), frame the term *remnant* within their respective frameworks, thereby offering analytical insights that make the use of the term *remnant* slightly different and not easily comparable. This highlights the precision (and the difficulty) required in addressing terminology in the analysis of a phenomenon studied by multiple frameworks, such as ellipsis (see also Section 3.2).

### 3.1.4 Identity condition

A crucial aspect to consider in the analysis of ellipsis is the relationship between the ellipsis<sub>(E)</sub> and its antecedent<sub>(A)</sub>. This relationship is generally referred to as the identity condition, also termed parallelism by McShane (2005: 20) and identity constraint by Merchant (2012). This condition requires that the content of the ellipsis (E) be recoverable from its antecedent<sub>(A)</sub>.

For example, in sentence (6):

6) *La luna tesserebbe<sub>(A)</sub> i capelli e il viso, e il polline di Dio<sub>(E)</sub>, di Dio il sorriso.*

The moon would weave<sub>(A)</sub> the hair and the face, and the pollen of God<sub>(E)</sub>, God’s smile.

the reconstruction of the elided material  $(E)$  is identical to (or, in McShane's terminology, parallel to) its antecedent  $(A)$ . In this case, the reconstruction verb *tesserebbe* 'would weave' corresponds to its antecedent in all morphosyntactic features. By reconstructing the ellipsis, the sentence results in (7):

- 7) *La luna tesserebbe<sub>(A)</sub> i capelli e il viso, e il polline di Dio [tesserebbe], di Dio il sorriso.*  
The moon would weave<sub>(A)</sub> the hair and the face, and the pollen of God [would weave],  
God's smile.

It should be specified (a point discussed in greater detail in Section 4) that the reconstruction, which exclusively aims to make explicit content that is present at the level of meaning but remains unexpressed at the textual level, is not intended to operate at the same level as the original text. Rather, it operates at a fictional, reconstructive level and is therefore not involved in the dimension of actual textual realization. Consequently, the reconstruction process does not entail any modification of the original text, even when the resulting structure, that is, the structure obtained through reconstruction, may appear ill-formed. Note, for instance, that in (7) the reconstruction of *[tesserebbe]*, as in *e il polline di Dio tesserebbe, di Dio il sorriso*, would result in an ill-formed structure, since the position of the reconstructed verb is incompatible with the comma placed between the verb and the object. However, it is important to stress that reconstruction is a heuristic and fictional process, intended to make implicit content explicit at an analytical level rather than to propose a genuine surface-level version of the original text.

Moreover, the reconstruction of the elided material should not be understood as a simple copy-and-paste of the antecedent. Sag (1976) notes that restricting identity to surface-level correspondence is too limiting, emphasizing that the elided material is not necessarily a direct replication of the antecedent's form, since  $(E)$  may differ from its antecedent  $(A)$  in features such as agreement or voice.<sup>103</sup>

The following example (8), taken from Sag (1976:289), illustrates a case of agreement mismatch:

- 8) My brothers have all gone to the circus, and my sister  $(E)$  to the carnival.

Here, the elided material  $(E)$  should be 'has gone' to agree with its subject, 'my sister', in contrast to the antecedent, which features the plural 'have gone' to match its subject, 'my brothers'.

---

<sup>103</sup> For further insight into voice mismatches between  $(E)$  and  $(A)$ , see Merchant (2012: 30-35).

Concerning the nature of the identity between ellipsis and its antecedent, this remains a topic of ongoing debate, with divergent analyses depending on the theoretical framework adopted. This issue is closely linked to the question of the nature of the ellipsis site itself, specifically, how, and whether, it should be represented at some linguistic level (see Section 3.1.5). To clarify what is meant by (i) the ‘nature of the identity between ellipsis and its antecedent’ and (ii) the ‘nature of the ellipsis site’, two diagrams are provided below to graphically illustrate these aspects:

- i)      A  $\leftarrow(?)\rightarrow$  (E)                  nature of the identity between ellipsis and its antecedent
  
- ii)     (E)?    nature of the ellipsis site

In (i), the empirical question concerns the type of relationship, indicated by (?) in the scheme (i), between the antecedent A and the ellipsis (E), which is placed in brackets to underscore its invisible nature. As discussed in this paragraph, the antecedent and the ellipsis are connected through an identity or parallelism condition. Investigating the relationship between A and (E) therefore entails analyzing the linguistic nature of this identity, whether it is semantic, syntactic, or of a different kind. The second question, illustrated in (ii), concerns the nature of the ellipsis site itself. Since ellipsis involves absent material, scholars debate its status by asking what kind of linguistic representation, if any, the ellipsis site possesses, that is, whether and what kind of structure occupies the position where ellipsis occurs. This issue will be addressed in Section 3.1.5.

The following section, instead, provides a general overview of the different approaches to question (i) (the identity between ellipsis and its antecedent) together with references to key sources that offer major insights into the topic. The identity between the ellipsis and its antecedent has been analyzed either as syntactic identity (Chung, Ladusaw and McCloskey 1995; Chung 2005; Merchant 2008, among others), whereby the antecedent and the elided material share the same syntactic structure, or as semantic identity (Merchant 2001; AnderBois 2010; Thoms 2013), whereby ellipsis and its antecedent share the same meaning. A third line of research, developed within referential theories (Hardt 1993; Kehler 2000; Poppels 2022), challenges the very notion of identity, proposing instead that ellipsis is licensed by the same mechanisms that govern other forms of discourse reference (Poppels 2022).

### 3.1.5 Structure and (non-)structural approaches

Another widely debated aspect of ellipsis concerns the identity of the ellipsis site (Phillips et al. 2014:80), specifically how, at what level of linguistic abstraction, it should be represented, if at all.

This issue centers on whether the elided material in the ellipsis site (*E*) actually exists and, if so, what kind of structure it assumes (see the diagram (ii) in Section [3.1.4](#)). Building on this question, the literature on ellipsis is divided into structural and non-structural approaches: structural approaches posit the existence of an underlying structure for the elided material, while non-structural approaches reject such a structure and offer alternative explanations.<sup>104</sup>

Specifically, non-structural approaches (Ginzburg and Sag 2000; Culicover and Jackendoff 2005) deny the presence of any underlying structure, adopting instead a ‘what you hear is what you get’ (WYHIWYG) framework (Merchant 2012:4). Structural approaches, by contrast, posit the existence of an unpronounced structure within the ellipsis site. These approaches diverge, however, in how they characterize such a structure: some argue for a syntactic nature, others for a semantic one, and still others adopt a hybrid perspective that incorporates both. The following paragraphs provide an overview of these two main approaches.

### *3.1.5.1 Structural approaches*

Although all structural accounts share the hypothesis that a structure exists at the ellipsis site, they differ in how they analyze the nature of the elided content and its relationship to the antecedent. The syntactic account holds that the elided material has a syntactic structure identical to its non-elliptical counterpart but remains unpronounced, either through deletion or as a result of prosodic algorithms. Some proposals (Merchant 2012) posit the existence of an ellipsis feature, the so-called E-feature, described as the “sole repository of all the information about the ellipsis” (Merchant 2012:21). This feature licenses ellipsis and plays a crucial role in guaranteeing its recoverability. The concepts of licensing and recoverability are central to syntactic accounts of ellipsis (McShane 2005:4). Licensing refers to the criteria or mechanisms that permit ellipsis in specific contexts, determining where it is possible and explaining why it is excluded in others (Lobeck 1995, among others). Recoverability, by contrast, concerns the process of recovering the elided material, requiring an analysis of its relationship to the antecedent, therefore connected with the identity condition.

Within structural approaches, not all accounts posit a strictly syntactic relation between the antecedent and the elided material. Semantic accounts<sup>105</sup> propose instead the existence of a null element or empty

---

<sup>104</sup> The distinction between structural and non-structural accounts is examined in detail by Merchant (2012). The same issue, albeit with different terminology, is addressed in Phillips et al. (2014), who frame it in terms of the “nature of the antecedent”, distinguishing between syntactic nature (structural accounts) and semantic/discourse-based nature (non-structural accounts).

<sup>105</sup> It is worth noting that the systematization of accounts differs slightly between Merchant (2012) and Phillips et al. (2014). Merchant (2012) offers a more fine-grained classification, distinguishing between studies that state the existence of syntax in the ellipsis site but analyze the identity between antecedent elided material as semantic and those that reject the existence of a syntactic structure altogether and likewise analyze identity as semantic. By contrast, Phillips (2014: 81) groups all such accounts under the broader label of semantic accounts.

placeholder node, consisting either of a copy of the antecedent structure or of material interpreted via anaphoric mechanisms (Hardt 1993; Lobeck 1995, among others).

### 3.1.5.2 *Non-structural approaches*

Non-structural approaches reject the existence of a syntactic structure associated with the elided material, instead offering alternative interpretations. Some accounts propose the presence of a free variable that directly yields a semantic interpretation (Culicover and Jackendoff 2005), while others (Dalrymple et al. 1991) posit the existence of a property P, determined by the antecedent and linked to the meaning of the predicate.

### 3.1.6 Discourse approach

Among the various theories on ellipsis, there are also mixed approaches that integrate both syntactic and semantic/anaphoric perspectives.

Halliday and Hasan (1976), in their seminal work *Cohesion in English*, analyze ellipsis as a cohesive device. Their focus, however, is not on ellipsis within the sentence but rather on inter-sentential relations, “where there are no structural relations, and this is where the study of cohesion becomes important” (Halliday and Hasan 1976:146).

A discourse-oriented account of ellipsis is also offered by Kehler (2000; 2002), who incorporates discourse structure into the study of ellipsis and challenges approaches confined to a single linguistic module. Within his framework of discourse coherence, Kehler integrates syntactic and anaphoric analyses and emphasizes the role of coherence relations between the clause containing the ellipsis (the target clause) and the clause containing its antecedent (the source clause). He identifies three main types of coherence relations: Resemblance (including Parallel and Contrast), Cause–Effect (Result, Explanation, Violated Expectation, and Denial of Preventer), and Contiguity (Narration). Crucially, Kehler (2000) shows that Resemblance relations require syntactic identity between source and target clauses, whereas Cause–Effect relations do not, since their interpretation is guided by semantic analysis.

Another framework situating ellipsis at the interface between syntax and discourse is the dialogic syntax model proposed by Du Bois (2014). This approach seeks to unify syntax and discourse by analyzing syntax in dialogic interaction, thereby extending syntactic analysis beyond the intra-sentential level to the inter-sentential one, in line with Halliday & Hasan’s (1976) perspective. The central syntactic unit in dialogic syntax is the diagraph, defined as “a higher-order, supra-sentential syntactic structure that emerges from the structural coupling of two or more utterances” (Du Bois 2014: 376). The diagraph represents multiple levels of linguistic structure, not only syntax, and

enables the real-time mapping of words and structures produced by dialogic co-participants. Du Bois contrasts the diagraph with traditional linear syntax, which “isolates its sentences (whether attested or invented)” (2014: 394), thereby narrowing the range of observable syntactic phenomena. Ellipsis is discussed as a key example of a structure that operates across sentence boundaries, making it “a promising test case for the integration of insights about intra-sentential syntax from linear syntax with analyses of supra-sentential syntactic structures from dialogic syntax” (Du Bois 2014: 398). Nonetheless, dedicated studies of ellipsis within the dialogic syntax framework are not yet available.

### 3.2 Classification of ellipsis

In the study of ellipsis, the notion is treated as a “cover term” (Phillips et al. 2014: 78) for linguistic phenomena involving the omission of material that would otherwise be expected in a sentence. Consequently, delineating the full scope of what this cover term encompasses is a challenging task. As noted in [Section 3](#), this difficulty arises from the inherently absent nature of ellipsis: being defined by omission, it is particularly difficult to identify and retrieve. Moreover, a further challenge lies in the fact that research of ellipsis in the existing literature has been strongly shaped by the theoretical frameworks adopted.

It is widely acknowledged that the terminology employed in scientific inquiry is deeply tied to the theoretical framework in which it is situated (Kuhn 1962). With respect to ellipsis, the phenomenon has predominantly been examined within transformational frameworks, where analyses largely focus on syntactic movement, an aspect not addressed in dependency syntax (see Section 1.3.2). By contrast, dependency-based research has offered only limited examinations of ellipsis, primarily focusing on specific cases such as gapping (see [Section 3.1.4](#) for further discussion of ellipsis in dependency syntax). The dominance of transformational approaches has produced a classification system in which ellipsis labels are closely tied to the framework itself, with the risk that the terminology remains specific and meaningful only within that approach. A case in point is the so-called Right-Node Raising (RNR)<sup>106</sup> (Wilder 2019:700-718), whose label explicitly refers to the upward movement of nodes within the tree structure. The term RNR refers to constructions in which material is absent from non-final conjuncts, as illustrated in example (9):<sup>107</sup>

- 9) [We are not responsible for ~~any losses incurred~~] and [we will not reimburse you for any losses incurred]

---

<sup>106</sup> It should be noted that RNR has been considered a non-canonical ellipsis phenomenon (Wilder, 2019:706) and will therefore not be included in the present classification.

<sup>107</sup> Example cited from Wilder (2019:682).

In example (9), conjuncts are represented within brackets. In the first conjunct, the element shared with the final conjunct (“any losses incurred”) is omitted. The phenomenon was analyzed by Ross (1967) in terms of the so-called across-the-board movement (ATB), specifically triggered by a rightward extraction rule. Although Wilder (2019:682) notes that the label RNR can be applied to the phenomenon without presupposing a specific analysis, the term nonetheless carries a clear theory-related connotation. This exemplifies how the definition of the phenomenon remains closely tied to the framework in which it is formulated.

Furthermore, another challenge in investigating ellipsis lies in the fact that different frameworks may employ the same label to denote slightly different concepts. This is evident in the case of remnant discussed in [Section 3.1.3](#): in transformational accounts, it refers to a constituent from which material has been extracted, whereas in dependency accounts, specifically in Universal Dependency v.1, it designates a syntactic relation between the remaining nodes in the elided sentence and their correlates. In light of these considerations, the cross-framework review of phenomena requires careful clarification of terminology.

In this section, an overview is provided of the major types of ellipsis discussed in the literature, without the aim of offering an exhaustive account of the vast body of research on the topic. The description does not engage with technical aspects such as theories of constituent movement (for which references are supplied where relevant),<sup>108</sup> as such fall beyond the scope of the present study. Instead, the goal is to provide readers with a descriptive overview.

For clarity, a distinction is first made among predicate ellipsis (Section 3.2.1), nominal ellipsis (Section 3.2.2), and clausal ellipsis (Section 3.2.3), following the classification proposed by Halliday and Hasan (1976). Specific cases and subtypes within each category are then examined in greater detail. Since the present study focuses on a corpus-based classification and analysis of predicate and object ellipsis in the *Divine Comedy*, a detailed account of these types is provided in Chapter 5. A summary table (Table 3) of all types of ellipsis is included in the Appendix.

### 3.2.1 Predicate ellipsis

Among the various types of ellipsis, predicate ellipsis (Aelbrecht and Harwood 2018: 504-525) specifically targets the main predicate of the clause, often including the object and, optionally, its adjuncts. In the following Sections, the types of ellipsis that involve the predicate and have been classified as such in the literature are presented.

---

<sup>108</sup> It should be noted that each type of ellipsis has been the subject of numerous studies, with many works devoted to individual cases and some engaging in debates over the classification of specific phenomena.

As noted in [Section 3.2](#), the aim of this classification is to provide a descriptive overview of the main ellipsis typologies of ellipsis as presented in previous studies, with references to the cited sources for further insights into the specific characteristics of each type. Appendix presents a table (Table 4) summarizing all the instances of predicate ellipsis discussed in these sections.

### 3.2.1.1 Gapping

Gapping, whose name and main insights are attributed to Ross (1970), involves the omission of the verb (including both the lexical verb and any auxiliaries, if present) in the second conjunct of a coordination (Johnson 2019:562-604). It has been described to typically occur only in conjunctions or disjunctions; therefore, it is also considered a central conjunction reduction phenomenon (Johnson 2019:563). Example 10 reports a gapping construction:<sup>109</sup>

10) Jones likes seafood and Smith, bread.

In (10), gapping is realized in the second conjunct ‘Smith, bread’, as the main verb, which is identical to the one in the first conjunct (‘likes’), is omitted. As shown in example (10), gapping is considered essentially anaphoric (Johnson, 2019:573), as the meaning of the omitted verb is the same as that of the verb in the first conjunct.

It is also important to note that gapping may occur not only in the second conjunct (as shown in example 10) but also in the first one. The former case, where gapping occurs in the second conjunct, is called forward gapping (Example 10), whereas the latter, where gapping appears in the first conjunct, is usually referred to as backward gapping. An example of backward gapping is illustrated by a German sentence in Example 11:<sup>110</sup>

11) *Ich weiß, dass die Kinder Fisch ~~LIEBEN~~ und die Eltern Fleisch lieben.*

I know that the children fish ~~LOVE~~ and the parents love meat.

As pointed out by Krisch (2009), forward and backward gapping correlate with word order and with the directionality of verbal government. Specifically, forward gapping occurs in SVO (subject-verb-object), SOV, and VSO languages, whereas backward gapping appears only in SOV languages (Krisch 2009:193). Among the analysis of gapping, Kuno (1976) and Ramat (1982) consider this phenomenon to be pragmatic rather than syntactic. They argue that gapping affects discourse

---

<sup>109</sup> Example 10 is cited by Johnson (2019:562).

<sup>110</sup> Example 11 is cited by Krisch (2009:193).

strategies and cognitive processes related to memory. A hybrid approach is taken by Krisch (2009), who, in addition to associating gapping with word order and verbal government, also considers factors related to functional sentence perspective.

### 3.2.1.2 *Pseudo-Gapping*

Pseudo-gapping is closely related to gapping (hence its name) as is a type of ellipsis that may retain the internal argument (object) of the elided predicate, as observed in [Section 3.2.1.1](#) for gapping constructions. However, unlike gapping (and in this respect resembling another type of ellipsis<sup>111</sup>, VP-ellipsis, described in [Section 3.2.1.4](#)), it does not affect the “inflectional domain” (Aelbrecht and Harwood 2018: 504-505). In other words, finite auxiliaries are not elided together with the lexical verb, as illustrated in Example 12 (Aelbrecht and Harwood 2018: 504):

12) Some brought roses, and others did lilies

In Example 12, pseudo-gapping appears in the second conjunct ‘others did lilies’, where the lexical verb ‘buy’ is omitted, leaving only the auxiliary ‘did’ and the object ‘lilies’.

It should be noted that in Modern Standard Italian, pseudo-gapping is generally regarded as unacceptable (Example 13):

13) \**Qualcuno ha comprato le rose, e altri hanno i mughetti.*

Some brought roses, and others did lilies.

Evidence of analyses of pseudo-gapping outside generative frameworks remains scarce.

### 3.2.1.3 *VP-ellipsis*

VP-ellipsis (also VPE) refers to the omission of the verbal phrase (VP). In this construction, “the predicate and its internal arguments” are omitted (Aelbrecht and Harwood 2018: 504), while modals, finite auxiliaries, and the subject remain. VP-ellipsis is among the most extensively studied ellipsis phenomena, particularly in English (Aelbrecht and Harwood 2018: 519). Sentence (14) provides an example of VP-ellipsis (Aelbrecht and Harwood 2018: 506):

---

<sup>111</sup> The fact that pseudo-gapping shares properties with two other types of ellipsis, namely gapping and VP-ellipsis, has been discussed by Gengel (2013), who also provides a detailed analysis of pseudo-gapping within a generative framework.

14) Mr HoneyPie was happily eating pumpkins, but his wife certainly wasn't.

In (14), the lexical verb and its object 'eating pumpkins' are omitted in the second conjunct, 'his wife certainly wasn't'. A specific type of VPE is verb-stranding VPE, where, unlike classical VPE, the lexical verb is not elided; only the material following the lexical verb (i.e., the internal arguments) is omitted. This phenomenon is attested in languages such as European and Brazilian Portuguese, Hebrew, and Irish. (Aelbrecht and Harwood 2018: 519).

#### 3.2.1.4 Other predicate ellipsis

Aelbrecht and Harwood (2018: 522-525) also describe other types of verbal ellipsis, namely "British English do" and "modal complement ellipsis".

British English do (also "BE do") is similar to VP-ellipsis but includes an additional non-finite form of the verb 'do' (Aelbrecht and Harwood 2018: 525):

15) Luis will run the race and Nana will do, too.

As in VP-ellipsis, the lexical verb and its object 'run the race' are omitted in the second conjunct 'Nana will do, too'. However, unlike VP-ellipsis, an additional instance of 'do' is present. This type will not be included in summary Table 3, as it is language-specific.

Modal complement ellipsis, or modal ellipsis, is a type of VP-ellipsis that elides the lexical verb following modal verbs. It occurs only with root modals and not with aspectual auxiliaries (Aelbrecht and Harwood 2018: 506):

16) I want to help you, but I can't.

This type of ellipsis is also attested in Italian when the subjects of the modal and the matrix clause are coreferential (Dagna 2010:161):

17) *Maria legge tutti i libri che può.*

Maria reads all the books that she can.

The Appendix, in addition to a general overview table of ellipsis (Table 3, Section 3.2), includes a table (Table 4) summarizing all the instances of predicate ellipsis discussed above.

### 3.2.2 Nominal ellipsis

Nominal ellipsis is a “somewhat vague label” (Saab 2019: 526) used to describe gaps within the nominal phrase. In the following Example 18, a case of nominal ellipsis is illustrated in which the noun is omitted, leaving the adjective to function as the head of the nominal phrase<sup>112</sup>:

18) *tali erano i nomi dei due ordigni, il primo destinato a Hiroshima;*  
such were the names of the two devices, the first (one) intended for Hiroshima;”

In (18), the adjective *primo* “first” functions as the head of a nominal phrase, with the noun (*ordigno* ‘device’) omitted. In a non-elliptical noun phrase such as *il primo ordigno* ‘the first device’, the noun *ordigno* ‘device’ would serve as the head of the phrase, while the adjective *primo* ‘first’ and the article *il* ‘the’, would function as its modifiers.

Classifications of nominal ellipsis vary according to the analytical perspective adopted. For instance, Halliday and Hasan (1976), who treat ellipsis as a cohesive device (see [Section 3.1.6](#)), classify nominal ellipsis on the basis of the linguistic material that remains overt in the elliptical sentence rather than what is omitted. Their classification focuses on the grammatical category of the promoted modifier, distinguishing among: deictic, typically determiners; numerative, corresponding to numerals or other quantifiers; epithet, consisting of adjectives; classifier, corresponding to nouns; and qualifier, typically a relative clause or a prepositional phrase (Halliday and Hasan 1976: 147).

Conversely, other classifications emphasize the linguistic material that is omitted rather than on what remains overt. Saab (2019), for instance, proposes a classification of nominal ellipsis based “on the size of the elliptical site” (Saab 2019:559). This approach involves reconstructing the missing material and predicting the licensing conditions, which Saab attributes to a specific ellipsis features *e* (Saab 2019:550) that determines which phrases can be elided and which cannot. On this basis, Saab distinguishes three types of nominal ellipsis: *nP* ellipsis, *NumP* ellipsis and *RootP* ellipsis.

Another classification of nominal ellipsis, which focuses on the omitted nominal element, emphasizes the syntactic function that the missing element would otherwise fulfilled. Studies in this area typically refer to null arguments (Haug and Luraghi 2012), analyzing ellipsis in terms of valency gaps. Valency, a concept introduced by Tesnière (1959), refers to the argument structure of the verb. Since the verb denotes an event, it inherently introduces participants, known as verbal arguments (Luraghi and Roma 2021). In cases of ellipsis, or argument omission, one of these participants (i.e., the subject or the object) is not overtly expressed.

---

<sup>112</sup> The example is taken from the ISDT treebank (isst\_tani-1170) (UD v. 2.15).

[Sections 3.2.2.1](#) and [3.2.2.2](#) briefly address the two main types of null arguments, subject and object ellipsis, with particular emphasis on the latter.

### 3.2.2.1 Subject ellipsis

**Subject ellipsis** is attested in so-called *pro-drop* languages, where the subject can be omitted because its identity is recoverable from the verbal morphology. Languages that allow subject ellipsis include Italian and Old Italian. Examples of subject ellipsis in Modern Italian (19)<sup>113</sup> and in Old Italian (20)<sup>114</sup> are shown below:

19) *non ho finito il liceo artistico, mi sono ritirato alla maturità.*

(I) didn't finish high school art, (I) withdrew at high school graduation.

20) *Nel mezzo del cammin di nostra vita mi ritrovai per una selva oscura (...)*

When (I) had journeyed half of our life's way, (I) found myself within a shadowed forest (...)

In both examples, the subject *io* 'I' is omitted and inferred from the first-person singular verb inflection. In the Modern Italian example, this distinction is indicated by the auxiliaries *ho* 'have.1SG' and *sono* 'am', the latter disambiguated by the singular suffix of the past participle *ritirat-o* 'withdrew.SG.MASC' to avoid potential ambiguity with the third-person plural, which also employs the auxiliary *sono* 'are'. In the Old Italian example, subject recovery instead relies on the verb form *ritrova-i* 'found.1SG', which carries the first-person singular inflection.

Subject ellipsis has attracted considerable scholarly attention, particularly in *non-pro-drop* languages, where subject omission is generally disallowed except under specific conditions. One example is found in written and spoken English, where subject omission is associated with register variation (Haegeman 2007; Bonsignori 2018). In particular, it is frequently attested in diary-style writing and in registers associated with younger speakers (Bonsignori 2018:68-72), underscoring the crucial role of diamesic and diaphasic variation (Berruto 1993:8-11) in enabling subject omission. With regard to Italian, Marengo (1984) analyzed the omission of subject personal pronouns. To the best of my knowledge, no specific study of subject omission in Old Italian has yet been undertaken.

### 3.2.2.2 Object ellipsis

---

<sup>113</sup> The sentence is taken from the ISDT treebank (isst\_tanl-2239) (UD v. 2.15).

<sup>114</sup> *Divine Comedy* of Dante Alighieri, *Inf.* I, vv. 1-2. English translation by Robert P. Mandelbaum (<https://digitaldante.columbia.edu/dante/divine-comedy/inferno/inferno-1/>)

A specific type of nominal ellipsis, or null argument, is object ellipsis, which involves the omission of the object of transitive verbs. This phenomenon has received limited attention in linguistic studies, specifically compared to the other types of ellipsis. Before providing a detailed description of object ellipsis, it is important to clarify that the term may overlap with, but it is not directly equivalent to, another widely discussed concept in ellipsis studies, namely null complement anaphora.<sup>115</sup> This term refers to cases in which the complement of certain verbs is omitted but recoverable through anaphoric interpretation. In null complement anaphora (Depiante 2019), the omitted complement may also take the form of a clause, as illustrated in the following example 21 (Depiante 2019:657):

21) When mother told him to clean up his room, Tommy refused.

In null complement anaphora, the omitted material corresponds to a previously expressed clause, in (21) ‘to clean up his room’. Moreover, this label does not extend to cases where a transitive verb is used intransitively, as in (22):<sup>116</sup>

22) John was hungry, so he ate.

In such instances, the omitted object is not interpreted anaphorically but instead conveys a more general meaning.

By contrast, the term object ellipsis also referred to as object drop, object omission, or null object, is used here to denote instances where the object is omitted in a prototypical transitive verb pattern. This involves an event that normally requires two participants, an agent and a patient/theme, where the patient/theme-like participant is left unexpressed. Examples (23) and (24), taken from a well-known passage of *The Little Prince* (De Saint-Exupéry 2015: 6), illustrate the verb ‘understand’ with and without an object:

23) Grown-ups never understand anything on their own.

24) I then drew the insides of the boa constrictor, so that the grown-ups could understand.

In example (23), the object of ‘understand’ is explicitly expressed (‘anything’). By contrast, in (24), the object is not overtly expressed and must be inferred from the context: the grown-ups are expected to understand that the picture shown by the Little Prince is not, as they first perceive, a hat, but rather

---

<sup>115</sup> This phenomenon has been referred with different name in the literature, such as ‘definite constituent ellipsis’ by Shopen (1973:67); ‘definite null complement’ Fillmore (1986). Refer to (Depiante 2019).

<sup>116</sup> The example is cited by Depiante (2019:657).

an elephant inside a boa constrictor. Metaphorically, one might say that the invisible elephant inside the boa constrictor parallels the omitted object: although not visible, it is nonetheless present and recoverable.

But this raises the question: does the elephant exist at all? The absence of an element does not necessarily imply the non-existence of a referent, a point analyzed by Fillmore and Kay (1993) under the notion of null instantiation. Within constructional frameworks, instantiation is defined as “a characterization of the symbolic relation between syntactic elements and semantic components” (Croft 2001:276). Fillmore and Kay (1993) use the term null instantiation to describe cases in which a referent (regardless of its syntactic function, and thus not restricted to object ellipsis) is instantiated without being overtly expressed. By analyzing the pragmatic status of the referent, they distinguish three types of null instantiation:

- free null instantiation, where the referent can be freely identified by the hearer;
- indefinite null instantiation, where the referent is indefinite;
- definite null instantiation, where the referent must be accessible (also known as pro-drop or null anaphora).

Examples of each type of null instantiation are provided below<sup>117</sup>:

25) She took some blue stationery<sub>i</sub> out of the cupboard and wrote a letter 0<sub>i</sub>.

26) The dog just ate 0<sub>i</sub>.

27) 0<sub>i</sub> *Mangiano qualcosa insieme, in un ristorante.*

(They) eat something together, in a restaurant.

In example (25), the free null instantiation, marked in the sentence with 0<sub>i</sub> by the authors, is recoverable by the hearer through context. Specifically, following the author’s intuition, the reader/listener associates the ‘blue stationery’ with the letter that was written, even though this is not explicitly stated. In such cases, however, the referent may also remain entirely unmentioned. In example (26), the indefinite null instantiation corresponds to an omitted and unspecified object, interpreted as some generic food that the dog ate. Finally, in example (27), the definite null instantiation corresponds to an omitted subject, which is not overtly expressed but is recoverable by the hearer from the verb’s third-person plural inflection.

Returning to object ellipsis and focusing specifically on Modern Italian, this phenomenon has been extensively studied (Lo Duca 2000; Cennamo 2015, 2017). Cennamo’s (2015) analysis of object

---

<sup>117</sup> Examples (25) and (26) are cited by Croft (2001:276), whereas example (27) is taken from ISDT (isst\_tanl-1336).

omission in Modern Italian was conducted as part of a “large-scale cross-linguistic comparison of valency classes”<sup>118</sup>, which led to the creation of the open-access database ValPal (Hartmann et al. 2013), later expanded into PaVeDa (Luraghi et al. 2024). In this framework, object omission in Modern Italian is analyzed as a valency-changing strategy (Cennamo 2015: 422) involving an uncoded argument alternation (Cennamo 2015:440). In other words, verbs that are prototypically transitive display optionality in the expression of the object. This is interpreted as a change in the so-called basic valency pattern<sup>119</sup> of the verb (that is, the argument structure in which a verb most frequent), that remains unmarked (uncoded).

More specifically, a verb that prototypically occurs in a transitive construction, such as the Italian verb *cercare* ‘to search’, with both a subject and an object expressed (example 28), can also occur without the object (example 29<sup>120</sup>), resulting in an uncoded argument alternation:

28) *Anna cerca un documento importante.*  
Anne is looking for an important document.

29) *Marco ha cercato nel cassetto.*  
Mark has searched in the drawer.

Unlike in example (28), where the object (*un documento* ‘a document’) of the verb is expressed, in example (29) the object is omitted.

As Lo Duca (2000) observes, not all omitted objects are of the same type. In her discussion on the lexicographic treatment of argument alternation in Italian, she distinguishes three main classes of object omission, or verbs with a null object (Lo Duca 2000: 5), specifically referring to transitive verbs that can systematically occur without an object. The first class includes verbs with indefinite null objects, where the omitted object is indefinite and generic, conveying a self-sufficient action without requiring specification of the object, as shown in Example (30):

30) *Maria ama cantare, leggere e studiare.*  
Maria loves singing, reading, and studying.

---

<sup>118</sup> <https://valpal.info/project> (Consulted 12.03.2025).

<sup>119</sup> “In order to assess which coding frames are basic, we plan to give a greater weight to the frequencies of attested patterns” (Zanchi et al, 2022: 101).

<sup>120</sup> Example taken from PaVeDa database, specifically for Italian (<https://paveda.unipv.it/alternationvalues/106269>).

Although some literature describes similar cases using the term deletion (Allerton 1975: 214–215), thereby implying that an element (in this case, the object) has been deleted, it is important to note that for Lo Duca (2000) an indefinite null object does not entail an implied or elliptical object (Lo Duca 2000: 6). The second class consists of verbs with generalized null object, where the omitted object is easily reconstructed and it exhibits the following features: [+/- human], [+ generic], and [+ plural]:

31) *È una ragazza che affascina (tutti)*

‘She is a girl who fascinates (everyone).’

In such cases, it is acceptable to refer to an implied object. The presence of an implied object is indicated by the fact that possible object predicatives appear in the plural (Rizzi 1986: 505):

32) *Di solito, Gianni fotografa seduti.*

‘In general, Gianni photographs seated.’

In Example (32), the object predicative *seduti* (‘seated.PL.MASC’) explicitly refers to an implied object that is generalized; in other words, it is a generalized null object. This type of omission is allowed in habitual or atelic contexts but not in perfective ones.

Finally, the last class individuated by Lo Duca (2000) consists of verbs with defined null object, where the null object is evoked either by the linguistic context (Example 33) or by the situational context of the communicative act (Example 34). Alternatively, the defined null object is obligatorily evoked due to the nature of the event itself (Lo Duca 2000:10) (Example 35):<sup>121</sup>

33) *Le ho fatto una proposta e lei ha rifiutato*

‘I made her a proposal, and she refused (it)’

34) *Tira!*

‘Pull!’

35) *Maria sta apparecchiando (la tavola)*

‘Maria is setting (the table)’

---

<sup>121</sup> Examples are cited by Lo Duca (2000:10-11).

Building on Lo Duca's (2000) classification and linking it to the typological framework proposed by Fillmore and Kay (1993), and Croft (2001), Cennamo (2017) identifies the following subtypes of object omission in Modern Italian:

- indefinite null instantiation (in perfective and imperfective contexts), where the referential is indefinite or reconstructed by the context;
- indefinite null instantiation (in imperfective contexts only), which corresponds to the generalized null object in Lo Duca (2000);
- definite null instantiation, also known as anaphoric null object and deictic null object.

Moreover, Cennamo (2017) emphasizes that object omission in Italian is also connected with semantic constraints, lexical semantic and discourse-pragmatic perspective. More precisely, she notes that the omissibility of the object is sensitive to:

- **event structure templates**, as verbs with a low degree of aspectual specification are more likely to allow omission;
- **lexical semantics** as the meaning elements lexicalized in the verb root, interacting with notions such as animacy, definiteness and referentiality, influence object omission;
- **discourse-pragmatic factors**, as contextual recoverability and focus-topic distinction affect whether the object can be omitted.

Moreover, verbs with a simple event structure, such as activity verbs, are more likely to allow object omission than those with a complex event structure, such as accomplishments and achievements. Object omission in Old Italian will be discussed in Section 3.3.1.2.

### 3.2.3 Clausal ellipsis

This category includes all instances of ellipsis occurring at the clause level, where the omission is not limited to the verb phrase (and its internal argument) or the nominal phrase.

A brief overview of the main types of clausal ellipsis is provided, together with key literature references, to offer a comprehensive perspective on the phenomenon. Specifically, stripping/bare argument ellipsis is discussed in [Section 3.2.3.1](#); fragments in [Section 3.2.3.2](#), sluicing in [Section 3.2.3.3](#), and ellipsis in comparatives in [Section 3.2.3.4](#).

#### *3.2.3.1 Stripping/bare argument ellipsis*

Stripping, also known as bare argument ellipsis (Merchant 2012:3), has been defined as an ellipsis “that deletes everything in a clause under identity with corresponding parts of a preceding clause, except for one constituent (and sometimes a clause-initial adverb or negative)” (Hankamer and Sag 1976: 409). An example of stripping is shown in (36) cited by (Hankamer and Sag 1976: 409):

36) Alan likes to play volleyball, but not Sandy.

In Example (36), stripping occurs in the second conjunct, where the omission affects an entire portion of the clause, specifically, a structure similar to ‘does not like to play volleyball’, leaving only the subject and the negation.

### 3.2.3.2 *Fragments*

When discussing fragments, the literature (Merchant 2004; Hall 2018:605-623, Weir 2014, among others) refers to parts of sentences, such as short answers to questions or utterances without an overt antecedent. Examples of fragments are illustrated in (37) and (38):<sup>122</sup>

37) (A) ‘What did you buy?’

(B) ‘A new coat.’

38) [Uttered by a waiter displaying a bottle of wine to customers] From Italy.

Both the short answer in (37.B) and the utterance in (38) are isolated nominals. The debate in the literature concerns the sentential (Merchant 2004; Weir 2014)<sup>123</sup> versus non-sentential (Stainton 2006) nature of fragments. Specifically, it questions whether fragments involve the syntax of a full sentence subject to ellipsis or whether they consist solely of the overt utterance.

### 3.2.3.3 *Sluicing*

Sluicing is a type of clausal ellipsis (Johnson 2019: 599) that involves the omission of all elements except the *wh*-expression from an interrogative clause (Ross 1969; Merchant 2006; Vicente 2019: 479-503). The term sluicing is primarily associated with transformational accounts, though the same phenomenon is referred to as an “indirect *wh*-question” in other frameworks (Halliday and Hasan 1976). An example of sluicing is provided in (39):<sup>124</sup>

39) Someone just left—guess who [just left]!

---

<sup>122</sup> Examples are cited by Hall (2019:605).

<sup>123</sup> Merchant 2004 proposes an account of fragments that involves movement to the left periphery in the syntactic structure. In contrast, Weir 2014 demonstrated that movement is not necessarily implicated, at least not at the syntactic level, as he hypothesizes movement occurs at the phonological level.

<sup>124</sup> Example is cited by Vicente (2019:479).

In (39), the interrogative pronoun ‘who’ is interpreted as a remnant (i.e., what remains; see [Section 3.1.3](#)) of an elliptical clause, represented within the square brackets in Example 39. The antecedent can be provided either syntactically or as a pragmatic or discourse antecedent (Lobeck 1995), depending on the analysis.

#### *3.2.3.4 Ellipsis in comparatives and Comparative deletion*

Comparative structures are well known for their syntactic complexity (Bresnan 1973:275). However, with respect to ellipsis, the challenges posed by these constructions are not fundamentally different from those discussed thus far. Before analyzing ellipsis in comparatives, it is important to clarify that “ellipsis in comparatives” differs slightly from the types of ellipsis listed previously, as it refers to the grammatical context in which various types of ellipsis occur. In fact, multiple types of ellipsis, including gapping, stripping, and VP-ellipsis, among others (Lechner 2019: 624), can be observed in comparative structures, just as they appear in other clause types.

This study does not engage with the debate (Lechner 2019: 638) on whether ellipsis in comparatives should be classified as a process peculiar to comparative degree constructions, i.e., whether it constitutes a distinct type of ellipsis unique to comparative clauses, or whether it should be treated as an independently attested ellipsis operation, thereby assimilating it to the ellipsis in other clause types. Consequently, although ellipsis in comparatives is treated separately here for descriptive purposes, the use of the term remains neutral with respect to this distinction.<sup>125</sup>

Examples of different types of ellipsis in comparatives, specifically gapping, stripping, and VP-ellipsis, are provided in (40), (41), and (42), respectively:<sup>126</sup>

- |   |               |
|---|---------------|
| 40) Ann liked Berlin more than Cleo [liked] Dubai     | (gapping)     |
| 41) Ann liked Berlin more than Cleo [liked Dubai]     | (stripping)   |
| 42) Ann liked Berlin more than Cleo did [like Berlin] | (VP-ellipsis) |

In all cases, the elided material, as indicated in the original examples, is represented within square brackets. As with other types of ellipsis, however, a central and much-debated question arises when examining the reconstructed material: What, if anything, is missing? A fundamental divergence in the literature emerges in the analysis of phrasal comparatives, such as the one in (43):

---

<sup>125</sup> See Bresnan (1973) for a discussion of comparative ellipsis.

<sup>126</sup> Examples cited by Lechner (2019:624).

43) Ann is taller than Bill.

In (43), the phrasal comparative ‘than Bill’ is subject to different analyses. The direct analysis treats the phrasal comparative as a prepositional phrase (Hankamer 1973)<sup>127</sup> and, therefore, a sentence as the one in (43) does not involve ellipsis. By contrast, the reduction analysis (Bhatt and Takahashi 2011) posits that the phrasal comparative embeds an elided syntactic structure, thus leading to a clausal analysis of the phrasal comparative. Typologically, it has been shown that phrasal comparatives do not form a homogeneous class, as different languages respond differently when tested about the presence of silent syntactic structure (Bhatt and Takahashi 2011). To the best of my knowledge, similar studies on Old Italian comparatives have not yet been conducted. In the case of Modern Italian, however, Napoli and Nespor (1986) analyze inequality comparatives, identifying both a prepositional phrase introduced by *di* ‘of’ and a clausal comparative with *che* ‘than’. Accordingly, in the following examples,<sup>128</sup> (44) illustrates a prepositional phrase, whereas in (45) *che* ‘than’ functions as a so-called “coordinator” (Napoli and Nespor 1986:634):

44) *Marta ama Luca più di Giacomo*

Marta loves Luca more than Giacomo

45) *Considero più intelligente Giorgio che Luigi*

I consider Giorgio more intelligent than Luigi

Thus, while sentence (44) is analyzed as a matrix clause with a comparative prepositional phrase *di Giacomo* ‘than Giacomo’, sentence (45) is structured as a matrix clause with a comparative clause introduced by *che*. This implies that ellipsis applies only to constructions introduced by *che*, as in (45).

#### 3.2.3.4.1 Comparative deletion

---

<sup>127</sup>Hankamer identifies two different types of the English *than*: one functioning as a conjunction (or “complementizer”), thereby supporting the clausal nature of the comparative, and the other functioning as a preposition, resulting in a prepositional phrase. Only clausal comparisons can be analyzed as “the remains of underlying full clauses which have undergone ellipsis” (Hankamer, 1973:182).

<sup>128</sup> The examples are taken from Napoli and Nespor (1986:634).

When addressing ellipsis in comparatives, a specific type that warrants mention is comparative deletion. Comparative deletion (Bresnan 1973; Lechner 2019: 624-656) occurs when the gradable property inside the comparative remains unpronounced<sup>129</sup>:

46) The ship is longer than the train is (\*long).

As with ellipsis in general, comparative deletion is subject to various approaches that analyze the gradable property in different ways, ranging from accounts that hypothesize syntactic ellipsis (Hankamer 1973, among others) to those that propose a semantic analysis (Lerner and Pinkal 1995; among others).<sup>130</sup>

Interestingly, this phenomenon (along with subdeletion) is distinguished from “comparative ellipsis” (Napoli 1983: 675), which is said to differ from the former two in that it involves the deletion of material outside the compared constituent (Napoli 1983: 675):

47) Mary wrote more books than John.

Comparative ellipsis

In sentence 47, the deletion process affects not only the compared constituent (‘x books’), but also the predicate ‘wrote’.

In conclusion, ellipsis is attested in comparative structures to the same extent as in other clausal domains. Phenomena such as VP-ellipsis, gapping, and nominal ellipsis are likewise observed in comparatives. Furthermore, comparatives exhibit a distinct type of ellipsis, comparative deletion, which accounts for the omission of the gradable property within the comparative construction.

---

<sup>129</sup> The example is taken from Lechner (2019:624).

<sup>130</sup> In transformational literature, the term *comparative deletion* also occurs with the label subdeletion, which is considered a distinct phenomenon. When referring to comparatives ‘subdeletion’, Napoli (1983) refers to the deletion of the “quantifier-like element of the compared constituent under identity with the quantifier-like element in the head of the comparative clause” (Napoli 1983:675):

- |   |                      |
|---|----------------------|
| (i) Mary wrote more books than John wrote 0 <sub>CD</sub> .           | Comparative deletion |
| (0 <sub>CD</sub> = x many books)                                      |                      |
| (ii) Mary wrote more books than John wrote 0 <sub>Sub</sub> articles. | Subdeletion          |
| (0 <sub>Sub</sub> = x many)   |                      |

### 3.3 Ellipsis in Diachrony

When examined from a diachronic perspective, ellipsis emerges as a phenomenon for which there exists only “relatively small body of research literature with regards to the historical component of ellipsis” (Gergel and Cyrino forth: 1). The scarcity of contributions to the diachronic analysis of ellipsis can be attributed to several factors.

First, as noted by Gergel and Cyrino (forth: 5), ellipsis “is synchronically rather well-studied, as is the wealth and diversity of ellipsis types”. Indeed, most research on ellipsis has predominantly focused on synchronic theoretical approaches ([see Section 3.1.5](#)), particularly concerning the structural properties of ellipsis and the mechanisms underlying antecedent recovery and licensing. Another major line of inquiry concerns the relationship between ellipsis and cognitive studies, with a specific focus on its processing in neurodegenerative conditions such as dementia (Moshtaghi, Dehghan and Raghieb Doust, 2020).

Second, investigating ellipsis from a diachronic perspective requires a corpus-based approach, thereby precluding reliance on native speakers’ linguistic intuitions. This reliance on textual evidence presents a fundamental challenge: extracting from the data something that is, by definition, absent. Although this issue also arises in corpus-based synchronic analyses, it becomes especially critical in the study of historical languages, where grammaticality judgements cannot be elicited from speakers and analysis must depend exclusively on textual attestations. This constraint gives rise to two main challenges. The first follows the principle of “the more, the better”, which assumes that increasing the size of the dataset, and thus expanding the textual evidence available, yields more reliable results. However, this assumption has been called into question (Anthony 2013). The second challenge concerns data accessibility and the need for annotations that facilitate and accelerate the analytical process (McEnery 2012). In the case of ellipsis, this specifically entails the availability of corpora that support systematic queries for the identification elliptical structures<sup>131</sup>.

To further complicate the scenario, in the case of ancient texts transmitted through multiple editions or versions, it is often uncertain whether an omission constitutes a genuine instance of ellipsis, that is, a syntactically and pragmatically motivated omission, or rather a lacuna, a gap resulting from a copyist’s error rather than an intentional omission. Resolving this issue requires the expertise of philologists, who establish critical editions on the basis of rigorous philological analysis.

---

<sup>131</sup> This point and its complexity have also been addressed by Gergel and Cyrino (forth :30) from a cross-linguistic perspective. The authors highlight the need to include ellipsis in major typological databases such as WALS (Dryer & Haspelmath 2013) or APiCS (Michaelis et al. 2013), where it has not been incorporated so far. They also warn that corpus annotations reflect theoretical assumptions that influence both data collection and analysis, specifically referring to the Penn Treebank of English and other languages, thus aligning with what has been stated in Section 3.2.

For these reasons, the diachronic study of ellipsis remains relatively underexplored, with only a few notable exceptions, which will be briefly outlined here.

A recent diachronic investigation of ellipsis is offered by Gergel and Cyrino (forth), who examine types of verbal ellipsis in selected Romance and Germanic languages. Their study, however, excludes ellipsis types that are “quite widespread cross-linguistically (for example, fragments, gapping, sluicing)” owing to the lack of “an empirical basis to study their grammatical properties over time” (Gergel and Cyrino, forth: 6). Notably, although several Romance languages are included in their analysis, Old Italian is not among them (a brief overview of the literature dedicated to ellipsis in Old Italian is provided in [Section 3.3.1](#)). Among the specific predicate ellipsis phenomena studied in diachrony, gapping in ancient Indo-European languages, and its possible reconstruction in Proto-Indo-European, has been analyzed by Krisch (2009). His work highlights the interplay between syntax and functional sentence perspective in determining gapping (Krisch 2009:212).

While predicate and verbal ellipsis have received relatively little scholarly attention, object ellipsis (or object omission) has been more extensively investigated. Diachronic studies have been conducted on several historical languages, including Latin (Luraghi 1997; Luraghi 2004), Ancient Greek (Gaeta et al. 2001; Luraghi 2003; Keydana and Luraghi 2012) and Vedic Sanskrit (Keydana and Luraghi 2012), among others. In these languages, null objects display a tendency to be replaced by overt pronominal forms (Haug and Luraghi 2012), a development that aligns with the rise of pronominal clitics, as observed in Romance languages, in Greek and in English. This shift appears to correlate with factors such as semantic transitivity (e.g., the frequent use of generic objects and the loss of internal cognate objects), information structure (particularly focal status), and telicity. Additionally, coordination emerges as a syntactic environment that facilitates the licensing of null objects.

With regard to the evolution of the Romance languages, Luraghi (1997) conducted an in-depth corpus-based study of object omission in Latin. Her analysis reveals that object omission follows systematic patterns rather than occurring randomly. Specifically, it is obligatory in coordinated structures when the object is present in the first clause, due to syntactic constraints. By contrast, cases of omission driven by high topic continuity are less predictable and are influenced by an author’s stylistic choices. Furthermore, object omission is attested in contexts where the object is indefinite, in constructions involving nominal forms of the verb (such as gerundives and participles), in references to whole sentences, and in ablative absolute constructions. Luraghi (1997) also suggests that the likelihood of omission decreases depending on the syntactic properties of the antecedent.

The following Section 3.3.1 reviews the literature on ellipsis in Old Italian.

### 3.3.1 Ellipsis in Old Italian

An overview of ellipsis in Old Italian is provided by Marelo (Marelo 2010:1369-1387). Marelo (2010) observes that, while many elliptical structures in Old Italian align with those found in Modern Italian, Old Italian also exhibits distinctive features, including omissions that would be considered ungrammatical in Modern Italian. These include:

- i) the use of bare auxiliary (without the lexical verb) in coordination or responses;
- ii) the omission of the lexical verb;
- iii) the omission of clitics in specific contexts;
- iv) subject and object ellipsis.

These features will be briefly reviewed in the following subsections.

### i) Bare auxiliaries in coordination or answer

Old Italian allows the auxiliary verb to appear without the lexical verb and its complements, particularly in coordinated structures or response contexts. An example of lexical verb omission is illustrated in sentence (48):<sup>132</sup>

48) – *Che è ciò, messer Rinieri, che voi non siete partito di Sardegna? – Certo – disse messere Rinieri, – sì, sono; ma sono tornato, per li scappini delle calze.*

‘– What is this, messer Rinieri, that you have not departed from Sardinia? – ‘Certainly’ – said messer Rinieri, – **‘yes, I have**; but I have returned, by the foot of my stockings.’

Anonymous, *Novellino*, 1315

In this example, the lexical verb *partire* ‘to depart’ is omitted in Messer Rinieri’s response, which consists only in the affirmative *sì* ‘yes’ and the bare auxiliary *sono* ‘am’ (lit. ‘yes, am’). This omission would render the sentence ungrammatical in Modern Italian, where the inclusion of the lexical verb is mandatory.

### ii) The omission of the lexical verb

Additionally, the lexical verb can also be omitted with modal or phasal verbs, as discussed by Marelo (2010:1376). This phenomenon is exemplified in sentence (49):<sup>133</sup>

49) *E poco tardante, la reina venne alla fontana, ed a ventura li avvenne un bel pensiero: che guardò il pino, e vide l'ombra più spessa che non solea. Allora la reina dottò, e dottando ristette.*

---

<sup>132</sup> [http://gattoweb.ovi.cnr.it/\(S\(sagxv42dhfge42xbmirhpu0w\)\)/CatForm21.aspx](http://gattoweb.ovi.cnr.it/(S(sagxv42dhfge42xbmirhpu0w))/CatForm21.aspx)

<sup>133</sup> [http://gattoweb.ovi.cnr.it/\(S\(zqtm4g0vcwbxlrffoyefhkvz\)\)/CatForm21.aspx](http://gattoweb.ovi.cnr.it/(S(zqtm4g0vcwbxlrffoyefhkvz))/CatForm21.aspx)

‘And a little later, the queen came to the fountain, and by chance a good thought occurred to her: she looked at the pine tree, and saw the shadow thicker than it **used (to be)**. Then the queen suspected, and suspecting rested’

Anonymous, *Novellino*, 1315

As in sentence (48), the omission of the lexical verb *essere* ‘to be’ in the construction *solea* ‘used’ in sentence (49) would render the sentence ungrammatical in Modern Italian.

### iii) Omission of clitics in specific contexts

In coordinated structures with conjunctions or disjunctions, the second clitic pronoun can optionally be omitted in Old Italian, as illustrated in Example (50)<sup>134</sup>:

50) *E io vi dico e \_prometto che se queste pene e fatiche in pace portate...*

‘And I tell you and promise (you) that if these pains and hardships are endured in peace...’

Bono Giamboni, *Il Libro de' Vizî e delle Virtudi*, 1292

This omission occurs under conditions of structural symmetry, meaning that it is found when the subject and its complements remain the same (Egerland et al. 2010: 464). Additionally, semantic affinity between the coordinated verbs is required, with the verbs typically describing different aspects of the same event. The omission of the clitic is only attested in simple tenses, whereas in complex tenses, the clitic is generally present. However, the omission of the clitic is optional, as minimal pairs of both presence and absence of the clitic can be found in the same text, as shown in sentences (51)<sup>135</sup> and (52):<sup>136</sup>

51) *veggendolo il duca, gli fae lo maggio onore del mondo, e molto **l'abbraccia e bacia** \_  
teneramente*

‘seeing him the duke, does him the greatest honor in the world, and much embraces **him** and kisses (him) tenderly’

52) *la reina Ginevra sì **la abbraccia e baciala** strettamente;*

‘the queen Ginevra so embrace **her** and kiss **her** strictly;’

Anonymous, *Tavola ritonda o l'Istoria di Tristano*, sec. XIV

The two sentences depict a similar scene of hugging and kissing. However, while sentence (51) does not include the object clitic *lo* ‘him’ with the second coordinated verb *bacia* ‘kisses’, sentence (52)

<sup>134</sup> [http://gattoweb.ovi.cnr.it/\(S\(fr2esvpu4wc52l0ckkzdn4w\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(fr2esvpu4wc52l0ckkzdn4w))/CatForm22.aspx)

<sup>135</sup> [http://gattoweb.ovi.cnr.it/\(S\(nzvudiubnrbcng5df25qu2l3\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(nzvudiubnrbcng5df25qu2l3))/CatForm22.aspx)

<sup>136</sup> [http://gattoweb.ovi.cnr.it/\(S\(lo4frgw3tvnvenupftfqwz5y\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(lo4frgw3tvnvenupftfqwz5y))/CatForm22.aspx)

exhibits a different pattern, presenting the object clitic *la* ‘her’ in enclitic position of the verb *bacia* ‘kisses’. This contrast shows the optionality of clitic omission in coordinated structures.

#### iv) Subject and object ellipsis

Regarding subject ellipsis, omissions in Old Italian generally align with those in Modern Italian, with one key difference: while the subject can be omitted in main clauses, it cannot be omitted in subordinate clauses (Benincà and Poletto 2010:66-67), as shown in sentence (53)<sup>137</sup>:

53) *cosìe l'anima muta la sua forza per la propietade di quello corpo a cui **ella** si congiunge.*

‘Thus, the soul changes its strength according to the nature of the body to which **it** is joined.’

Brunetto Latini, *Rettorica*, 1260-61

Furthermore, subject omission in main clauses is possible when the pronominal subject is the theme of the sentence, a phenomenon referred to as *tema nullo* ‘null theme’ (Benincà and Poletto 2010:66-67):<sup>138</sup>

54) *E quella ripensa e riguardollo e rafiguollo e venelene sì gran vergogna ch'ella nol potte patire. Incontanente **morìo**. Questo Secondo, vegendo che per lo suo parlare la madre era morta, ...*

‘And she reflected and looked at him and envisioned him, and such great shame came over her that she could not bear it. (He) **died** immediately. This Second, seeing that his mother had died because of his words, ...’

Anonymous, *Fiori e vita di filosafi e d'altri savi e d'imperadori*, 1271-1275

Another notable difference from Modern Italian is the possibility of omitting the object in coordinated structures. Specifically, in the second conjunct, the object clitic is optionally omitted (Example 51). This property is not limited to object clitics but applies more generally to clitic pronouns, including locative and reflexive clitics, as previously discussed (see (iii): omission of clitics in specific contexts). The discussion of object ellipsis will be resumed in [Section 3.3.1.2](#).

The following Sections 3.3.1.1 and 3.3.1.2 examine in greater depth: i) verbal ellipsis, and ii) object ellipsis.

#### 3.3.1.1 Verbal ellipsis in Old Italian

<sup>137</sup> [http://gattoweb.ovi.cnr.it/\(S\(ogxjb5wyantifa3yuglvyp0f\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(ogxjb5wyantifa3yuglvyp0f))/CatForm22.aspx)

<sup>138</sup> [http://gattoweb.ovi.cnr.it/\(S\(bpcoljwxhihiwsejpfy1ugds\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(bpcoljwxhihiwsejpfy1ugds))/CatForm22.aspx)

When speaking of verbal ellipsis in Old Italian, cases of gapping are attested (Marello 2010: 1375; Molinelli 2010: 271) and they are interpreted as syntactic parallelism that consents the omission of the second verb, as illustrated in (55)<sup>139</sup> and (56):<sup>140</sup>

55) *Nel riposare, la sera, e' cavalieri si cominciaro a vantare, in sull'allegrezze loro, chi [SI VANTAVA] di bella giostra, chi [SI VANTAVA] di bello castello, chi [SI VANTAVA] di bello astore, chi [SI VANTAVA] di ricca ventura, e 'l cavaliere non si poteo tenere che non si vantasse ch'amava sì bella donna*

'As they rested, in the evening, the knights began to boast, in their merriment, who [WOULD BOAST] of fine jousting, who [WOULD BOAST] of fine castle, who [WOULD BOAST] of fine goshawk, who [WOULD BOAST] of rich fortune, and the knight could not be held who did not boast that he loved such a beautiful woman'

Anonymous, *Novellino*, 1315

56) *Sì com' altr' uomini vanno, / ki [VA] per prode e chi [VA] per danno, / per lo mondo tuttavia, / così m'andava l'altra dia...*

'Just as other men go, / some [GO] for gain and some [GO] for loss, / throughout the world, / so I was going the other day...'

Anonymous, *Detto del gatto lopesco*, 1275

In both examples, the elided verbs are represented in square brackets. As it is stated in Marello (2010:1376), sometimes the border between coordination and ellipsis is difficult to define.

Predicate ellipsis is not limited to coordination but also occurs in comparative structures (Marello 2010:1381; Belletti 2010:1138), where the predicate of the comparative clause can be omitted due to syntactic parallelism with the predicate of the main clause. Interestingly, Pelo (2012: 443) analyzes as elliptical comparatives also cases where the second element of the comparative is nominal, prepositional, or an adjectival phrase, thus implying an elided predicate. Sentence (57)<sup>141</sup> illustrates an instance of verbal ellipsis in a comparative structure. Notably, ellipsis appears in the first redaction of the text (57), whereas the vulgata redaction does not exhibit ellipsis (example 58)<sup>142</sup>:

57) *Et Marcho rispuose: - E' non è altro, se non che tue trovasti p[i]ue delli tuoi che io delli miei-*

'And Mark answered: - It is not for anything else, except that you found more of yours than I of mine -.'

first red.

<sup>139</sup> [http://gattoweb.ovi.cnr.it/\(S\(pgexhdzhtihmk5qhklj3zzmm\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(pgexhdzhtihmk5qhklj3zzmm))/CatForm22.aspx)

<sup>140</sup> [http://gattoweb.ovi.cnr.it/\(S\(mjo3ixju31sfcmcgx5px4naw\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(mjo3ixju31sfcmcgx5px4naw))/CatForm22.aspx)

<sup>141</sup> [http://gattoweb.ovi.cnr.it/\(S\(5cjr1pdhtg34jphqpqqezlbk\)\)/CatForm21.aspx](http://gattoweb.ovi.cnr.it/(S(5cjr1pdhtg34jphqpqqezlbk))/CatForm21.aspx)

<sup>142</sup> [http://gattoweb.ovi.cnr.it/\(S\(5cjr1pdhtg34jphqpqqezlbk\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(5cjr1pdhtg34jphqpqqezlbk))/CatForm22.aspx)

58) *E Marco rispuose: - Non è per altro, se non che tu trovasti più de' tuoi che io non trova' de[l]li miei-*

'And Mark answered: - It is not for anything else, except that you found more of yours than I found of mine -.'

vulgata red.

Anonymous, *Novellino*, 1315

Ellipsis is also attested in relative clauses, connected with the ellipsis of the relative pronoun itself (De Roberto 2012: 220-221).

### 3.3.1.2 Object ellipsis in Old Italian

Omission of the object is attested in Old Italian. Among the contexts in which it occurs, one notable case is in responses to questions. In this scenario, the omitted element is typically thematic, privileging the permanence of the focus (Marello 2010:1380). An example is reported in (59):

59) *-Or sè tu ancor qui? non avestù la torta?- Messere sì: ebbi.* –

“‘Now are you still here? did you not have the cake?’ ‘Messer yes (I) had.’”

Anonymous, *Novellino*, 1315

In example (59), the thematic element *la torta* ‘the cake’ is omitted in the response. This contrasts with Modern Italian, where, to be grammatical, the object must be explicitly expressed, either in its pronominal form (*la*, ‘it.SG.FEM’) or in its full lexical form (*la torta*, ‘the cake’). To the best of my knowledge, no in-depth or systematic study of object omission in Old Italian has not been conducted. Luraghi (1998), however, offers preliminary insights into its evolution from Latin to Old Italian, noting that the omission of the direct object gradually decreases from Latin to Old Italian and further into Modern Italian. In Latin, three types of object omission were possible: i) free omission, ii) omission in coordinate with multiple constituents; and iii) omission in coordination with de-emphatic (i.e., non-accented) pronouns. In Old Italian, only omission in coordination contexts remained possible. Modern Italian further restricts this pattern, allowing object omission exclusively in coordination with de-emphatic pronouns. As Luraghi (1998) points out, this decrease in object omission is not specific to direct object but reflects a broader pattern affecting clitics in general. Similar restrictions are observed with reflexive and locative clitics. This trend aligns with the grammaticalization of Romance clitics (Giacalone Ramat 1990), which display less freedom in the syntactic position, showing a behavior that led to compare Romance clitics to affixes rather than clitics (Miller 1992).

### 3.4 Ellipsis and Syntax

A preliminary question to be asked before engaging with the analysis of ellipsis is the following: *Is ellipsis a matter of syntax?* While this question may initially appear rhetorical, it raises important considerations. Yet, to approach it more effectively, the question might be better reversed: *Can syntax (and syntactic representation) do without ellipsis?* Consider the following sentence elliptical sentence:

60) ‘Jones likes seafood and Smith bread.’

Within a model designed to capture syntactic relations, it is undeniable that in a sentence like (60) ‘Smith’ assumes the syntactic role of subject and ‘bread’ that of object, even though the verb is omitted. The existence of these syntactic relations legitimizes the existence of ellipsis, which operates at the syntactic level and should therefore be accounted for within syntactic frameworks. From this perspective, several important questions arise: *Should ellipsis be represented in the syntactic structure? And if so, in what way?*

The answer to the first question is straightforward: if ellipsis is a syntactic phenomenon, it naturally follows that, for a model aiming to capture syntactic information, it should be represented in the syntactic structure. In contrast, the second issue to be addressed poses a more complex challenge: *How is ellipsis represented in the syntactic structure?* A reasonable answer is that, in order to represent ellipsis, the missing element is accounted for by introducing syntactic empty nodes. Although this solution may be controversial in frameworks that prefer to represent only positive evidence (i.e., words that are actually present and uttered in a sentence), if the aim is to capture the syntactic level, empty slots cannot be disregarded in the analysis of ellipsis.

In the following sections, the treatment of ellipsis will be examined within dependency syntax, a framework that eschews the use of empty nodes (Section 3.4.1), and within the Universal Dependencies framework (Section 3.4.2).

#### 3.4.1 Ellipsis and Dependency Syntax

In dependency syntax, ellipsis has not been investigated as extensively as in constituency-based accounts. One possible reason for this is that dependency syntax does not typically incorporate empty nodes and is even less aligned with the notion of movement.

Nevertheless, the concept of empty nodes is not entirely absent from dependency syntax and has been explored in various studies. In his theory of transfer, Tesnière (2015) introduced elements into his *stemmas*<sup>143</sup> that could be interpreted as empty nodes (Osbourne 2018:143).

Similarly, within the dependency-based Meaning-Text Theory, Mel'čuk (2009) argues for the necessity of empty nodes<sup>144</sup> for analyze cases of gapping, such as “*Alan went to Paris and Leo to Coruña?*”, where “*and Leo to Coruña?*” is defined as a ‘mutilated’ expression (Mel'čuk 2009:82).

Another contribution to the discussion of empty nodes and ellipsis comes from Hudson (2007:172-181), who introduces the concept of “unrealized words”, referring to covert words that lack pronunciation or spelling, with their unique distinction from overt ones being their inaudibility.

An additional significant perspective on ellipsis in dependency grammar (DG) is provided by Plewnia (2013), who examines ellipsis (alongside anacoluthia) in the context of German verbal omission. Plewnia critiques the result-oriented approach of DG and its de-temporalization<sup>145</sup>, emphasizing the importance of considering the listener's grammatical knowledge, particularly in elliptical constructions<sup>146</sup>. He argues that incorporating the predictability of syntactic projections and the likelihood of sentence continuation is crucial for accurately representing ellipsis. To formalize these constructions, Plewnia introduces empty nodes within the structure.

Osbourne's work (Osbourne et al. 2012; Osbourne, 2018; Osbourne 2019) offers another valuable contribution to the study of ellipsis in DG, employing the notion of *catena* to analyze elliptical constructions. He acknowledges the presence of an empty structure in the form of a null node, even though this null node “does not have the same syntactic status as the corresponding overt words” (Osbourne 2018:148). Regarding recoverability, that is, the relationship between the elided material and its antecedent, Osbourne argues that ellipsis mechanisms are heterogeneous, varying by ellipsis type: they are flexible for VP-ellipsis, syntactic for gapping and stripping, and pragmatic for left-edge ellipsis. As for licensing, i.e., the conditions under which ellipsis can or cannot occur, Osbourne identifies *catena* as the fundamental syntactic unit of ellipsis. He proposes that a necessary (though not sufficient) condition for ellipsis is that the elided material forms a *catena*. One advantage of Osbourne's analysis is that the concept of *catena* eliminates the need for movement, a mechanism used in constituency-based accounts but absent in dependency-based frameworks. However, this

---

<sup>143</sup> Refer to Stemma 331 (Tesniere 2015:523).

<sup>144</sup> “The reason for which I prefer to keep an explicit indication of ellipsis in the SSyntS of a sentence is my intuitive feeling that a complete sentence and a sentence with ellipsis are syntactically different and, therefore, this difference should be reflected at the SSynt-level.” (Mel'čuk 2009:82).

<sup>145</sup> “Die Dependenzgrammatik ist eine Resultatsgrammatik, sie bietet Ex-post-Analysen von Äußerungsprodukten und ist damit sozusagen entzeitlicht.” (Plewnia 2013:229)

<sup>146</sup> Plewnia specifically focused on *right-edge* ellipsis, a term defined in Osbourne (2018:146).

theory does not account for ellipsis in fragments or echo questions (Osbourne 2018:160), as these phenomena fall outside the scope of the *catena*-based framework.

### 3.4.2 Ellipsis in Treebanks

The representation of ellipsis in treebanks varies according to the formalism adopted.

In constituency-based treebanks, such as the Penn Treebank (Marcus et al. 1993; Taylor et al. 2003), ellipsis is explicitly annotated through the insertion of empty nodes. Penn Treebank marks null elements by co-indexing these null elements with the lexical material for which the null element stands (Markus et al. 1994: 116). As stated in Markus et al. (1994), the “scheme happens to use two symbols for null elements: \*T\*, which marks WH-movement and topicalization, and \* which is used for all other null elements”, as shown in Image 32:

```
(SBARQ (WHNP-1 What)
      (SQ is
        (NP-SBJ Tim)
        (VP eating
          (NP *T*-1)))
      ?)

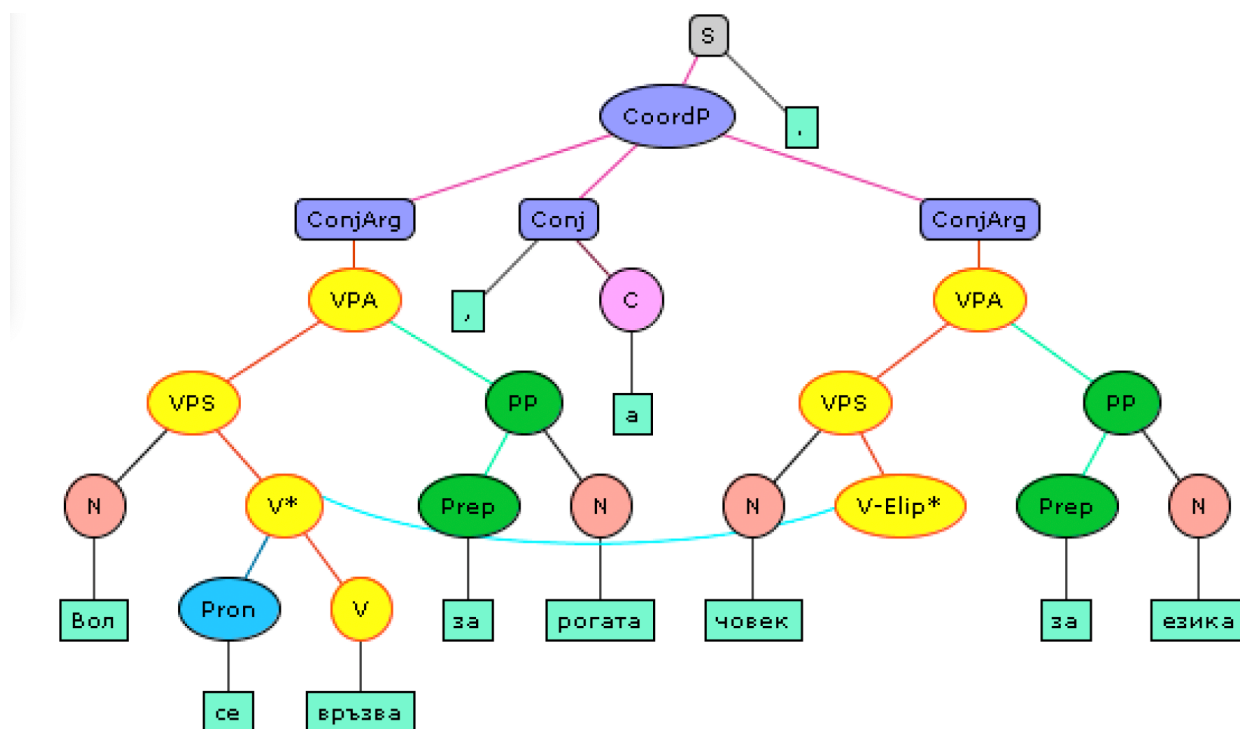
Predicate Argument Structure:
eat(Tim, what)
```

**Image. 32** Penn Treebank’s annotation of null elements

In 32, the use of \*T\*, standing for trace, marks the position originally occupied by the NP that has been moved to the WHNP position ‘What’ (for a detailed analysis of the tree, see Marcus et al. 1994:116). The annotation of empty categories facilitates their automatic retrieval, thereby supporting research in this area. For example, in his analysis of VP ellipsis, Hardt (1997) made use of the -NONE- category in the Wall Street Journal section of the Penn Treebank, which enabled the systematic identification of such instances (Hardt 1997: 526).

A solution that may be considered as a bridge between constituency- and dependency-based approaches is offered by the BulTreeBank (Osenova et al. 2004), which is grounded in Head-Driven Phrase Structure Grammar (Pollard and Sag 1994). In the BulTreeBank, ellipsis is encoded (Osenova et al. 2004: 69-74) by means of empty and coreferential nodes, and is defined as “a slot in the elliptical phrase which determines what information is missing and where it can be copied from within the context” (Osenova et al. 2004: 69). A fine-grained distinction is drawn between sentence-internal ellipsis, in which a reference mechanism “establishes a connection with the overt material that supports the ellipsis” (Osenova et al. 2004: 69), as illustrated in Image 33, and discourse ellipsis, i.e.

cases that are not recoverable within the sentence. The latter are further subdivided into world-knowledge-recoverable and discourse-recoverable ellipsis. An even more fine-grained level of annotation records morphological variation (when present) between the elided material and its copied counterpart, which is specified through the attributes *type*, *gram* and *form* (see Osenova et al. 2004: 69;72 for details).

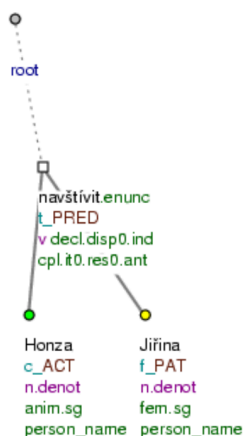


**Image 33.** Annotation of an elliptical sentence in BulTreeBank (Osenova et al. 2004: 70): “Вол се връзва за рогата, а човек за езика” ‘An ox is tied by its horns, but a man by his tongue’

As mentioned in 3.4.1, the representation of ellipsis in dependency treebanks presents a greater challenge. This difficulty arises from the principle that, in dependency-based annotation, the number of nodes in the tree should correspond exactly to the number of tokens in the sentence, thus precluding the use of empty nodes. An example of dependency treebank is the Prague Dependency Treebank (Hajič et al. 2017), which handles ellipsis explicitly through a dedicated annotation layer, the tectogrammatical layer (t-layer), which represents a deep syntactic<sup>147</sup>. The treatment of ellipsis in Prague Dependency Treebank is accurately detailed in Mikulová (2014), in which is carefully explained how the t-layer is linked with the a-layer (analytical-layer, representing the surface syntactic) to convey the “transition from the (linguistic) meaning of a sentence to its form, transition from the deep syntactic structure to the surface expression” (Mikulová 2014: 131). When a direct

<sup>147</sup> <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html>

counterpart at the a-layer is absent, additional nodes are introduced in the t-layer. These are classified as either copy, “a t-node that has the same values of certain attributes (...) as an expressed t-node”, or substitute, defined as “an artificial t-node” (Mikulová 2014: 133). When applicable, co-reference is indicated in the tree by an arrow linking the added t-node to the t-node representing the expressed unit (the anaphor). Image 34 illustrates a case of verbal ellipsis, where the t-node of the elided verb *navštívil* ‘visited’ is represented by a squared node.



(*Jirka navštívil Marii.*) *Honza Jiřinu.* (=lit. (*George visited Mary.*) *John Henriette.*)

**Image 34.** Dependency trees with an elided verb<sup>148</sup>

To conclude this overview, Section 3.4.3 provides a detailed discussion of the formalism adopted by the state-of-the-art dependency framework, Universal Dependencies.

### 3.4.3 Ellipsis in Universal Dependencies

Ellipsis is not explicitly accounted for in the basic UD framework<sup>149</sup>. At this level of annotation, UD does not seek to systematically signal the presence of ellipsis; instead, it only intervenes in cases where ellipsis lead to a problematic syntactic structure. The guidelines propose three possible approaches for treating elliptical constructions in both nominal and clausal contexts (as specified in the online ellipsis guidelines, cited below)<sup>150</sup>:

1. “If the elided element has no overt dependents, we do nothing.
2. If the elided element has overt dependents, we promote one of these to take the role of the head.

<sup>148</sup> <https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/ch06s12s01.html#pic2312elipsa1>, access 07/06/2025.

<sup>149</sup> <https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>, access 05/11/24.

<sup>150</sup> <https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>, access 05/11/24.

3. If the elided element is a predicate and the promoted element is one of its arguments or adjuncts, we use the *orphan* relation when attaching other non-functional dependents to the promoted head.”

The annotation of the second scenario (promotion) and the third scenario (orphan) is described in detail in Sections 3.4.3.1 and 3.4.3.2, respectively. At first glance, it is clear that none of the three annotation strategies explicitly mark the presence of ellipsis. Rather, they provide workarounds for representing syntactic structures without directly encoding ellipsis, which ultimately leads to a loss of information about the phenomenon. In the first scenario, ellipsis is disregarded altogether, with no additional information provided. In the second scenario, it is obscured through the promotion of elements, thereby indirectly suggesting the absence of omitted material. In the third scenario, ellipsis is only partially captured through the `orphan` relation, which allows for the inference of an elliptical predicate. However, even in this case, the relation is not intended to explicitly indicate ellipsis but rather to correct the atypical syntactic structure it produces.

This approach to annotating elliptical constructions makes the retrieval of ellipsis considerably more difficult, rendering its systematic extraction particularly challenging (see Chapter 4 for details on the extraction process). The basic UD guidelines do not offer an explicit rationale for the lack of systematic treatment of ellipsis as a syntactic phenomenon within the annotation scheme. One possible explanation is that the basic UD layer does not employ empty nodes, delegating such annotations to the enhanced dependency layer (see Section 3.4.3.3). Nevertheless, the marking of ellipsis does not necessarily require the use of empty nodes in the syntactic structure. The feasibility of alternative approaches is discussed in Corbetta et al. (2025 a)<sup>151</sup>, where an implementation of the annotation scheme for basic UD, based on the use of subtypes and avoiding empty nodes, is proposed.

### 3.4.3.1 Promotion

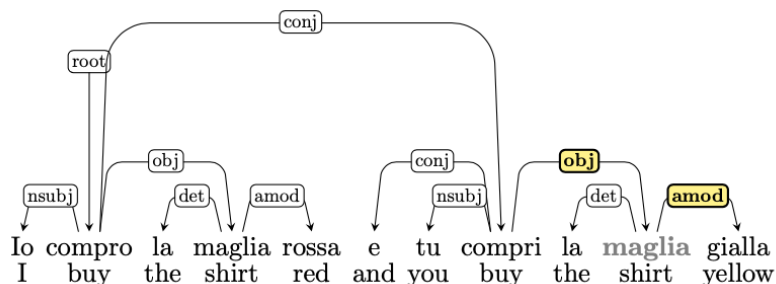
This section provides a detailed account of the second strategy outlined in the UD guidelines: promotion. Under this approach, a dependent of the elided element is promoted to replace the omitted node and thereby assumes the role of the head. In cases of verbal ellipsis, the promoted element becomes the head of the clause, whereas in nominal ellipsis, it serves as the head of the nominal phrase. In the case of **nominal ellipsis**, where the omitted element is typically a noun, promotion follows a predefined hierarchy<sup>152</sup>: `amod` (adjectival modifier) > `nummod` (numeral modifier) > `det`

---

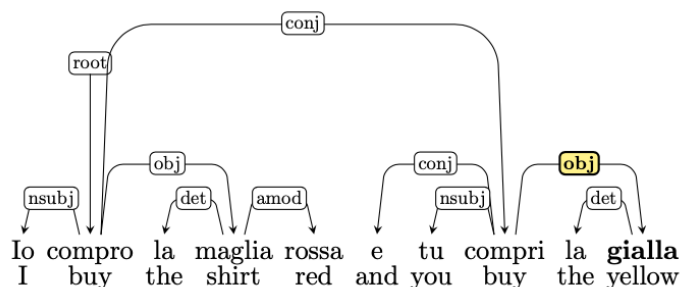
<sup>151</sup>[https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general\\_meetings:3rd\\_unidive\\_general\\_meeting:23\\_how\\_to\\_ellipsis\\_a\\_proposal\\_poster\\_material.pdf](https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general_meetings:3rd_unidive_general_meeting:23_how_to_ellipsis_a_proposal_poster_material.pdf) and [https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general\\_meetings:3rd\\_unidive\\_general\\_meeting:23\\_how\\_to\\_ellipsis\\_a\\_proposal.pdf](https://unidive.lisn.upsaclay.fr/lib/exe/fetch.php?media=meetings:general_meetings:3rd_unidive_general_meeting:23_how_to_ellipsis_a_proposal.pdf)

<sup>152</sup><https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>

(determiner) > nmod (nominal modifier) > case. According to this hierarchy, when the noun (i.e., the head of the nominal phrase) is elided, the element assigned the amod deprel (typically an adjective) becomes the new head. If no adjective is present, the head is the element bearing the nummod relation (typically a numeral), and so forth, following the established ranking. Image 35 and Image 36 provide examples of Italian sentences with and without nominal ellipsis: *Io compro la maglia rossa e tu compri la (maglia) gialla* ‘I buy the red shirt, and you buy the yellow (shirt)’.



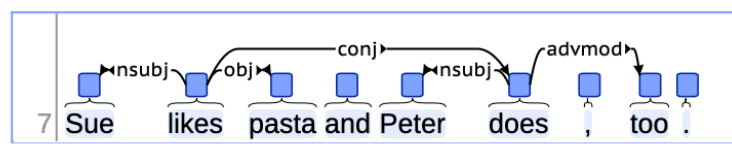
**Image 35.** Italian sentence without nominal ellipsis



**Image 36.** Italian sentence with nominal ellipsis and promotion

Image 35 illustrates the non-elided nominal (*la maglia gialla* ‘the yellow shirt’, lit. ‘the shirt yellow’), where the noun *maglia* ‘shirt’ receives the object relation (obj), depending on the verb *compr-i* ‘buy.2SG’. The noun *maglia* further takes two dependents: the article *la* ‘the’, attached with the det relation, and the adjective *gialla* ‘yellow’, attached with the amod relation. Image 36, by contrast, represents the elided nominal (*la gialla* ‘the yellow’), in which the noun *maglia* is omitted. According to the promotion hierarchy outlined above, the amod is promoted to the head of the nominal. Thus, in Image 36, the adjective *gialla* ‘yellow’ is promoted to the head status, assuming the syntactic role that the elided element (*maglia*) would otherwise fulfill (in this case, the obj relation). while also taking the article *la* ‘the’ as its dependent.

With respect to clausal ellipsis, promotion is applied “only if an auxiliary, a copula, or, in the case of an infinitival construction, a marker is present”<sup>153</sup>. An example of an elliptical clause annotated through promotion, provided in the guidelines, is reported in the following Image 37 (‘Sue likes pasta and Peter does, too.’):



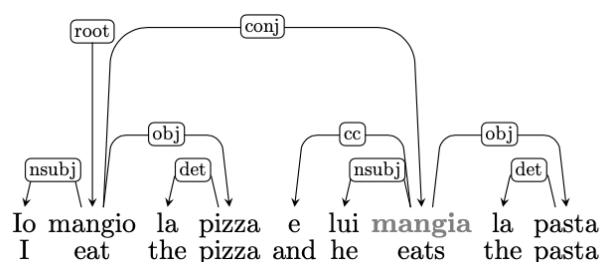
**Image 37.** Promotion in UD

In this sentence, the auxiliary ‘does’ is promoted to the head of the coordinate clause (`conj`), with the leaf nodes depending on it through their canonical relations, namely, `nsubj` (nominal subject) for the node ‘Peter’ and `advmod` (adverbial modifier) for ‘too’<sup>154</sup>.

### 3.4.3.2 Orphan

Turning to the third strategy described in the UD guidelines, the `orphan` relation is particularly relevant for cases of predicate ellipsis. It is used when the promotion of an element would lead to unnatural syntactic relations. In such cases, one element is promoted to be the head of the clause, while the remaining elements are attached to the promoted head via `orphan` `deprel`. The hierarchy for the promotion of orphan dependents is as follows<sup>155</sup>: `nsubj` (nominal subject) > `obj` (object) > `iobj` (indirect object) > `obl` (oblique) > `advmod` (adverbial modifier) > `csubj` (clausal subject) > `xcomp` (open clausal complement) > `ccomp` (clausal complement) > `advcl` (adverbial clause modifier) > `dislocated` > `vocative`.

Images 38 and 39 illustrate Italian sentences with and without an elided predicate: *Io mangio la pizza e lui (mangia) la pasta* ‘I eat pizza and he (eats) pasta’:

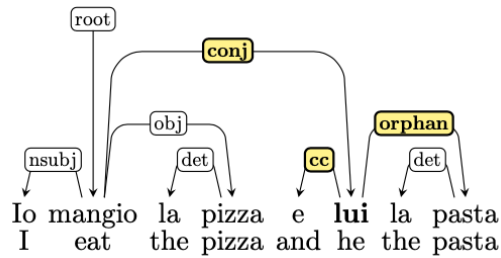


<sup>153</sup> <https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>

<sup>154</sup> Constructions such as the one shown in Image 37 do not occur in Modern Italian, where clauses with a bare auxiliary are unattested. However, as discussed in Section 3.3.1, in Old Italian, the presence of bare auxiliaries is possible.

<sup>155</sup> <https://universaldependencies.org/u/overview/specific-syntax.html#ellipsis>

**Image 38.** Italian non-elided sentence.



**Image 39.** Italian elided sentence.

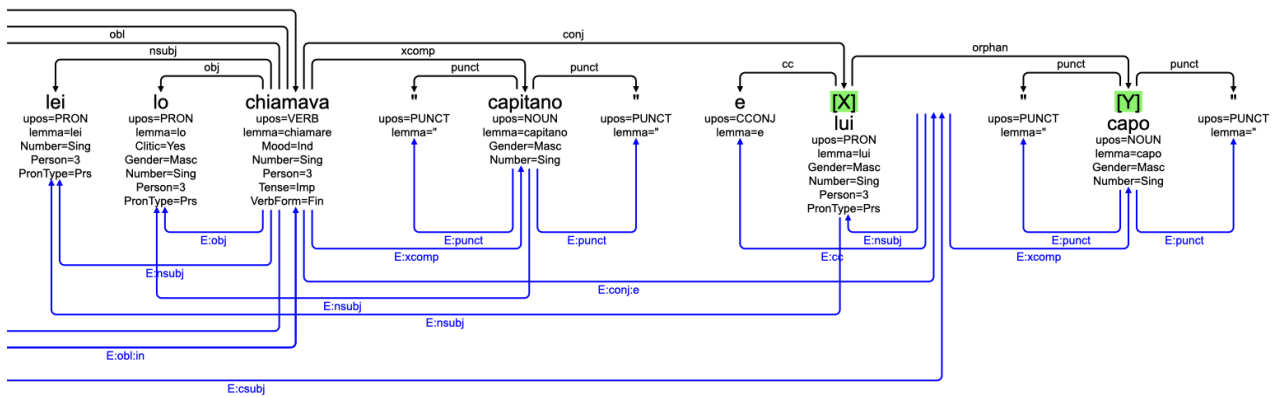
Image 38 does not represent predicate ellipsis but rather illustrates a coordination relation between the two verbs: *mangio* ‘eat’, which is the `root` of the sentence, and *mangia* ‘eats’, which depends on the `root` via a `conj` (conjunction) relation. In the coordinated clause *e lui mangia la pasta* ‘and he eats pasta’, all arguments are attached to the head of the clause, namely the verb *mangia* ‘eats’: *lui* ‘he’ is attached through an `nsubj` relation, while *pasta* is attached as an `obj`. By contrast, Image 39 shows *lui* ‘he’ as the head of the coordinated clause, having been promoted to fill the role of the omitted predicate *mangia* ‘eats’. In this elided structure, what was originally the object of the now-elided predicate is attached to the promoted head *lui* ‘he’ with an `orphan` relation. This mechanism is applied to avoid creating an unnatural syntactic structure, such as one in which a noun (*pasta*) would be attached as an `obj` to another noun (*lui*, ‘he’).

As shown by the comparison of Images 38 and 39, annotating the elliptical clause with the `orphan` relation results in the loss of two syntactic relations: the `nsubj` relation of *lui* ‘he’ and the `obj` relation of *pasta*.

### 3.4.2.3 Enhanced Dependency and empty nodes

As mentioned in [Section 1.4.4](#), UD also includes an enhanced annotation scheme called *Enhanced Dependencies*, in which the presence of empty nodes is attested. In enhanced dependencies, elliptical nodes are reconstructed within the tree structure, and syntactic dependencies that are lost in the basic annotation are restored. Image 40 presents an example of a sentence in the enhanced dependency annotation of the ISDT treebank: *lei lo chiamava “capitano” e lui “capo”* ‘She called him “Captain” and he “Boss”’:<sup>156</sup>

<sup>156</sup> The sentence is taken from the ISDT treebank (instance *isst-tanl-1052*). For a graphical visualization, see: [https://universal.grew.fr/?corpus=UD\\_Italian-ISDT@2.15#](https://universal.grew.fr/?corpus=UD_Italian-ISDT@2.15#).



**Image 40.** ISDT elliptical sentence in Enhanced Dependencies

In the example, enhanced relations are marked in blue and preceded by “E:” before the dependency label. As illustrated in Image 40, an empty node is introduced (i.e., a node with no attested form), and the dependent nodes are assigned their relations in accordance with the presence of the elided verb. Specifically, *lui* ‘he’ is annotated as the subject (E:nsbj) of the empty verbal node, *capo* ‘Boss’ is marked as E:xcomp, functioning as a direct completive of the omitted verb, and the conjunction *e* ‘and’ is attached to the omitted verb.

### 3.5 Summary

This [Chapter](#) provides an overview of ellipsis. [Section 3.1](#) reviews key terminology and major approaches to ellipsis in the literature. [Section 3.2](#) introduces a general classification of ellipsis, distinguishing between predicate, nominal, and clausal ellipsis. Reference tables are provided in the [Appendix](#). [Section 3.3](#) discusses studies on ellipsis in diachrony, with particular attention to research conducted on Old Italian. Finally, [Section 3.4](#) examines the relationship between ellipsis and syntax, focusing on its treatment in dependency syntax, and in particular within UD.

This overview reveals a complex scenario that nonetheless converges on two main issues:

1. What counts as ellipsis in dependency grammar, particularly in light of the extensive literature on ellipsis within constituency grammar?
2. How can ellipsis be retrieved from a dependency treebank that does not encode empty nodes and is there a way to semi-automatically enhance annotation?

The first question is theoretical, as it concerns the definition of ellipsis in dependency grammar and the extent to which empty nodes should be introduced, thereby raising also the issue of the boundary between ellipsis and coordination. The second question is practical, focusing on the extraction of

ellipsis from a dependency treebank and on strategies to improve its annotation in order to facilitate subsequent analysis. An attempt to answer these questions will be provided in Chapter 4, which is devoted to the extraction and annotation of ellipsis, and in Chapter 5, where a contextualized analysis will be offered with reference to Old Italian data.

## 4) Ghostbusting Ellipsis

Dealing with what is absent can at times feel like hunting for ghosts. As discussed in Chapter 3, ellipsis can only be detected and analyzed through its effects on the surrounding material and through the careful reconstruction of what occupies the gap, much like observing and studying a black hole, to paraphrase Merchant's well-known comparison (Merchant 2012:8). As outlined in Chapter 3, attempts to study and classify ellipsis have often led to the development of highly specific taxonomies, with distinct categories designed to capture different manifestations of the phenomenon. Yet the proliferation of approaches, each differing in scope and granularity to meet particular analytical needs, has at times yielded a rich but unbalanced catalogue of ellipsis, often constructed or elicited for the specific purpose of confirming or refuting a given hypothesis. In some cases, the objective has been to identify a unified pattern capable of explaining why instances of ellipsis behave in certain ways, how they interact with their syntactic environments, and whether a coherent set of properties can be used to characterize the phenomenon. Substantial progress has been made in this direction (see Chapter 3).

The contribution of the present research is neither to compile an inventory of ellipsis types by constructing hypothetical scenarios and contexts in which they might occur, nor to identify a single underlying mechanism, together with its case-specific variants, that would provide a comprehensive account of ellipsis in language. The latter undoubtedly represents the long-term objective of broader research efforts pursued in synergy, especially given that ellipsis itself has been characterized in such terms (Kehler 2019). The modest, and admittedly incomplete from a broader scientific perspective, aim of this research is to investigate ellipsis within a specific context, by systematically analyzing a single text and exploring the features that characterize it. While this scope may appear limited, and indeed it is, it nonetheless represents a first step toward further, comparable studies. Ultimately, progress arises from the cumulative efforts of a research community rather than from a single, all-encompassing solution.

In the present study, the analysis focuses on a single text, the *Divine Comedy* (Section 1.1), with the aim of systematically identifying and examining instances of ellipsis. The investigation is not carried out manually. Instead, systematic methods are developed and tested through a detailed analysis. Elliptical phenomena are examined both through positive evidence, namely their observable effects in context, and through negative evidence, obtained by partially reconstructing the omitted material. The objective is to produce a comprehensive and reproducible description of ellipsis as it occurs in the text.

The present [Section 4](#) outlines a method for “ghostbusting” ellipsis<sup>157</sup>, with a focus on predicate and object ellipsis and through a corpus-based approach. More specifically, [Section 4.1](#) provides a theoretical guide for detecting and analyzing ellipsis in a text, while [Section 4.2](#) details and puts into practice the principles outlined in [Section 4.1](#) by applying them to the text under analysis, the *Divine Comedy*, for which the treebank *Italian-Old* is available. Finally, [Section 4.3](#) address reconstruction with a specific focus on verbal predicate ellipsis.

#### 4.1 How to ghostbust ellipsis: a theoretical method

This Section outlines a theoretical method for retrieving and analyzing ellipsis in texts, detailing the general steps involved in its examination. The method has a practical goal: the extraction of instances of ellipsis, specifically predicate and object ellipsis, from a given text. Although the theoretical framework presented here primarily illustrates cases of predicate ellipsis, it can also be extended to account for object ellipsis. The latter, however, presupposes an involvement of valency and transitivity, which are further discussed in detail in Section 4.1.4.

The following steps are identified as potentially useful for carrying out this task<sup>158</sup>:

- 1) Detection
- 2) Identification
- 3) Reconstruction

Each step is described in detail in [Sections 4.1.1](#), [4.1.2](#), and [4.1.3](#), respectively. In addition, potential challenges associated with each task are discussed. Before addressing the details of each step, it should be emphasized that their formulation is tailored to the specific aims and practical objectives of the present work (outlined along Chapter 4). Accordingly, the description and partitioning of the steps do not entail any cognitive assumptions, nor do they imply claims regarding possible psycholinguistic implications<sup>159</sup>.

To illustrate each task, the following excerpt from an Italian song by De André is used as example (Example 61):<sup>160</sup>

<i>Se ti tagliassero a pezzetti</i>	If they cut you into pieces
<i>il vento li raccoglierebbe</i>	the wind would gather them
<i>il regno dei ragni</i>	the kingdom of spiders
<i>cucirebbe la pelle</i>	would stitch the skin

---

<sup>157</sup> Chapter 4 is an extension of Corbetta et al. (2025c).

<sup>158</sup> The steps identified here correspond to those highlighted in Čavar et al. (2024a:51), who treats these three points as separate tasks for addressing ellipsis using state-of-the-art NLP technologies.

<sup>159</sup> By cognitive assumptions is meant the branch of research concerned with the mechanisms by which a reader or a listener processes ellipsis (e.g., Martin & McElree 2009; 2011; Kertz 2010, among others).

<sup>160</sup> Translation provided by the author.

<i>e la luna tesserebbe</i>	and the moon would weave
<i>i capelli e il viso</i>	the hair and the face
<i>e il polline di Dio</i>	and the pollen of God
<i>di Dio il sorriso.</i>	of God the smile.
<i>Ti ho trovata lungo il fiume</i>	I found you by the river
<i>che suonavi una foglia di fiore</i>	playing a flower's leaf
<i>che cantavi parole leggere</i>	and you sang light words,
<i>parole d'amore</i>	words of love,
<i>ho assaggiato le tue labbra</i>	I tasted your lips
<i>di miele rosso rosso</i>	of deep red honey.
<i>ti ho detto dammi quello che vuoi</i>	I told you: give me what you want,
<i>io quel che posso.</i>	I, what I can.

**Example 61.** excerpt from an Italian song by De André

As for the text under analysis in this study, the *Divine Comedy*, Example (61) is considered a non-binding text, that is, a text in which the author's communicative constraint is not stringent with respect to the addressee (Sabatini 1990, 1999).

#### 4.1.1 Detection

The term detection refers to the identification of the ellipsis site within the text, namely locating the position of the gap representing ellipsis and marking it with an underscore (‘\_’) (see [Section 3.1.1](#)). The retrieval of ellipsis requires awareness that a linguistic element has been omitted or dropped from the sentence. However, the degree of awareness regarding such omissions, as well as the granularity of their reconstruction<sup>161</sup>, is somewhat arbitrary and sometimes open to interpretation. This is particularly true for nominal ellipsis, referred to as object ellipsis in the present study (as will be shown in Section 5.2), whereas predicate ellipsis tends to be less controversial in this respect<sup>162</sup>.

In Example 61, it is possible to detect two predicate ellipsis sites, marked with the underscore in the Example 61(a) below:

<i>(...) la luna tesserebbe</i>	(...) and the moon would weave
<i>i capelli e il viso</i>	the hair and the face
<i>e il polline di Dio _</i>	and the pollen of God _
<i>di Dio il sorriso.</i>	of God the smile.
<i>(...)</i>	<i>(...)</i>
<i>ti ho detto dammi quello che vuoi</i>	I told you: give me what you want,

<sup>161</sup> This issue is closely related to the problem of over-reconstruction, which is discussed in Section 4.1.3.

<sup>162</sup> For this reason, this Section focuses on predicate ellipsis.

*io* \_\_\_ *quel che posso*

I \_\_\_ what I can.

**Example 61(a).** predicate ellipsis sites marked with the underscore

The retrieval of the two omitted predicates is facilitated by the presence of subjects (*il polline di Dio* ‘the pollen of God’ in the first case, and *io* ‘I’ in the second), as well as their respective arguments (*il sorriso di Dio* ‘the smile of God’ and *quel che posso* ‘what I can’), which appear in the sentence without an overt verb, thereby signaling the presence of predicate ellipsis. However, the degree of awareness mentioned above can be challenged in specific cases. For instance, in Example 61, one might hypothesize the presence of predicate ellipsis in the verse *parole d’amore* ‘words of love’, by interpreting an omitted verb, as in Example 61(b):

*che cantavi parole leggere, \_\_\_ parole d’amore* ‘as you sang light words, \_\_\_ words of love’.

**Example 61(b).** possible interpretation of ellipsis

However, within the sentence, *parole d’amore* ‘words of love’ functions to emphasize and specify *parole leggere* ‘light words’, acting as a specification or apposition to the preceding nominal and thereby excluding the possibility of predicate ellipsis. In such cases, the interpretation of the text plays a crucial role in ellipsis detection.

Another type of challenging case arises in the context of coordination. For instance, a potential instance of predicate ellipsis may be hypothesized in the following line, shown in Example 61(c):

*la luna tesserebbe i capelli e \_\_\_ il viso* ‘and the moon would weave the hair and \_\_\_ the face’

**Example 61(c).** possible interpretation of ellipsis

One might detect a predicate ellipsis following the coordinating conjunction *e* ‘and’, analyzing the construction as coordination of two clauses: one headed by the overt predicate *tesserebbe* ‘would weave’, and the other headed by an omitted predicate (\_\_\_). This interpretation contrasts with a simpler syntactic reading in which *i capelli* ‘the hair’ and *il viso* ‘the face’ are coordinated noun sharing the same verb. The issue of coordination will be revisited in Section 4.2.1.1.2.2. For the purposes of the present work, the simpler syntactic interpretation is adopted whenever possible, in preference to more complex analyses.

#### 4.1.2 Identification

The term identification refers to the process of identifying the antecedent (see [Section 3.1.2](#)) which triggers the third step in the ellipsis reconstruction mechanism, i.e. the reconstruction of the elided material (see Section 4.1.3). As previously discussed, the antecedent of ellipsis may occur at a sentence level, appearing in the same sentence as the ellipsis, or at the discourse level, occurring in a preceding sentence. In both cases, the antecedent is anaphoric (or cataphoric), that is, it is retrievable from the text. Alternatively, the antecedent may not be present in the text at all and must instead be inferred from world knowledge.

In Example 61, the identification of antecedents appears relatively straightforward, with antecedents marked in bold in Example 61(d):

(...) la luna <b>tesserebbe</b> i capelli e il viso e il polline di Dio __ di Dio il sorriso. (...)	(...) and the moon <b>would weave</b> the hair and the face and the pollen of God __ of God the smile. (...)
<b>ti ho detto dammi</b> quello che vuoi io __ quel che posso	<b>I told you: give me</b> what you want, I __ what I can.

**Example 61(d).** identification of antecedents

In the first case of ellipsis, the antecedent can be straightforwardly identified as the verb *tesserebbe* ‘would weave’. In the second instance, however, the identification of the antecedent is more problematic, as it may plausibly oscillate between a broader reconstruction, such as *ti ho detto dammi* ‘I told you: give me’, and a narrower one, namely *dammi* ‘give me’. This issue will be addressed in Section 4.1.3.

As already noted in relation to detection, this point highlights the degree of arbitrariness to which the task may be subject. It emphasizes the importance of establishing clear guidelines for the treatment of ellipsis. Further details on this aspect will be provided in the section on reconstruction.

### 4.1.3 Reconstruction

The last step in handling ellipsis in a text is the reconstruction task, which aims to facilitate the analysis of ellipsis together with its elided material. It is important to emphasize that the process of reconstruction necessarily involves the creation of “fictional material”, i.e., linguistic elements of a “shady” nature, as they are *reconstructed*, *created*, rather than explicitly present in the text, and therefore carry with them all the implications that such a status of “fictional material” entails. Reconstruction is as risky as it is useful. At first glance, one may question whether such a creative operation, which may initially appear to manipulate the original text, is necessary at all. The answer

lies in the very aim of the task: since the goal is to analyze ellipsis, the risk involved in reconstructing the elided material cannot justify excluding the reconstruction process. However, and this is a fundamental point, this risk can be mitigated by adopting appropriate precautions: measures aimed at limiting, as much as possible, the degree of arbitrariness inherent in the reconstruction operation, as well as in the other two steps of the process (detection and identification).

Before turning to the precautions outlined in Section 4.2, it is necessary to first describe the reconstruction process itself. Continuing with Example 61, the final step of reconstruction proceeds as follows in Example 61(e):

<p>(...) <i>la luna tesserebbe</i>  <i>i capelli e il viso</i>  <i>e il polline di Dio [tesserebbe]</i>  <i>di Dio il sorriso.</i>          (...)</p>	<p>(...) and the moon <b>would weave</b>          the hair and the face          and the pollen of God [would weave]          of God the smile.          (...)</p>
<p><b>ti ho detto dammi</b> <i>quello che vuoi</i>  <i>io [ti ho detto: ti dò/ti dò] quel che posso</i></p>	<p><b>I told you: give me</b> what you want,          I [ <i>told you: I give you / I give you</i>] what I can.</p>

#### Example 61(e). reconstruction step

While the first reconstruction ([*tesserebbe*] ‘[would weave]’) is uncontroversial, since the omitted verb carries not only the same meaning but also the same morphological features as its antecedent (given that the Italian subjects *luna* ‘moon’ and *polline* ‘pollen’ both require a third person singular verb), the second reconstruction leaves more room for debate. More specifically, as discussed in the identification step, the material undergoing ellipsis in the second instance is more complex to reconstruct, not only because it does not involve a straightforward copy of its antecedent, as in the previous case, but also because determining what should count as the elided material may require positing more than a single verb, thereby giving rise to multiple possible reconstruction strategies.

Let us analyze this instance in more detail. As discussed in Section 4.1.2, at first glance the antecedent for this instance may be either *ti ho detto dammi* ‘I told you: give me’ or simply *dammi* ‘give me’. Accordingly, there are two reconstruction options (as indicated by the slash in Example 61(e): [*ti ho detto: ti dò / ti dò*] ‘I told you: I give you / I give you’). In the first scenario, syntactically schematized in Example 61(f), where the antecedent is *ti ho detto dammi* ‘I told you: give me’, the reconstruction extends to both the main clause (E<sub>1</sub>) and its complement clause (E<sub>2</sub>), yielding a reconstruction such as *ti ho detto: ti dò* ‘I told you: I give you’. This option involves the coordination of two main clauses (indicated in the schema by the symbol &), one of which is elided (E<sub>1</sub>). Each main clause selects its own complement clause: one is overt (*dammi quello che vuoi* ‘give me whatever you want’), while the other is omitted (E<sub>2</sub>).

[ *io ti ho detto [dammi quello che vuoi ] ] & [ io <sub>(E1)</sub> [ <sub>(E2)</sub> quello che posso ] ]  
 [ I told you [give me what you want] ] & [ I <sub>(E1)</sub> [ <sub>(E2)</sub> what I can ] ]*

**Example 61(f).** Syntactic schema illustrating the first reconstruction option.

Conversely, the alternative hypothesis, illustrated in Example 61(g), takes only the complement clause *dammi* ‘give me’ as the antecedent and yields to the reconstruction *ti dò* ‘I give you’. Under this analysis, the ellipsis gives rise to a coordination between two complement clauses, both governed by a single main clause:

[ *io ti ho detto [dammi quello che vuoi ] ] & [ io <sub>(E2)</sub> quello che posso ] ]  
 [ I told you [give me what you want] & [ I <sub>(E2)</sub> what I can ] ]*

**Example 61(g).** Syntactic schema illustrating the second reconstruction option.

Although both scenarios illustrated in 61(f) and 61(g) are theoretically plausible, the first option involves a more extensive reconstruction, which makes it comparatively more “costly” and repetitive in terms of linguistic material, as well as in terms of interpretive effort. From a syntactic perspective, the first option, 61(f), requires the duplication of both the main clause and its complement clause, thus entailing a double reconstruction of <sub>(E1)</sub> and <sub>(E2)</sub> and resulting in a more complex syntactic structure, whereas the second option, 61(g), allows the ellipsis to be resolved within a more local domain, involving only the complement clause <sub>(E2)</sub>. Moreover, reconstructing also the main clause <sub>(E1)</sub> as in 61(f) would not contribute any distinctive information; rather, it would amount to a redundant repetition that could be more economically captured by simple coordination. For this reason, the second reconstruction option, 61(g), in which only <sub>(E2)</sub> is reconstructed, appears to be syntactically more economical and interpretively less demanding.

To conclude with some more general remarks on the reconstruction of this second instance of ellipsis, it can be observed that its complexity lies in the grammatical shifts that the reader or listener must apply during the process of ellipsis resolution. More specifically, this instance involves a clear shift in the roles of the participants in the event, thereby creating a sharp contrast between them. In the antecedent, the subject, the giver, is a fictional ‘you’ (the woman to whom the song is dedicated), while the addressee is the author (*mi* ‘me’). By contrast, the ellipsis entails a clear reversal of these roles: in the elided material, the giver is now the author, and the addressee is the woman, who receives *quel che posso* ‘what I can’. The sharp opposition between the participants ‘you’ (which, due to the pro-drop nature of Italian, is not overtly expressed but is recoverable from the context) and ‘I’ *io* may

at first glance appear to increase the complexity of ellipsis resolution; however, it in fact constitutes a crucial cue for identifying and delimiting the ellipsis itself. In this case, ellipsis appears to operate through clear contrasts, thereby bringing about a shift in thematic focus, which is made explicit by the opposition between *tu* and *io*.

To sum up, complexity of ellipsis resolution in most cases extends far beyond a simple copy-and-paste of the antecedent. This task will be examined from a practical/computational perspective in Section 4.2.3, with particular attention to the extent of manual intervention still required and to the importance of adopting the most explicit annotation possible.

Given the complexity of this reconstruction, the next step is to address the question of *what* should actually be reconstructed. Although reconstruction is inevitably an intrinsic part of the analysis, and therefore a required/necessary task, the level of granularity in the reconstructed material remains open to discussion. Since the linguistic material in question is elided, and explicit cues about what should be reconstructed are not always available in the text, the scope and the depth of the reconstruction process cannot be governed by a single, universally correct solution. Rather, it is inherently subject to a degree of arbitrariness. In such cases, there is no “one-size-fits-all” approach, nor a uniquely “correct” analysis applicable across all instances. Therefore, for a task of this nature, no universal rules can be established; instead, only consistent principles aligned with the goals of the task may be applied. In filling the gap, the delicate task of the ‘reconstructionist’ is to provide sufficient information without overinterpreting or introducing creative or unnecessary elements. A key element that helps restrain interpretative freedom is the antecedent itself (when present), as it delineates the linguistic space within which the elided material is to be retrieved. Adhering to the antecedent in terms of both syntactic structure and lexical content, that is, limiting the creative attempt by following what is actually present in the text (see Section 4.2.3), can be seen as a reliable strategy that prevents drifting into overtly creative interpretations. However, as discussed in [Section 4.1.2](#), identifying the antecedent remains part of the very same challenge.

Once again, no single unique solution emerges, which is why Section 4.2.3 outlines the practical guidelines adopted in this work to clarify the reconstruction procedure.

#### 4.1.4 Summary

[Section 4.1](#) discusses a theoretical framework for handling ellipsis in corpus-based studies, with illustrative examples focusing specifically on predicate ellipsis. The method is divided into three steps: i) detection, which involves locating the ellipsis site; ii) identification, understood as the process of determining the antecedent from which the meaning of the ellipsis is retrieved; and iii)

reconstruction, which concerns establishing what is the elided material that should be restored. It has been shown that each of these steps presents its own challenges, i.e. determining what qualifies as ellipsis; identifying which linguistic element, if any, counts as an antecedent; and, finally, deciding the extent to which reconstruction should be performed, that is, how much linguistic material should be reconstructed. In relation to these challenges, this study addresses the issue of what qualifies as ellipsis, with the primary aim of providing a corpus-based classification of predicate ellipsis and object omission instances in the text under examination. As for the other two points, the inherently elusive nature of the phenomenon makes it difficult to establish universal criteria that are not subject to a certain degree of arbitrariness, since ellipsis, by definition, involves the absence of linguistic material, rendering it difficult to describe what is not overtly present. To mitigate this arbitrariness, the present study adopts a set of operational guidelines designed to ensure consistency in the identification of antecedents and the scope of reconstruction, which are detailed in the next [Section 4.2](#).

## 4.2 How to ghostbust ellipsis: a practical guide

To describe the body of the shadows, the substance of what is absent, the first requirement is to establish clear criteria that provide more reliable foundation for analysis. This Section is therefore devoted to outlining the method adopted for addressing ellipsis in the present study. Specifically, it presents the workflow developed to handle two types of ellipsis that constitute the focus of the investigation: predicate ellipsis and object ellipsis. As stated in Chapter 2, the corpus under examination is the *Divine Comedy* (Section 1.1), for which the treebank *Italian-Old* is available. Before going into details about the structure of this Section, some preliminary remarks are in order. This study presents a pilot analysis aimed at investigating ellipsis in a specific text through a data-driven approach, that is, an analysis grounded in quantitative data, with the goal of developing a classification informed by the data itself. As noted in Chapter 3, studies on ellipsis are often based on introspective judgments or elicited sentences, with limited engagement in data-driven methodologies; notable exceptions include the works of Plewnia (2013) and Hardt (1997). A further contribution of this study lies in the methodology adopted: the retrieval and part of the analysis of ellipsis are conducted in a (semi-)automatic manner, in order to facilitate both verification and reproducibility of the data. This method has been developed on the basis of the UD *Italian-Old* treebank, which provides the linguistic foundation for the analysis. Designing a methodology based on a cross-linguistic framework as UD, encompassing over 150 languages (see Section 1.4), allows for the potential reuse and extension of this approach, albeit with the adjustments to accommodate language-specific features.

Given the extensive literature on ellipsis and the broad scope of the phenomenon (see Chapter 3), this study limits its focus to two specific types: predicate ellipsis and object ellipsis. These two types of ellipsis are selected based on their linguistic relevance (predicate ellipsis for its syntactic impact, and object ellipsis for its diachronic significance) as well as their methodological feasibility. More specifically, predicate ellipsis is particularly relevant due to the extensive literature devoted to this phenomenon and to its investigation across all major theoretical frameworks (Section 3.2.1). The omission of a predicate is generally more salient and widely acknowledged, as it can profoundly impact the syntactic structure of the sentence (see Section 3.4.3.2). Object ellipsis, by contrast, is less syntactically prominent, since its omission typically does not result in ill-formed structures but rather affects the valency structure. However, it offers valuable insights from a diachronic perspective, especially when compared with patterns observed in Modern Italian (see Section 3.3.1.2).

Furthermore, both types lend themselves well to systematic analysis within the framework of UD. In the case of predicate ellipsis, UD generally marks such instances, albeit not exhaustively<sup>163</sup>, using a specific syntactic relation (`orphan`), which, at least in some cases<sup>164</sup>, is applied consistently across treebanks. This consistency facilitates the detection and reliable identification of a subset of these instances (those marked with the `orphan` relation). Similarly, for object ellipsis, the existing annotations also facilitate its retrieval, enabling a targeted investigation of this phenomenon. This makes both types particularly suitable for (semi-)automatic detection and subsequent analysis.

Finally, before turning to the detailed description of the work, the following clarification should be noted. A limitation of this study lies in the specificity of the language and text under examination<sup>165</sup>, and thus it does not aim to provide an exhaustive analysis of predicate and object ellipsis in general. Rather, by focusing on a language-specific (Old Italian, Florentine) and text-specific (poetic text) investigation, it seeks to offer a vivid, albeit restricted, portrayal of these phenomena within the text. It is hoped that this work can serve as a foundation for future developments, including extensions to other languages, for which UD treebanks offer a valuable starting point. Further research, particularly cross-linguistic investigations, is envisaged.

This Section is organized as follows: [Section 4.2.1](#) describes the detection method, providing a detailed explanation of the extraction approach used for predicate ellipsis ([4.2.1.1](#)) and object ellipsis ([4.2.1.2](#)); [Section 4.2.2](#) describes the identification task, both for predicate ([4.2.2.1](#)) and object ellipsis

---

<sup>163</sup> This point will be addressed in the section on the extraction of predicate ellipsis (Section 4.2.1.1).

<sup>164</sup> For further insights into the issues surrounding the use of `orphan` dependency relation in UD treebanks, see Droганova et al. (2018).

<sup>165</sup> As mentioned in Chapter 2, *Italian-Old* is the first UD treebank to document Old Italian. Therefore no other data could be used.

(4.2.2.2), which were performed manually. Finally, [Section 4.2.3](#) will address the reconstruction, with a focus on verbal predicate ellipsis in [4.2.3.1](#). Section 4.3 provides a summary.

#### 4.2.1 Detection

This Section outlines the implementation of the detection process described in [Section 4.1.1](#). For practical reasons, however, the term *extraction* is used in the following sections instead of detection. The former may be understood as the practical counterpart of the latter, as it involves retrieving the linguistic contexts that potentially host an ellipsis and that serve as input for the subsequent steps of the theoretical workflow, namely identification and reconstruction.

Since the goal of this work is to minimize manual effort and to design a process that is shareable, reusable, and transparent, a semi-automatic method is adopted for the detection of predicate ellipsis ([Section 4.2.1.1](#)) and object ellipsis ([Section 4.2.1.2](#)). As ellipsis in basic UD, and consequently in *Italian\_Old*, is not explicitly marked (see [Section 3.4.2](#)), omitted elements are identified through structured queries exploiting the existing annotations. The detection of both types of ellipsis is carried out using Arborator-Grew<sup>166</sup> (Guibon et al. 2020), a collaborative tool for treebank annotation and graph querying, and Udapi (Popel et al. 2017), an API framework for processing UD data.

##### 4.2.1.1 Extraction of predicate ellipsis

Predicate ellipsis is here understood as the omission of a verb or an auxiliary in predicate position within a sentence, whether occurring in a main clause or in an embedded clause. Before turning to the specifics of the data extraction process, it is necessary to clarify how the underlying grammatical notions are approached and operationalized in this study. According to traditional grammar, as described by Serianni (1991: 92), Italian is said to present two types of predication<sup>167</sup>:

62) verbal predication:

*Mario rideva*                    ‘Mario was laughing’

63) nominal predication:

- a. *Marta è giornalista*    ‘Marta is a journalist’
- b. *Gino era felicissimo*   ‘Gino was very happy’

---

<sup>166</sup> <https://arboratorgrew.elizia.net/#/>

<sup>167</sup> The examples are taken from Serianni et. al (1991: 92).

According to Serianni, verbal predication consists of a verb with a complete meaning<sup>168</sup>, as illustrated in example (62); whereas nominal predication consists of the combination of the verb *essere* ‘to be’ and a noun or adjective, as shown in examples (63)<sup>169</sup>. Interestingly, some scholars do not consider the distinction between nominal and verbal predication at all, viewing it instead as a classification inherited from Latin, where the presence of morphological case distinctions gave rise to such a categorization. This view aligns with the perspective of Benveniste (1971), who treats predication as an intrinsic and defining property of the sentence<sup>170</sup> (Benveniste 1971:152). He uses the term *predication* at the functional level, drawing a distinction between *function*, linked to syntactic structure, and *form*, associated with morphological realization. Crucially, he argues that the verbal function can exist independently of the verbal form<sup>171</sup>. For the sake of clarity and simplicity, grammatical speculations are set aside at this stage of the extraction process and deferred to the analysis phase. Accordingly, the straightforward terminology used by Serianni is adopted, with the caveat that a more critical discussion of these terms will be provided in the interpretative analysis.

All these types of predication as defined by Serianni (predicate and nominal) are, with some effort, retrievable in UD annotation thanks to the part-of-speech and syntactic information. In particular, part-of-speech tags help identify instances of verbs and auxiliaries, enabling the retrieval of both verbal predication, where the verb is tagged with the POS `VERB`, and nominal predication, where the verb *essere* ‘to be’ is tagged as `AUX`. Syntactic annotation, in turn, allows for distinguishing these cases more precisely by analyzing the dependencies between elements, thereby helping to disambiguate between true predicates and nominal uses of verbs (see Section 2.1). Moreover, since UD is predicate-centric, the retrieval of predicates is facilitated, as they occupy a head position of the clause in the syntactic tree. This aspect will be detailed in Section 4.2.1.1.2.

While the extraction of the verbal predicate ellipsis is relatively straightforward, as the omission directly involves the verb, ellipsis of the nominal predicate is more complex to address as the elided

---

<sup>168</sup> “Il predicato verbale è formato da un verbo predicativo, ossia da qualunque verbo dotato di un proprio senso compiuto che possa essere adoperato senza l’ausilio di un complemento predicativo” (Serianni 1991: 92). ‘The verbal predicate is formed by a predicative verb, that is, any verb with a complete meaning (*sensu compiuto*) that can be used without the support of a predicative complement’.

<sup>169</sup> There is also an intermediate type of predication, which includes instances of so-called effective verbs (e.g., *sembrare* ‘to seem’, *parere* ‘to appear’, etc.) and similar (Serianni 1991: 93). These are classified as either nominal or verbal depending on the grammatical tradition followed.

<sup>170</sup> “Possiamo segmentare la frase, ma non usarla per integrare. (...) Ciò dipende in primo luogo dalla peculiare caratteristica, inerente alla frase, di essere un *predicato*” (Benveniste 1971:152). ‘We can segment the sentence, but not use it for integration. (...) This depends primarily on the peculiar characteristic inherent in the sentence, namely that of being a predicate’.

<sup>171</sup> Specifically, the verbal function is distinct from the verbal form, and the two do not necessarily coincide (Benveniste 1971:182). The verbal function, particularly in assertive utterances, is grounded in its cohesive function, as it organizes the elements of the utterance into a complete structure, and in its assertive function, as it conveys a predication about reality. Consequently, a verbal form does not always correspond to a verbal function.

material could interest more than one linguistic element (the verb *essere* ‘to be’ and its noun/adjective). Ellipsis in nominal predication as defined in (63) usually affects only the verb *essere* ‘to be’, not involving the noun or adjective. In fact, omitting the noun or adjective would result in ungrammaticality (64.c), as shown in the following examples (64a-c):

- 64) a. *Marta è una giornalista. È felicissima (di esserlo).* ‘Marta is a journalist. She is very happy (to be one).’  
b. *Marta è una giornalista. Felicissima (di esserlo).* ‘Marta is a journalist. Very happy (to be one).’  
c. \* *Marta è una giornalista. È (di esserlo).* \* ‘Marta is a journalist. She is (to be one).’

However, one might question whether cases such as the one presented in (65) can be interpreted as elliptical:

- 65) *Marta è una giornalista abile e curiosa.* ‘Marta is a skilled and curious journalist’

A possible reconstruction would be: *Marta è una giornalista abile ed è una giornalista curiosa* ‘Marta is a skilled journalist and is a curious journalist’. However, this reconstruction goes beyond the intended scope and is therefore treated as a simple case of coordination.

Another preliminary issue that arises when dealing with nominal predication concerns whether the omission of the verb *essere* ‘to be’ in cases such as (64.b) is genuinely implied and whether such implication is *necessary*. Once again, given the very nature of ellipsis, which lacks overt markers, its presence cannot be stated with certainty. Therefore, these cases are provisionally treated as positive evidence, while a more fine-grained assessment is deferred to the analysis phase.

From what has been outlined so far, it is clear that one of the main challenges in addressing predicate ellipsis lies in drawing a boundary between genuine cases of predicate omission and other phenomena, such as coordination (as shown in example 65) and nominal sentences, defined as propositions in which non-verbal grammatical categories serve a predicative function (Serianni 1991:74).

These issues are addressed in the analysis phase rather than during the extraction stage, as the latter is intended to remain as neutral as possible with respect to grammatical interpretation, given that the classification is designed to be data-driven. Accordingly, during data extraction no attempt is made to differentiate between these cases; rather, all instances in which the verbal element of the predicate,

understood here as verbs of verbal predication and auxiliaries of nominal predication, are missing are extracted, with finer distinctions left to the analysis phase.

Having clarified the approach to terminology from the traditional grammatical framework, the next step is to describe how the data are used to extract the information relevant to the analysis. The extraction of predicate ellipsis is performed by automatically retrieving all verbs and auxiliaries in clausal head positions, leveraging UD annotation. Since UD does not explicitly mark ellipsis (refer to [Section 3.4.2](#)), in some cases, specific workarounds are required. Accordingly, on the basis of UD annotation and with the aim of providing a clearer overview of the queries applied, the description of the extraction process is divided into the following groups:

- > extraction of `orphan` dependencies, addressed in [Section 4.2.1.1.1](#);
- > extraction of promoted elements, i.e., extraction of what is not explicitly marked as ellipsis in UD, addressed in [Section 4.2.1.1.2](#):

The first group (extraction of `orphan` dependencies) addresses the most explicit and evident cases of predicate ellipsis (refer to [Section 3.4.2.2](#)), whereas the second group (extraction of promoted elements) captures cases in which ellipsis is handled through the promotion mechanism (refer to [Section 3.4.2.1](#)), being therefore more implicit. To address the latter (cases of promotion), and following the discussion of the challenges involved in their detection (Section 3.4.2.2), plausible contexts in which such cases are likely to occur are outlined, and a key parameter is established to support their extraction (Subsection 4.2.1.1.2.1). The query employed for their extraction is then described (Subsection 4.2.1.1.2.2).

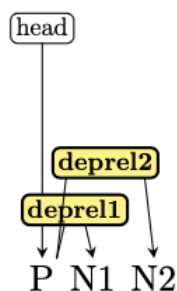
#### 4.2.1.1.1 Extraction of `orphan`

Concerning the extraction of `orphan` dependency relations, the retrieval process is facilitated by the dependency label itself, which explicitly signals the presence of ellipsis through the `orphan` relation. As discussed in [Section 3.4.2.2](#), the `orphan` relation is employed in contexts where promotion, i.e., instances in which a dependent of the elided element replaces the omitted node and assumes its syntactic role (see [Section 3.4.2.1](#)), could lead to ambiguous or misleading representations. In such cases, the `orphan` label connects the non-promoted dependents to the node that has been promoted. Therefore, ellipsis can be efficiently retrieved by querying for this specific `deprel`.

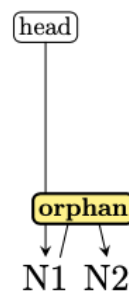
However, it is important to clarify that when extracting `orphan` relations, the focus is specifically on cases of predicate ellipsis in which *at least two nodes* function as arguments of the omitted predicate. This reflects the nature of the `orphan` `deprel`, which indicates a dependency relation between a promoted element, that assumes the function of the omitted predicate, and another argument, that would have originally depended on the elided predicate. To detect instances of

predicate ellipsis involving *fewer than two such arguments*, where a single argument is promoted without an explicit marker of ellipsis, a more fine-grained extraction process is needed. These cases are examined in detail in Section 4.2.1.1.2.

To clarify, a simplified annotation of a tree containing the `orphan` relation, previously introduced in Section 3.4.2.2, is presented in order to compare an instance of predicate ellipsis, indicated by the `orphan` `deprel`, with the corresponding non-elliptical structure, as shown in Images 41 and 42.



**Image 41.** Tree with predicate



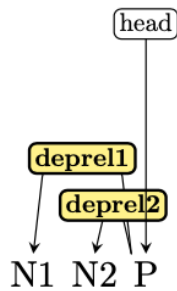
**Image 42.** Tree with omitted predicate (`orphan`)

As previously mentioned, the `orphan` `deprel` arises only when there are *at least two* nodes (N1 and N2) that would have depended on an elided predicate (P). More specifically, Image (41) illustrates a syntactic tree without predicate ellipsis, in which the two nodes N1 and N2 depend on the overt predicate P via their respective dependency relations (`deprel1` and `deprel2`). Since UD predicate-centric, the predicate P functions as the syntactic head of the clause, here conventionally marked as “head”. By contrast, Figure (42) illustrates a syntactic tree exhibiting predicate ellipsis: the predicate P is absent, and node N1 is promoted (see Section 3.4.2.1 for details on the promotion mechanism) to assume the syntactic role of the elided P. As a result, N1 functions as the head of the clause and governs N2, which depends via the `deprel orphan`.

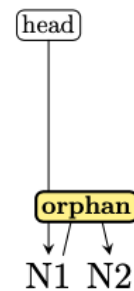
It is important to note that, since Old Italian exhibits relatively free word order<sup>172</sup>, the omitted predicate may appear after its dependent nodes N1 and N2. However, in the extraction of `orphan` relations, word order does not affect the retrieval process<sup>173</sup>. Even when the omitted predicate is head-final (see Image 43), the annotation procedure remains consistent with head-initial structures (see Image 44): N1 is still promoted to be the head position, while N2 is marked with the `orphan` relation (Image 43). Thus, the same mechanism applied in head-initial cases yields a structurally similar result in head-final configurations (compare Images 43 and 44).

<sup>172</sup> Poletto (2014) demonstrates that Old Italian exhibits a greater number of elements in the left periphery compared to Modern Italian.

<sup>173</sup> As shown in Section 4.2.1.1.2.2, this factor is not relevant even in the extraction of promotion cases.



**Image 43.** Tree with head-final predicate



**Image 44.** Tree with omitted predicate (*orphan*)

The query performed to retrieve `orphan` is the following<sup>174</sup>:

Pattern { N1 `-[orphan]` → N2 }

**Query 1.** Orphan retrieval

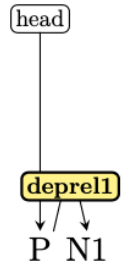
**Query 1** identifies the existence of two nodes, N1 and N2, where N1 is the head of N2, and N2 depends on N1 via the `orphan` `deprel`. By executing this query, all instances of `orphan` relations in the treebank are extracted and listed.

#### 4.2.1.1.2 Extraction of promoted elements

As discussed in [Section 3.4.2.1](#), when the absence of an elided predicate does not result in “unnatural syntactic relations”<sup>175</sup>, the solution adopted by UD for ellipsis is to promote one of the overt dependents to function as the head. However, this approach does not explicitly signal the presence of ellipsis, for example, through the use of `orphan` relations, thus potentially obscuring the underlying elliptical structure. Such cases arise when only a single argument or adjunct depends on the (elided) predicate. As in the `orphan` case illustrated above, a graphical representation of the annotation strategy based on promotion is provided in Images (45) and (46), which is applied when only one argument or adjunct is attached to the (elided) predicate:

<sup>174</sup> The Arborator-Grew formalism (Guibon et al, 2020) is privileged for illustrating the method adopted. However, the same procedure can be applied with other tools, provided that the corresponding formalism is appropriately implemented.

<sup>175</sup> <https://universaldependencies.org/u/dep/all.html#al-u-dep/orphan>



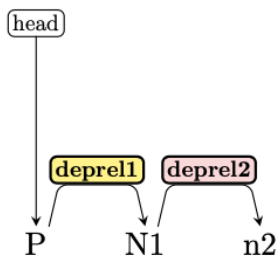
**Image 45.** Tree with predicate



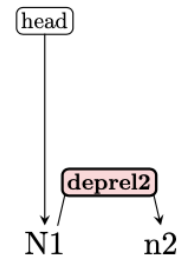
**Image 46.** Tree with omitted predicate (promotion)

In contrast to Image (42), Image (46) does not display any orphan `deprel` that would facilitate the detection of ellipsis by explicitly signaling its presence. Since such cases are not marked, they cannot be easily retrieved from the treebank. Although abstracting the graph in Image (46) might suggest a solution for extracting these cases based on identifying a single node `N1`, the scenario is further complicated by the fact that a node may, of course, serve as the head of additional dependents, thereby functioning as the parent of multiple nodes.

This scenario is illustrated in the Images (47) and (48):



**Image 47.** Tree with predicate



**Image 48.** Tree with omitted predicate (promotion)

Image (47) displays the same tree of (45), with the only difference that this time `N1` has a children node `n2` as its dependent, taking a `deprel2`<sup>176</sup>. If the predicate is omitted, the resulting structure is shown in Image (48): `N1` is promoted to head, assuming the role of the omitted predicate. The graph in Image (48) displays a structure similar to that in Image (44), with the key difference that `n2` does not receive the orphan `deprel`, once again resulting in a loss of information regarding the ellipsis. Moreover, unlike in Image (46), there is no single isolated node: `N1` is not an isolated element, as nodes typically do not occur independently. Rather, it brings along its possible dependents, resulting in deeper syntactic structures that are not explicitly marked as elliptical. Once the possible configurations of promotion have been illustrated, it becomes evident that recovering unmarked cases

<sup>176</sup> The node `n2` is indicated by the lowercase and shows a pink `deprel` label as visual aid.

in the syntactic annotation requires identifying the contexts in which such cases are likely to occur and, where possible, defining parameters to guide their retrieval.

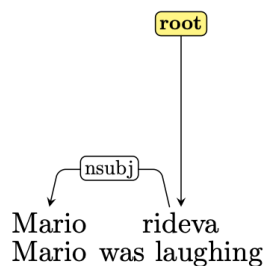
#### 4.2.1.1.2.1 Theoretical workarounds for promotion

The description begins with the contexts in which ellipsis is likely to arise, outlining the relevant variables to be considered in their characterization. The effectiveness of these variables is then assessed, and finally, a parameter that has proven particularly decisive in the extraction process is introduced.

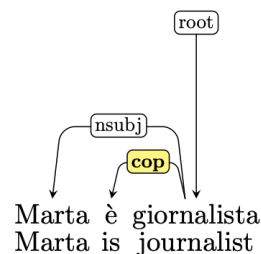
##### 1) Type of the elided predicate

One essential variable to consider when designing queries for retrieving promoted elements is the type of the elided predicate itself. As previously discussed in [Section 4.2.1.1](#), predicates ellipsis can involve either verbal or nominal predicates, each of which entails different annotation strategies within the UD framework.

Images (49) and (50) present examples of how UD annotates verbal and nominal predicates. The sentences are the same provided in the examples in [Section 4.2.1.1](#):



**Image 49.** Verbal predicate



**Image 50.** Nominal predicate

As shown in Image (49), verbal predicates serve as the heads of the syntactic structures in which they appear. In contrast, in nominal predication (Image 50), the syntactic head is nominal or adjectival element, while the verb *essere* ‘to be’ functions as a dependent, linked to the head through the `cop` (copula) dependency relation. In such constructions, the element that is usually elided is the verb *essere* ‘to be’ itself (refer to [Section 4.2.1.1](#)). Consequently, the omission affects a leaf node rather than the syntactic head of the clause, as it does in verbal predication, leaving no explicit cues of the elision, since the node is simply absent, and no promotion takes place.<sup>177</sup>

<sup>177</sup> It is important to stress that, in cases of nominal predication, the omission of the verb *essere* (‘to be’) is not handled in UD through promotion. Since *essere* functions as a leaf node in these constructions, its elision does not trigger the promotion of any other element. Rather, the verb *essere* is simply omitted from the syntactic structure, resulting in its complete absence from the tree.

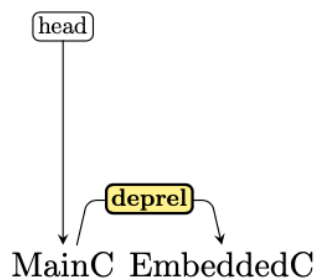
This difference in annotation must be considered when designing queries to extract elided predicates.

## 2) Type of clause with the elided predicate

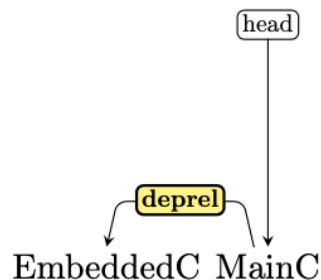
Another additional variable to consider is the type of the clause in which the predicate is omitted. This includes both the clause's position (main or embedded) and its syntactic function (e.g., coordinated clause, adverbial clause, etc.). Specifically, retrieving an omitted predicate in main clause is relatively straightforward, as the predicate, though elided, is associated with the head of the clause, typically in `root` position. In contrast, identifying omitted predicates in embedded clauses is more complex. In such contexts, the syntactic role of the missing predicate may vary depending on the syntactic function of the embedded clause, resulting in a broader range of structures that must be queried (`cconj` for coordinated clauses, `advcl` for adverbial clauses, etc.).

## 3) Position of the embedded clause

Moreover, when the omitted predicate is in an embedded clause, another variable that should be considered in the extraction is the position of the embedded clause relative to the main clause. As illustrated in Images (51) and (52), the embedded clause (EmbeddedC) can either follow (51) or precede (52) the main clause (MainC), further expanding the range of cases to search through.



**Image 51.** Following embedded clause

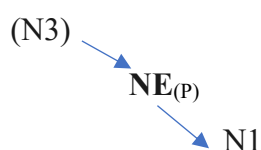


**Image 52.** Preceding embedded clause

All these variables, namely the type of predicate ellipsis, the type of clause in which the predicate is elided, and the position of the embedded clause, will be taken into consideration in the formulation of the queries designed to extract case of promotion. In addition to these variables, another crucial parameter plays a central role in the detection process: the morpho-syntactic mismatch.

## 4) Morpho-syntactic mismatch:

One key indicator of promotion is the presence of morpho-syntactic mismatches, particularly when a part-of-speech receives a non-prototypical dependency relation. This parameter operated at the interaction of two distinct levels: the morphological and the syntactic. The mismatch arises because the promoted node retains its original part of speech while assuming a syntactic role that would otherwise be filled by the elided predicate. For instance, this may occur when a noun<sup>178</sup>, typically associated with syntactic relations such as *nsubj*, *obj*, or *obl*, is assigned a relation more commonly associated with verbs, such as *root*, *advcl*, or *conj*. Such mismatches may serve as indicators that the noun has been promoted from the head of a nominal phrase to the head of a clause. To further clarify the concept, the following are presented: i) the syntactic structure of a sentence with an elided predicate (Image 54), alongside its reconstructed counterpart, that is, the same sentence with the missing predicate restored (Image 53); and ii) a relational table that provides the morpho-syntactic information for both constructions (Table 5). Here, N3 is an optional node functioning as the head, while  $NE_{(P)}$  denotes the reconstructed elided predicate. N1 is the node promoted to fulfill the syntactic role of the missing predicate.



**Image 53.** Reconstructed predicate ellipsis



**Image 54.** Promoted sentence

Reconstructed predicate ellipsis	Promoted sentence
N3 = pos3; deprel3	N3 = pos3; deprel3
$NE_{(P)}$ = pos <sub>(P)</sub> ; deprel <sub>(P)</sub>	-----
N1 = pos1; deprel1	N1 = pos1; deprel <sub>(P)</sub>

**Table 5.** relational table providing morpho-syntactic information

As shown in Table 5, the promoted sentence does not fully convey all the relevant information. In particular, both the part of speech of the elided element ( $pos_{(P)}$ ) and the original dependency relation (*deprel1*) of the promoted element (N1) are absent, resulting in a loss of information. In the promoted structure, N1 exhibits the previously discussed morpho-syntactic mismatch: while it retains its original part of speech (*pos1*), it is assigned the dependency relation associated with the elided

<sup>178</sup> Working within the UD framework, references to parts of speech adhere to the specific definitions and conventions established by the framework. For instance, when referring to nouns, the definition of a noun is as outlined in UD (see <https://universaldependencies.org/u/pos/index.html>).

predicate ( $\text{deprel}_{(P)}$ ). This mismatch serves as a crucial feature for detecting promoted elements within the annotation.

Building on these considerations regarding the relevant variables and the outlined parameter, the next step is to describe the queries.

#### 4.2.1.1.2.2 Queries for promotion

Once the possible scenarios and the relevant caveats for extracting promotion cases have been outlined, these considerations are applied to the formulation of the queries. The present subsection considers all the identified variables and translates them into practical, operational queries.

- i) **Type of predicates:** since predicates can be either verbal or nominal (Section 4.2.1.1), it is necessary to account for both scenarios. In the case of verbal predicates, the retrieval process must target nodes that are *not* verbs, as the elided element is the verbal predicate itself, i.e., a verb. In contrast, for nominal predicates, the retrieval condition focuses on cases where the omission involves the verb *essere* ‘to be’, which functions as a copula.
- ii) **Type of clause:** this variable intersects both the position of the clause (main or embedded) and the syntactic function of the promoted node or the elided predicate ( $\text{deprel}_{(P)}$ ). Regarding the position of the elided predicate, it may occur as the head of a main clause, i.e., attached to a fictional node with a `root` dependency relation (see Section 4.2.1.1.2.1), or within an embedded clause, in which case it is governed by a higher node through a specific clausal relation. However, clause position alone is not sufficiently informative for retrieval. It must be combined with the syntactic function of the elided predicate ( $\text{deprel}_{(P)}$ ). Since UD adopts a predicate-centric approach (see Section 1.4), and predicates function as the heads of clauses, the criterion adopted is to target all dependency relations associated with clause heads. This approach applies to cases of verbal predicate ellipsis. In contrast, nominal predication requires a different treatment. In such cases, the target node is not the clause head, which is the nominal or adjectival element, but the verb *essere* (‘to be’), which consistently takes the `cop` dependency relation. Its omission can therefore be detected by identifying contexts in which the copula is expected but absent.
- iii) **Position of the embedded clause:** although this factor is theoretically relevant, it does not influence the query process. This is because dependency relations in UD are not

directional, meaning that nodes can be queried regardless of their linear position relative to their governor.

- iv) Finally, the last parameter that aids in the extraction of promoted elements is the **morpho-syntactic mismatch**. Unlike the previous variables, this one emerges naturally from the observation that the promoted node retains its original part-of-speech category while assuming a syntactic function typically associated with a different category, specifically, that of the elided predicate.

The queries used to retrieve such cases are provided below. As for the `orphan` extraction (Section 4.2.1.1.1), they are expressed in the formalism of Grew-Arborator, although alternative tools may be employed, provided that the appropriate formalism is adopted.

```
pattern {N1 -[*]deprel->N2 }  
without {N2 [upos="VERB"] }  
without {N2-[cop]->X}  
without {N2-[orphan]->X}
```

(\*)deprel = root, parataxis, advcl\*, ccomp\*, csubj\*, conj, acl\*

### Query 2. Orphan extraction

Query 2 posits the existence of a node N1 (corresponding to N3 in the formalism used in Table 5) that governs a node N2 (corresponding to N1 in Table 5) via a dependency relation generally labeled “deprel” (referred to in Table 5 as  $\text{deprel}_{(p)}$ ). The term “deprel” here refers to one of the UD relations that mark clause-level<sup>179</sup>. The presence of an asterisk (\*) following some deprels (e.g., `advcl*`, `ccomp*`, `csubj*`, `acl*`) indicates that both the base relations and their subtypes are included in the query. The only clausal head relations excluded from the query are `xcomp` and `advcl:pred`, as they are typically used to mark secondary predication, which falls outside the scope of this study. To verify which dependency relations actually occur in the *Italian-Old* treebank, a script<sup>180</sup> provided by the Universal Dependencies<sup>181</sup> developers was run to generate statistics on the CoNLL-U files, including the dependency relations used and their frequency. The list of dependency relations that mark clause-level in *Italian-Old* are reported in Table 6:

<sup>179</sup> <https://universaldependencies.org/u/dep/index.html>

<sup>180</sup> The script, named `conllu-stats.pl`, is designed to collect and provide various statistics on the CoNLL-U files (<https://github.com/universaldependencies/tools?tab=readme-ov-file#conllu-statspl>).

<sup>181</sup> <https://github.com/universaldependencies/tools>

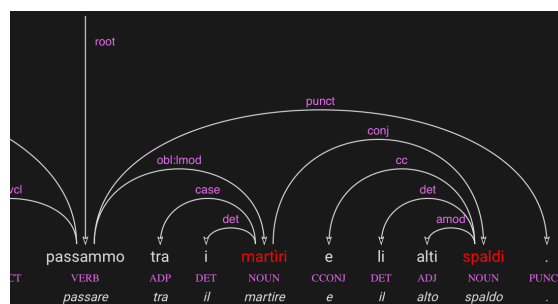
Clausal dependency relation	Occurrences in the <i>Comedy</i>
acl	690
acl:relcl	3250
advcl	3829
advcl:cmp	770
advcl:pred	378
advcl:relcl	119
ccomp	725
ccomp:reported	936
conj	3849
csubj	361
csubj:pass	3
parataxis	1835
root	3419
xcomp	1460

**Table 6.** dependency relations in the *Comedy* and their frequency

The query further specifies that N2, namely the promoted node, is not a verb, i.e., its part-of-speech tag is not VERB. Additionally, the query excludes cases where N2 has a dependent node with the `cop` `deprel`, targeting nominal ellipsis, as well as cases where N2 has a dependent node with the `orphan` `deprel`, since such cases are retrieved and analyzed separately (see Section 4.1.1.1).

A special treatment was applied to cases of coordination. This is because `conj` is not exclusively used to mark clausal coordination, but is also employed to mark coordination between phrases, including arguments or adjuncts. For this reason, an additional restriction had to be introduced. More specifically, for a coordinated element, the parent of the node from which the `conj` `deprel` originates must bear a clausal syntactic `deprel`. This restriction is essential to filter out false positives, such as those illustrated in Image (55), where the `conj` relation, although occurring between two nouns and not involving a verbal POS, and thus potentially qualifying for the query in 1, actually coordinates two oblique phrases, thereby yielding a false positive.

*Inf. IX, v. 133*  
*passammo tra i martiri e li alti spaldi*  
 ‘we passed between the torments and high walls’



### Image 55. False positive

Image (55) shows two nodes, N1 (*martiri* ‘torments’) and N2 (*spaldi* ‘walls’). The node N2 bears a non-verbal part-of-speech tag (i.e., NOUN) and depends via a `conj` relation on node N1, which also carries the NOUN tag. These characteristics would make the example appear to meet the conditions defined in Query 1 for the extraction of predicate ellipsis cases annotated through promotion. However, although this example ostensibly satisfies the query requirements, the `conj` relation here does not operate at the clausal level, where it would signal predicate ellipsis, but rather at the nominal level, indicating coordination between two nominal elements. Therefore, such cases should be excluded from the results through an additional constraint. For this reason, the query applied in presence of coordination is reported in Query 2:

```
pattern {N0 –
[acl | acl:relcl | advcl | advcl:relcl | advcl:cmp | ccomp | ccomp:reported | csu
bj | csubj:pass | parataxis | root]->N1}
pattern {N1 –[conj]->N2 }
without {N2 [upos=“VERB”] }
without {N2-[cop]->X}
without {N2-[orphan]->X}
```

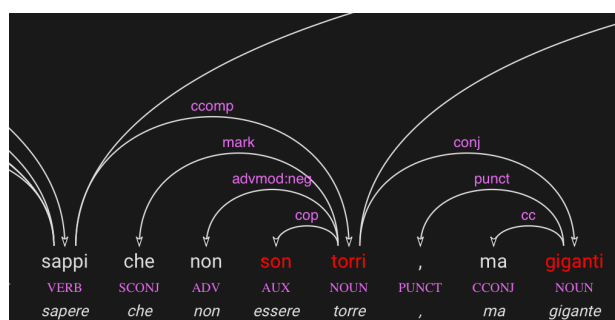
### Query 2. Coordination query

The first line of the query imposes a restriction on the type of syntactic relation in which node N1 occurs, limiting the contexts in which coordination applies to phrases (as in the cases shown in Image 55) and returning only those where the relation involves a clause, as illustrated in Image (56).

*Inf. XXXI, v. 31*

*sappi che non son torri, ma giganti,*

‘I’d have you know they are not towers, but giants’



### Image 56. coordination

In Image (56), although node N1 (*torri* ‘towers’) is a noun governing another noun N2 (*giganti* ‘giants’) via a `conj` relation, unlike the case illustrated in Image (55), N1 is the head of a completive clause, as indicated by its `ccomp` deprel.

#### 4.2.1.2 Extraction of object ellipsis

Object ellipsis is understood here as instances in which a canonical transitive predicate lacks one of its core arguments, specifically the argument corresponding to the syntactic relation of the object. In clauses headed by a transitive predicate that typically follows a Subject-Predicate-Object (S-P-O) structure, the object (O) is omitted, yielding a Subject-Predicate (S-P) structure. For example, in an Italian sentence with a prototypical transitive verb such as *Io mangio la pizza* ‘I eat pizza’, the object may be omitted, resulting in *Io mangio \_* ‘I eat \_’. To retrieve all instances of object omission, two types of information are essential in the data:

- i) the transitivity status of the predicate (i.e., whether it is transitive or intransitive);
- ii) the ability to verify the presence or absence of an overt object depending on the predicate.

With regard to the second criterion, treebanks by design provide syntactic annotations that allow for verifying the presence or absence of an object by querying for predicate nodes that have an object among their dependents. As for the first criterion, predicate transitivity is not annotated in UD. However, as discussed in Section 2.1.2, *Italian\_Old* is partially derived from a conversion of a previous corpus DanteSearch (Tavoni, 2022), which marks predicate transitivity<sup>182</sup> in its annotation scheme. In *Italian\_Old*, the original DanteSearch tagset is preserved in the `xpos` field, thereby making transitivity information accessible within the CoNLL-U format of the treebank. The combination of these two annotations enables the systematic investigation of object ellipsis within the treebank.

To extract instances of object ellipsis, Udapi (Popel et al. 2017) was employed, an open-source application programming interface (API) framework specifically designed for processing UD data. By means of Udapi modules, a script<sup>183</sup> was developed that integrates the two key criteria, predicate transitivity and the presence or absence of an overt object, to extract all occurrences of object

---

<sup>182</sup> To date, the project has not released any official documentation detailing the annotation criteria for transitivity and intransitivity. The only source of information currently available is Tavoni (2022).

<sup>183</sup> The script will be made available at: <https://github.com/ClaudiaCorbe>.

omission. As previously noted, this automatic extraction serves as a preliminary step for the manual validation process outlined in Section 5.2. The procedure for object extraction is detailed below.

In DanteSearch, and, by extension, in the `xpos` field of the CoNLL-U representation, predicate transitivity is encoded in the second character of the alphanumeric tag (see Section 2.1.2). This character follows the part-of-speech marker (e.g. “v” for verb) and uses “t” to denote transitive verbs and “i” for intransitive verbs. Image (57) presents an excerpt of the CoNLL-U format containing an example with the transitive verb *vedere* ‘to see’ (specifically, *vedi* ‘you saw’).

```
# sent_id = Canto1-22
# text = Vedi la bestia per cu' io mi volsi; aiutami da lei, famoso saggio, ch'ella mi fa tremar le vene e i polsi».
1  Vedi  vedere  VERB  vta2mps2  Mood=Imp|Number=Sing|Person=2|Tense=Pres|VerbForm=Fin|Voice=Act 0  root  0:root  Canto=1|Verso=88
2  la  il  DET  rdfs  Definite=Def|Gender=Fem|Number=Sing|PronType=Art  3  det  3:det  Canto=1|Verso=88
3  bestia  bestia  NOUN  sf1fs  Gender=Fem|Number=Sing  1  obj  1:obj  Canto=1|Verso=88
```

**Image 57.** CoNLL-U of a transitive verb

In this example, taken from *Inf.* I, v. 88, the `xpos` field, found in the fifth column, encodes transitivity in its second character, with the tag `vt...` indicating that the verb (v) is transitive (t).

As previously noted, transitivity information was combined with the syntactic annotation provided by the dependency trees. To this end, all transitive verbs, namely all nodes whose `xpos` tag begins with `vt`, were selected if they lacked any “object-like” nodes among their immediate syntactic dependents. “Object-like” nodes are understood here as those marked by the following dependency relations: `obj`, indicating direct objects; `ccomp`, denoting clausal complements; and `xcomp`, referring to open clausal complements. Subtypes of these `deprels`, such as `ccomp:reported`, which marks reported speech, were also excluded from consideration when present. An additional constraint involved the exclusion of all expletive child nodes (`expl`, `expl:pv`, `expl:impers`, `expl:pass`), as these typically occur in constructions featuring the clitic *si*, a special marker used to express reflexive, middle, impersonal, and passive voice<sup>184</sup>. The analysis of such constructions lies beyond the scope of the present study. Moreover, child nodes labeled `nsubj:pass` (passive subjects) were excluded, since the absence of an object in passive constructions does not yield relevant insights for the goals of this investigation. Finally, the extraction focuses exclusively on finite verb forms, excluding verbs in the participial, infinitival, and gerundial forms.

For illustration, the script is designed to retrieve cases such as Example (66) and Image (58), in which the verb is marked as transitive but lacks any “object-like” dependency relation among its child nodes.

<sup>184</sup> Refer to Cennamo, 1993; Cennamo & Jezek 2011; Pescarini 2015, among others.

In contrast, instances such as Example (67) and Image (59), where an object child (e.g., `obj`) is explicitly present, are excluded from the extraction.

66) *Inf.* V, vv. 73-77:

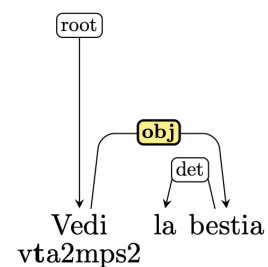
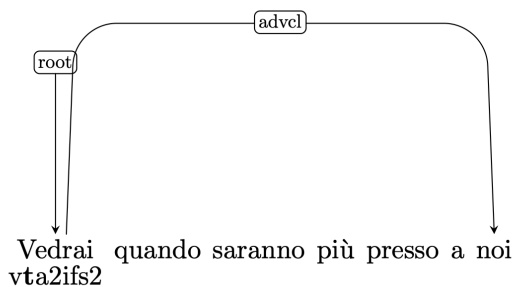
*I' cominciai: «Poeta, volontieri parlerei a quei due che 'nsieme vanno (...)». Ed elli a me: «Vedrai quando saranno più presso a noi; e tu allor li priega (...)»*

‘My first words: “Poet, I should willingly speak with those two who go together there (...)”. And he to me: “You’ll see when they draw closer to us, and then you may appeal to them (...).”’

67) *Inf.* I, v. 88:

*Vedi la bestia per cu' io mi volsi*

“You see the beast that made me turn aside”



**Image 58.** Tree of a transitive verb without object    **Image 59.** Tree of a transitive verb with object

In Example (66), as illustrated in Image (58), the transitivity of the verb *vedrai* ‘you will see’ (see.FUT.2SG) is indicated by the letter `t` in the `xpos` tag (as shown in brackets in Image 57). Additionally, this transitive verb lacks any “object-like” child nodes, having only one child node, *noi* ‘we’, which functions as the head of an adverbial clause (`advcl`), thereby making this example suitable for extraction. By contrast, in Example (67), the verb *vedi* ‘you see’ (see.PRES.2SG), although marked as transitive in the `xpos` tag (see Image 59), has an object-dependent node (`obj`), *bestia* ‘beast’, as illustrated in Image 59. Consequently, this instance is excluded from the script’s extraction.

#### 4.2.2 Identification

Once the extraction step was completed, the cases were manually checked, and the identification process was carried out. As discussed in [Section 4.1.2](#), identification involves determining the antecedent, specifically, whether an antecedent is present in the text, and if so, identifying it. The task

of identifying the antecedent of a given element in a text is related to the broader domain of coreference and anaphora resolution (Poesio et al. 2016). However, for the purpose of this study, no automatic coreference resolution was applied. Instead, antecedents were identified manually. This choice was motivated by the fact that identifying antecedents in elliptical constructions is particularly challenging, and no pre-annotated coreference datasets are currently available for Old Italian. [Section 4.2.2.1](#) outlines the identification process for predicate ellipsis, whereas [Section 4.2.2.2](#) addresses that of object ellipsis.

#### 4.2.2.1 Predicate ellipsis

The identification of the antecedent in cases of predicate ellipsis was carried out by manually inspecting the sentences retrieved through the queries described in the predicate extraction section ([Section 4.2.1.1](#)). Once the potential elliptical site was retrieved and its status as an instance of ellipsis was confirmed, the process involved reading the surrounding context and identifying the “antecedent”<sup>185</sup>, whether it precedes (antecedent) or follows (postcedent) the ellipsis site.

More specifically, with regard to the position of the antecedent relative to the ellipsis site, the antecedent (highlighted in bold in the examples) may either precede the ellipsis site, as illustrated in (68), or follow it, i.e., function as a postcedent, as shown in example (69):

68) *Inf.* VII, vv. 56-57

*questi **resurgeranno** del sepulcro col pugno chiuso, e questi \_\_\_ coi crin mozzi.*

‘These from the sepulchre shall rise again With the fist closed, and these \_\_\_ with tresses shorn.

69) *Inf.* XXX, v. 57

*l’un \_\_\_ verso ’l mento e l’altro in sù **rinverte**.*

‘One \_\_\_ tow’rds the chin, the other upward turns.’

Concerning the textual position of the antecedent, it may be found in the same sentence as the ellipsis site, as illustrated in examples (68) and (69), in a preceding sentence (70), or it may be entirely absent from the text (71). Examples of the latter two scenarios are provided below:

70) *Inf.* VI, vv. 40-43

*«O tu (...), mi **disse**, «riconoscimi (...). E io \_\_\_ a lui: «L’angoscia che tu hai (...)*

---

<sup>185</sup> As noted in [Section 3.1.2](#), the term “antecedent” is used in accordance with the literature to include both antecedents and postcedents.

“O you (...)” he said to me, “recall me, (...)”. And I \_\_\_ to him: “It is perhaps your anguish (...)

71) *Inf. I*, vv. 129-131

(...) *oh felice colui cu' ivi elegge!* » . *E io \_\_\_ a lui: «Poeta, io ti richeggio per quello Dio che tu non conoscesti (...)*

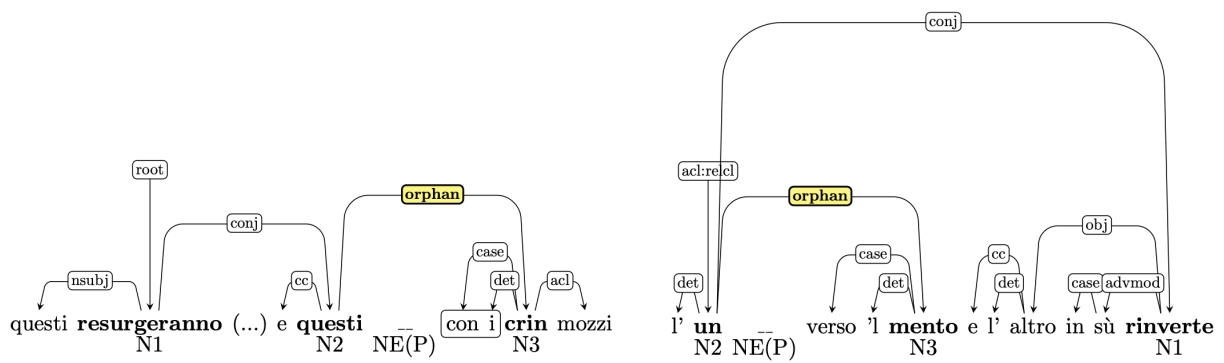
‘(...) o happy those He chooses to be there!’ And I \_\_\_ to him: “Poet, I thee entreat, by that same God whom thou didst never know (...)’

In examples (68) and (69), the antecedents are respectively the verb *resurgeranno* ‘shall rise’ and *rinvertite* ‘turns’, which are attested in the same sentence of the ellipsis site, though it occurs in the preceding/following clause. In example (70), by contrast, the antecedent, namely the verb *disse* ‘said’, is not located in the same sentence as the ellipsis site, but must be identified in the preceding one, thereby crossing the boundary imposed by the full stop and the segmentation of discourse into separate sentences. This fact let us understand once again the wide scope of ellipsis, which is only partially stuck to the syntactic context. Finally, in example (71), no explicit antecedent is present. The reconstruction of the verb relies on a shift in speaker, from Virgil to the first-person narrative voice of Dante’s character. Even if one were to consider tracing the antecedent back to the point at which Dante the author introduces Virgil’s reported speech, this would entail returning to verse 92, whereas the ellipsis site occurs in verse 130, implying a considerable distance between the potential antecedent and the ellipsis site, and thereby making such a retrieval highly unlikely. Therefore, instances like the one in (71), namely cases in which the antecedent is not explicitly present or its identification requires parsing a long stretch of text, were classified as involving world knowledge. It is worth noting that no postcedents were observed outside the sentence boundary.

The fact that only a subset of antecedents occurs at the sentence level, that is, where the antecedent and the ellipsis site appear within the same sentence (as exemplified in (68) and (69)), restricts the applicability of UD syntactic annotation to these cases<sup>186</sup>, making it difficult to fully automate the process of antecedent identification and, consequently, its reconstruction. Focusing on these sentence-level cases, the examination of syntactic trees reveals that the antecedent can, in principle, be extracted by leveraging the syntactic annotation. However, the scenarios to be accounted for are complicated by the fact that the position of the antecedent is not fixed; it may either precede (ante-) or follow (post-) the ellipsis site. More precisely, Images (60) and (61) illustrate the syntactic trees corresponding to examples (68) and (69), respectively.

---

<sup>186</sup> As mentioned in Section 1.4, UD annotation is conducted at the sentence level, focusing on intra-sentential structures.



**Image 60.** Syntactic tree of example (68)    **Image 61.** Syntactic tree of example (69)

As evident from the trees, in Image (60) the antecedent (N1) functions as the syntactic *parent* of the node (N2) to which the `orphan` node (or the promoted ones, in cases of promotion) is attached, namely, N3. In contrast, Image 61 presents a different syntactic configuration: the antecedent (or postcedent, in this specific case), N1, is a *child* of the node N2 to which the `orphan` node (or again, the promoted nodes) is attached (N3), rather than its parent. This variation effectively doubles the configurations that must be taken into account if the goal is to automatically identify and retrieve the antecedent. Therefore, although syntactic information can be exploited to aid in the identification and reconstruction of potential antecedents, it is clearly insufficient as a standalone method for antecedent resolution. Manual verification remains necessary, highlighting the complexity and challenges of resolving coreference in elliptical constructions using syntactic annotation alone.

To add further complexity, it is important to note that some instances present a challenging identification of the antecedent. This can be attributed to the interpretive freedom afforded to the reader, as well as to the inherently ambiguous and deliberately open-ended nature of poetic language, which must always be taken into consideration and never overlooked. An example of such challenging identification is shown in (72):

72) *Inf.* XVII, vv. 10-12

*La faccia sua era faccia d'uom giusto, tanto benigna avea di fuor la pelle, e d'un serpente  
 \_\_ tutto l'altro fusto*

The face was as the face of a just man, Its semblance outwardly was so benign, And of a serpent all the trunk beside.

In such an example, the identification of the antecedent remains uncertain between the two verbs in the sentence, *avea* 'had' and *era* 'was', as both could plausibly be interpreted as the elliptical material that fills the ellipsis site. Although the symmetry of the sentence could suggest reconstructing the

verb *era* ‘was’, the principle of economy, favoring the closest possible verb, results in the selection of *avea*. Thus, identifying the antecedent (and reconstructing) the ellipsis is not a straightforward process.

Finally, an additional consideration that warrants attention concerns the nature of the antecedent, an aspect that intersects with the reconstruction stage discussed in [Section 4.2.3.1](#). Specifically, the antecedent does not necessarily replicate the form of the elliptical site. While examples (68) and (69) display identical structures for the antecedent and its elided materials, other cases exhibit a more complex relationship, deviating from a straightforward copy-and-paste pattern, as previously illustrated in examples (70) and (71), and further exemplified below in example (73):

73) *Inf. X*, vv. 1-3

*Ora sen va per un secreto calle,/ tra 'l muro de la terra e li martiri, lo mio maestro, e io \_\_\_  
dopo le spalle.*

Now onward goes, along a narrow path/ Between the torments and the city wall, My Master,  
and I (follow) at his back.

The reconstruction of the elided material, along with the issues it raises, will be addressed in detail in [Section 4.2.3](#). However, it is worth noting here that in identifying the antecedent in example (73), namely, the verb *va* ‘goes’ in the third person singular, a mismatch becomes apparent, as the elliptical sentence requires a different person agreement, specifically the first person.

To summarize, considered all the challenges highlighted before, the identification of the antecedent has been conducted manually.

#### 4.2.2.2 Object ellipsis

Similarly to the identification of predicate antecedents, the identification of the antecedent of the omitted object was performed manually by examining the instances retrieved through the queries described in [Section 4.2.1.2](#). For each case of object omission, the following aspects were evaluated and annotated as additional information:

- i) the contextual transitivity of the verb;
- ii) the identification of the antecedent of the omitted object, along with its classification as either referential or generic, based on the criteria proposed by Lo Duca (2000) and Cennamo (2017) for Modern Italian (see Section 3.2.2.2);

The following section outlines each step in detail. Regarding the contextual evaluation of verb transitivity, this step served as a verification, since the present study builds upon a pre-existing annotation of transitive and intransitive verbs, as discussed in Section 2.1.2. The original annotation was carried out by scholars involved in the DanteSearch project. Unfortunately, no documentation is available concerning the criteria adopted in that process, specifically, whether a particular dictionary was consulted (and, if so, whether it was a modern or historical one) to determine the transitivity of a given verb. As a result, the existing annotation was relied upon, under the assumption that it had been conducted with sufficient scholarly rigor to warrant reuse. In positive cases, however, the verb's transitivity was re-evaluated in context. An example of such reassessment is provided in Example (74):

74) *Inf.* I, vv. 13-18

*Ma poi ch'i' fui al piè d'un colle giunto, là dove terminava quella valle che m'avea di paura il cor compunto, **guardai** in alto e vidi le sue spalle vestite già de' raggi del pianeta che mena dritto altrui per ogne calle.*

'But after I had reached a mountain's foot, At that point where the valley terminated, Which had with consternation pierced my heart, Upward I **looked**, and I beheld its shoulders Vested already with that planet's rays Which leadeth others right by every road.'

In Example (74), the verb *guardai* 'looked' was annotated as transitive, although its use with an oblique phrase (*in alto* 'upward'), indicating location, is more appropriately interpreted as intransitive. In this context, the meaning is closer to *volgere lo sguardo*, i.e., directing one's gaze, which supports an intransitive interpretation. This reading is also corroborated by the *Grande Dizionario della Lingua Italiana*<sup>187</sup>. Therefore, for each positive attestation, the verb's transitivity was manually reassessed in context, regardless of the existing annotation, in order to filter out cases such as the one illustrated in Example (74).

Once contextual transitivity had been verified, the subsequent steps involved identifying the antecedent of the omitted object and classifying it as either generic or referential. Examples illustrating generic (75) and referential (76) omitted objects are presented below.

75) *Inf.* I, vv. 55-56

*E qual è quei che volentieri **acquista**, e giugne 'l tempo che perder lo face*

And as he is who willingly acquires, And the time comes that causes him to lose

---

<sup>187</sup> Refer to: <https://www.gdli.it/sala-lettura/vol-vii/7>

76) *Inf.* VI, v. 18

*graffia li spirti ed iscoia ed isquatra.*

He rends the spirits, flays, and quarters them.

In Example (75), the omitted object of the transitive verb *acquista* ‘acquires’ is generic and unspecified. These verses (*Inf.* I, vv. 55-56) appear at the beginning of the poem and introduces a striking comparison between the *avaro* ‘greedy’ (namely, *quei che volentieri acquista e giugne ‘l tempo che perder lo face* ‘who willingly acquires, and the time comes that causes him to lose’), described through an explicative periphrasis, and the character of Dante. This comparative structure contributes to the psychological characterization of Dante’s character. In these verses, the omitted object refers to something not overtly mentioned in the text and must be inferred by the reader or listener through world knowledge and is thus annotated as [wk]. In Example (76), by contrast, the antecedent of the omitted object is explicitly present in the sentence. This verse is part of the description of Cerberus, a ‘cruel and uncouth monster’ (*fiera crudele e diversa*), whose purpose is to *graffia li spirti ed iscoia ed isquatra* ‘to rends the spirits, flays, and quarters them’. In this line, the last two verbs *iscoia* and *isquatra*, though transitive, do not overtly express the object. The antecedent of the omitted object is *li spirti* ‘the spirits’, which is explicitly mentioned only with the first verb *graffia* ‘rend’.

In addition to this information concerning the omitted object and its referent, further information relevant to the analysis was automatically extracted, including morpho-syntactic information of the verb licensing the object omission (for further details, refer to Section 5.2).

#### 4.2.3 Reconstruction

Before turning to the specifics of reconstruction, it is important to address two fundamental questions in detail: is reconstruction necessary for the analysis of ellipsis? And more broadly, is reconstruction necessary at all?

Reconstruction is not intrinsic to analysis, as an analysis can exist independently of it, whereas a reliable reconstruction necessarily depends on an underlying analysis. While it is possible to discuss results without engaging in a reconstruction process, enabling cross-linguistic comparison of a language-independent phenomenon such as ellipsis remains a valuable and desirable goal. It fosters dissemination, promotes the discussion and comparison of ellipsis issues, and ultimately contributes to the advancement of knowledge. These are all core principles embraced by the UD framework,

which actively supports cross-linguistic comparison and encourages the use of data in NLP tools (De Marneffe et al. 2021).

For these reasons, there are multiple motivations for reconstructing ellipsis within UD. As discussed in Section 3.4.3, there is a clear need to improve ellipsis annotation, given that ellipsis is not explicitly addressed in the basic layer and that only a limited number of enhanced treebanks are currently available. This is largely due to the substantial effort required to develop and maintain a basic treebank, which often renders manual enhancement a lower priority. Moreover, automatically enhancing a UD treebank remains particularly challenging, especially when it involves generating empty nodes to account for ellipsis (Simi et al. 2018; Droганova et al. 2018). In addition, the need for reconstruction is further underscored by the fact that efforts to automatically parse ellipsis in dependency treebanks using large language models (LLMs) have so far produced unsatisfactory results (Čavar et al. 2024a,b). This has motivated the development of corpora specifically dedicated to ellipsis (Čavar et al. 2024a), again highlighting the importance in having data and reconstructing ellipsis. Therefore, while a fully automatic reconstruction of ellipsis remains unfeasible at the current state of the art, it is still possible to support human annotation through the design of workflows that facilitate manual annotation.

However, reconstruction does not come without efforts. Once established the importance of reconstruction, it is worth noting that the reconstruction step is arguably the most delicate to carry out. In the case of a historical text such as the *Divine Comedy*, for which no native speakers are available, the very notion of “reconstruction” may appear problematic, if not overtly speculative. As discussed before, however, the risk inherent in reconstruction is outweighed by its utility. Moreover, although it may seem self-evident, it is important to clarify that reconstructing empty nodes is an artificial procedure that does not, in any way, alter the original text. Another potential risk associated with reconstruction is that the process may turn into a creative exercise, leading to the introduction of elements that appear arbitrary or involve a degree of interpretative freedom. More specifically, one of the key issues that arises concerns the number of tokens that should be reconstructed. In some cases, the context leaves no room for doubt, as only one token is clearly implied, such as in the preceding Example (70) (repeated in (77) for clarity), where the elided material can only be the verb ‘to say’.

77) *Inf.* VI, vv. 40-43

«O tu (...)», mi *disse*, «riconoscimi (...)». E io \_\_ a lui: «L’angoscia che tu hai (...)

“O you (...)” he said to me, “recall me, (...)”. And I to him: “It is perhaps your anguish (...)

Other cases, however, are more complex and may involve a larger amount of material. This is the case in Example (78), where one may wonder whether it is permissible to reconstruct, along with the verb *resurgeranno* ‘shall rise again’, the locative phrase *del sepulcro* ‘from the sepulchre’ as well.

78) *Inf.* VII, vv. 56-57

*questi resurgeranno del sepulcro col pugno chiuso, e questi \_\_\_ coi crin mozzi.*

‘These from the sepulchre shall rise again With the fist closed, and these \_\_\_ with tresses shorn.’

Since the elided clause appears to share the same situational context, this would justify a more extensive reconstruction, namely not just *resurgeranno* ‘shall rise again’, but *resurgeranno del sepulcro* ‘from the sepulchre shall rise again’. However, although this method may appear intuitive, it presents a potential drawback: it can be difficult to determine where reconstruction should stop. In cases where the antecedent is the head of an embedded clause or a nominal modified by a relative clause, there may be a tendency to reconstruct the entire subtree of the antecedent, but this is not always warranted. Therefore, to avoid producing cemeteries of empty words, it is essential to adopt a clear methodology that ensures transparency and enables consistent reconstruction.

Finally, one last consideration is worth addressing before proceeding. Given that UD is a community-driven framework, the development of a valid cross-linguistic methodology is essential. For this reason, at the current stage of the work, reconstruction proposals, which are intended to be tested and potentially adopted by the community, will be limited to verbal predicate ellipsis. Although this choice may initially appear to narrow the scope of the research, it has been made with the intention of adopting a gradual approach, one that allows for community feedback and fosters dialogue around reconstruction strategies and ellipsis-related issues that may not emerge in a study for its nature limited to a single language. In this context, focusing initially on one type of ellipsis, verbal predicate ellipsis, appears to be a more judicious starting point, allowing the investigation to remain focused and manageable at this early stage. The choice to target verbal predicates for reconstruction is motivated by their central role in the UD framework, which adopts a predicate-centric approach, and by the substantial impact that verbal predicates have on syntactic structures. Since copulas and auxiliaries are treated as leaves in UD, their reconstruction is not included in this preliminary phase. Along with copulas and auxiliaries, the reconstruction of object ellipsis is also not addressed. This decision is motivated by two main reasons. First, it is consistent with the reconstruction approach to be outlined in [Subsection 4.2.3.1](#), which explicitly excludes the reconstruction of arguments of any kind. Second, it is primarily informed by the linguistic nature of

Old Italian, in which the absence of certain types of objects (the referential ones) is not a mere optional omission (Section 5.2). The following paragraph ([Section 4.2.3.1](#)) outlines the criteria applied in the reconstruction of verbal predicate ellipsis.

#### 4.2.3.1 Verbal predicate ellipsis

The reconstruction of verbal predicates involves the creation of an empty node in the treebank, assigned its own syntactic relation, thereby resulting in an enhanced dependency representation. However, this reconstruction follows a set of specific guidelines designed to address potential issues that may arise, particularly those related to the challenges of ellipsis discussed above (Section 4.1.3). The issues are addressed in three points: i) position; ii) antecedent information; iii) reconstructed material.

##### i) Position

With regard to the placement of the empty node, the proposed approach adheres to a principle of parallelism, namely positioning the empty node so as to mirror the structure of its antecedent. This entails replicating the word order found in the antecedent clause. Consider, for instance, Example (79):

79) *Inf.* III, vv. 115-117

*similmente il mal seme d'Adamo **gittansi** di quel lito ad una ad una, per cenni come augel  
\_\_ per suo richiamo.*

similarly, the evil seed of Adam descended from the shoreline one by one, when signaled, as a falcon—called—will come.

In Example (79), the antecedent is the verb *gittansi* ‘descended’, which follows the subject (*il mal seme d'Adamo* ‘the evil seed of Adam’). In accordance with the principle of parallelism, the empty node in the elided clause is likewise placed after the subject *augel* ‘falcon’: *come augel \_ per suo richiamo* (‘as a falcon \_ called’).

In cases where no explicit antecedent is available, it is appropriate to default to the canonical word order of the language, understood as the most frequently attested pattern. In Example (80), for instance, the antecedent is absent, and therefore the reconstruction cannot mirror its structure. In similar contexts, the empty node is inserted after the subject (*io* ‘I’ in Example (80)). This placement is motivated by the observation that similar constructions with an overt verb more frequently display

the verb in post-subject position. A search through the *Comedy*<sup>188</sup> confirms that, in constructions with an overt verb comparable to Example (80), the verb typically follows the subject.

80) *Inf.* I, v. 130

*E io \_\_ a lui: «Poeta, io ti richeggio*  
 And I \_\_ to him: “Poet, I thee entreat

ii) Antecedent information

When dealing with ellipsis, the antecedent constitutes a fundamental piece of information, and its identification is therefore essential for accurate analysis. In order to track antecedents regardless of their position in the text, the proposal is introducing a numerical identifier, referred to as “unique\_number”<sup>189</sup>. This identifier is assigned to each token in the treebank and encodes its position relative to the entire text, rather than to the individual sentence—as is the case of the `id` field in CoNLL-U (see Section 1.4.6). As such, the “unique\_number” remains independent of sentence boundaries and uniquely identifies the position of each token within the text. In the case of *Italian-Old*, the “unique\_number” is constructed by combining the `sent_id`, which encodes information about the *Canto* and sentence number, with the token’s *id*. Below, Example (81) of attribution of “unique\_number” in the *Misc* field is reported.

81) *Inf.* I, v. 1

*Nel mezzo del cammin di nostra vita*  
 ‘Midway upon the journey of our life’

```
# sent_id = Canto1-1
1-2 Nel (...) Canto=1|Comment=ne-1|Verso=1|UniqueNumber=Canto1-1_1-2
1 in (...) Canto=1|Comment=ne-1|Verso=1|UniqueNumber=Canto1-1_1
2 il (...) Canto=1|Comment=ne-1|Verso=1|UniqueNumber=Canto1-1_2
3 mezzo (...) Canto=1|Verso=1|UniqueNumber=Canto1-1_3
4-5 del (...) Canto=1|Comment=de-1|Verso=1|UniqueNumber=Canto1-1_4-5
4 di (...) Canto=1|Comment=de-1|Verso=1|UniqueNumber=Canto1-1_4
```

<sup>188</sup> The following query was designed to retrieve comparable structures:

```
pattern { Z [lemma="e"] }
pattern { X [lemma="dire"] }
pattern { X -[cc]-> Z }
pattern { X -[nsubj]-> Y; X </> Y }
```

Here, the expression `X </> Y` is varied to determine the position of the verb relative to the subject. The query using `X < Y`—indicating a pre-subject verb position—returned only one occurrence. However, in that instance, the reported speech precedes the main clause, and thus it does not correspond to the structure of Example (80). In contrast, the query with `X > Y`—where the verb follows the subject—returned 13 occurrences. These cases display a word order that mirrors that of Example (80), and can therefore provide frequency-based justification for reconstructing the example with a post-subject verb.

<sup>189</sup> It should be noted that, due to an oversight, the value “unique\_number” is currently labeled as “UniqueTokenId” in UD version 2.17; however, in the upcoming release it will be reverted to “unique\_number”, as also adopted in Corbetta (2025d).

5	il	(...)	Canto=1 Comment=de-1 Verso=1 UniqueNumber=Canto1-1_5
6	cammin	(...)	Canto=1 Verso=1 UniqueNumber=Canto1-1_6
7	di	(...)	Canto=1 Verso=1 UniqueNumber=Canto1-1_7
8	nostra	(...)	Canto=1 Verso=1 UniqueNumber=Canto1-1_8
9	vita	(...)	Canto=1 Verso=1 UniqueNumber=Canto1-1_9

As mentioned, the “unique\_number” (UniqueNumber) is formed by merging the sent\_id, e.g., “Canto1-1”, indicating the first sentence of *Canto* 1, with the token’s *id* (as encoded in the first column of the CoNLL-U file), using an underscore ( ) as a separator. In the case of multi-word tokens, the range of token *ids* is reported (as in Canto1-1\_1-2 for the token *Nel* ‘in the’ and Canto1-1\_4-5 for the token *del* ‘of the’).

Once the “unique\_number” is assigned to each token in the treebank, retrieving the antecedent becomes significantly more straightforward. Specifically, information about the antecedent is recorded in the *Misc* field of the reconstructed empty node. This includes an indication of whether the antecedent (*Antec*) is explicitly present in the text, encoded as *Antec=Yes* or *Antec=No*. In addition, the *Misc* field always specifies the position of the antecedent using the key *AntecPosit*, which points to the corresponding “unique\_number”. If the antecedent is present in the text, the value is set as *AntecPosit=<Antecedent’s\_unique\_number>*, identifying the precise token(s) corresponding to the antecedent. The value of *AntecPosit* may be either a single “unique\_number”, if the antecedent consists of a single word, as illustrated in Example (78) in Section 4.2.2.1 (here, Example (82)):

82) *Inf.* VII, vv. 56-57

*questi resurgeranno del sepulcro col pugno chiuso, e questi \_ coi crin mozzi.*

‘These from the sepulchre shall rise again With the fist closed, and these with tresses shorn.’

# sent_id = Canto7-247			
9	questi	(...)	Canto=7 Verso=56 UniqueNumber=Canto7-247_9
10	resurgeranno	(...)	Canto=7 Verso=56 UniqueNumber=Canto7-247_10
(...)			
20.1	_	(...)	<b>Antec=Yes AntecPosit=Canto7-247_10</b>

Alternatively, when the antecedent spans multiple words, *AntecPosit* includes a range of unique\_numbers, as shown in Example 79 in Section 4.2.2.1 (here, Example (83)):

83) *Inf.* III, vv. 115-117

*similmente il mal seme d'Adamo gittansi di quel lito ad una ad una, per cenni come augel \_ per suo richiamo.*

similarly, the evil seed of Adam descended from the shoreline one by one, when signaled, as a falcon—called—will come.

# sent_id = Canto3-105			
32	d'	(...)	Canto=3 SpaceAfter=No Verso=115 UniqueNumber=Canto3-105_32
33	Adamo	(...)	Canto=3 Verso=115 UniqueNumber=Canto3-105_33
34-35	gittansi	(...)	Canto=3 Verso=116 UniqueNumber=Canto3-105_34-35

34	gittano	(...)	Canto=3 Verso=116 UniqueNumber=Canto3-105_34
35	si	(...)	Canto=3 Clitic=Yes Verso=116 UniqueNumber=Canto3-105_35
(...)			
47	augel	(...)	Canto=3 Verso=117 UniqueNumber=Canto3-105_47
47.1	_	(...)	Antec=Yes  <b>AntecPosit=Canto3-105_34-35</b>
48	per	(...)	Canto=3 Verso=117 UniqueNumber=Canto3-105_48

Thanks to the use of the `AntecPosit` annotation, this method allows for the retrieval of antecedents across sentence boundaries, as illustrated in Example (84), where the antecedent is located in a preceding sentence (Example (77) in Section 4.2.2.1):

84) *Inf.* VI, vv. 40-43

«*O tu (...)»*, *mi disse*, «*riconoscimi (...)»*. *E io \_ a lui*: «*L'angoscia che tu hai (...)*  
 “O you (...)” he said to me, “recall me, (...)”. And I to him: “It is perhaps your anguish (...)”

```
# sent_id = Canto6-209
...
12    mi      (...)   Canto=6|Clitic=Yes|Verso=41|UniqueNumber=Canto6-209_12
13    disse  (...)   Canto=6|SpaceAfter=No|Verso=41|UniqueNumber=Canto6-209_13
...

# sent_id = Canto6-210
1      E      (...)   Canto=6|Verso=43|UniqueNumber=Canto6-210_1
2      io     (...)   Canto=6|Verso=43|UniqueNumber=Canto6-210_2
2.1   _      (...)   Antec=Yes|AntecPosit=Canto6-209_13
3      a      (...)   Canto=6|Verso=43|UniqueNumber=Canto6-210_3
4      lui   (...)   Canto=6|SpaceAfter=No|Verso=43|UniqueNumber=Canto6-210_4
```

In contrast, if the antecedent is not explicitly present in the text and must be inferred from general world knowledge, this is indicated by the value `AntecPosit=wk`, where `wk` stands for world knowledge. Example (85) (Example (80) in Section 4.2.2.1) illustrates a case in which the antecedent is reconstructed based on world knowledge:

85) *Inf.* III, v. 13

*Ed elli \_ a me, come persona accorta*:  
 ‘And he to me, as one who comprehends’

```
# sent_id = Canto3-74
1      Ed     (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_1
2      elli  (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_2
2.1   _      (...)   Antec=No|AntecPosit=wk
3      a      (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_3
4      me     (...)   Canto=3|SpaceAfter=No|Verso=13|UniqueNumber=Canto3-74_4
```

Referencing the “unique\_number” in this explicit manner enables the automatic retrieval of all morpho-syntactic information associated with the antecedent, as the direct link to the corresponding token(s) in the treebank also grants access to its full linguistic annotation. This approach ensures a precise, systematic, and linguistically rich representation of antecedents in elliptical constructions. It is important to highlight a key methodological decision: the retrieval and representation of the antecedent in the `Misc` field of the empty node does not include any argumental dependents. Rather,

the annotation of the antecedent is restricted to the predicate itself, meaning that only auxiliaries and clitics inherently associated with the lexical verb are taken into account. The decision to intentionally exclude arguments from antecedent retrieval<sup>190</sup> was made in order to avoid overcomplicating the annotation process. It is worth noting, however, that once the predicate antecedent is identified, the predicate-centric nature of UD syntactic annotation allows for the automatic retrieval of all its dependents as subtrees of the predicate node, thereby also accessing all the relevant morpho-syntactic information already encoded in the treebank.

For instance, the antecedent may be discontinuous in terms of linear word order, specifically, when a predicate and its modal verb/auxiliary are selected, an intervening word functioning as an argument or adjunct may occur between them. Since the annotation guidelines explicitly restrict antecedent marking to the predicate, along with its auxiliaries and clitics, any intervening elements, such as arguments or adjuncts, must be excluded, even in these cases. To indicate this discontinuity, a hyphen (-), is used within the antecedent span. An example is provided below:

86) *Inf.* XXIV, vv. 32-33

*ché noi a pena, ei \_\_ lieve e io \_\_ sospinto, potavam sù montar di chiappa in chiappa.*

‘For hardly we, he light, and I (pushed) upward, Were able to ascend from jag to jag.’

```
# sent_id = Canto24-862
14    ei          (...)   Canto=24|Verso=32|UniqueNumber=Canto24-862_14
14.1  _            (...)   Antec=Yes|AntecPosit=Canto24-862_20-22
15    lieve       (...)   Canto=24|Verso=32|UniqueNumber=Canto24-862_15
16    e           (...)   Canto=24|Verso=32|UniqueNumber=Canto24-862_16
17    io         (...)   Canto=24|Verso=32|UniqueNumber=Canto24-862_17
17.1  _            (...)   Antec=Yes|AntecPosit=Canto24-862_20-22
18    sospinto   (...)   Canto=24|SpaceAfter=No|Verso=32|UniqueNumber=Canto24-862_18
19    ,          (...)   _
20    potavam    (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_20
21    sù         (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_21
22    montar     (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_22
23    di         (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_23
24    chiappa    (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_24
25    in         (...)   Canto=24|Verso=33|UniqueNumber=Canto24-862_25
26    chiappa    (...)   Canto=24|SpaceAfter=No|Verso=33|UniqueNumber=Canto24-862_26
```

Example (86) illustrates a case of double predicate ellipsis<sup>191</sup>, where the omitted predicates share a postcedent composed of the verb *montar* ‘ascend’ and the auxiliary *potavam* ‘were able’. An adverb, *sù* ‘up’, depends on the verb and appears between the auxiliary and the lexical verb, thereby interrupting the linear sequence and complicating the tracking of the antecedent. To indicate this

<sup>190</sup> This methodological choice diverges from the Prague Dependency Treebank approach (refer to Section 3.4.2), which includes the reconstruction of arguments as part of its annotation strategy (<https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/ch06s12s01.html#elipsa1.1>).

<sup>191</sup> In this case as well, the frequency of the verb’s position relative to its oblique adjuncts has been taken into account in determining where to place the elided verbs. A query revealed that oblique adjuncts frequently follow the verb *montare* ‘ascend’, with 10 such occurrences attested out of 12.

discontinuity, a hyphen is used to separate the `unique_numbers` associated with the antecedent (`AntecPosit=Canto24-862_20-22`).

Finally, in order to remain consistent with the decision not to mark arguments or adjuncts, in cases involving complex predicates, such as ‘to make shield of’, where the tokens can be interpreted as a multi-word expression conveying a specific meaning, only the verb is marked as the antecedent.

### iii) Reconstructed material

Once an empty node has been placed in the syntactic structure, respecting, where possible, the principle of parallelism, and once information about its antecedent has been recorded in the `Misc` field, the final step is to specify, when permitted by the context, how many empty nodes are required and, for each one, which linguistic features should be reconstructed.

Concerning this aspect, the choice has been to conduct a partial reconstruction in order not to have too many empty nodes in the structure. More precisely, the approach involves creating a single empty node for each instance of ellipsis, regardless of the size of the antecedent in terms of syntactic nodes. The full extent of the antecedent is instead indicated in the `Misc` field by means of a range, as previously explained. The decision to reconstruct only a single empty syntactic node for each ellipsis, and to delegate information about the full antecedent span to the `Misc` field without syntactically reconstructing it, is motivated by the fact that, from a syntactic perspective, only the predicate is required for the tree to be structurally well-formed, since all other elements are dependents, leaves, of that predicate. This choice also represents a middle ground between constituency-based annotation, where empty nodes are widely adopted, and dependency-based annotation, where empty nodes are typically. Although the ultimate goal would be to move toward full reconstruction, a moderate solution has been adopted at this initial stage in order to facilitate its introduction within the community and to test its portability across other languages.

Concerning the linguistic information encoded in the empty node, the decision is to populate only the form, when explicitly recoverable from the text (example 87) or, for specific cases, inferable (example 88), as well as the syntactic fields (properly annotated in the `deps` field). If the surface form of the reconstructed element differs from that of the antecedent, it is not annotated directly, to avoid over-reconstruction. Instead, this discrepancy is indicated in the `Misc` field using the parameter `Form=Diff` (Example 88). Also in line to avoid over-reconstruction, the form of cases where the antecedent is not textually explicit, or in cases of dubious solution (Example 89) the form is not inserted, simply filled with syntactic information. The following are examples of empty node reconstruction based on the previously discussed scenarios, beginning (Example 87) with a case of

complete reconstruction, where form and deps are all specified, since the elided material is a copy of the antecedent:

87) *Inf.* VII, vv. 56-57

*questi resurgeranno del sepulcro col pugno chiuso, e questi \_\_\_ coi crin mozzi.*

‘These from the sepulchre shall rise again With the fist closed, and these \_\_\_ with tresses shorn.’

```
# sent_id = Canto7-247
10   resurgeranno (...)   Canto=7|Verso=56|UniqueNumber=Canto7-247_10
(...)
20.1 resurgeranno   _ _ _ _ _ 10:conj Antec=Yes|AntecPosit=Canto7-247_10
```

Example 2 (already seen in Section 4.2.3.1) presents a case in which the elided material differs in surface form from the antecedent (example 88). As a result, this discrepancy is recorded in the *Misc* field using the parameter *Form=Diff*.

88) *Inf.* X, vv. 1-3

*Ora sen va per un secreto calle, tra 'l muro de la terra e li martiri, lo mio maestro, e io \_\_\_ dopo le spalle.*

‘Now onward goes, along a narrow path Between the torments and the city wall, My Master, and I (follow) at his back.’

```
# sent_id = Canto10-345
(...)
4     va      (...)   Canto=10|Verso=1|UniqueNumber=Canto10-345_4
(...)
24    e       (...)   Canto=10|Verso=3|UniqueNumber=Canto10-345_24
25    io      (...)   Canto=10|Verso=3|UniqueNumber=Canto10-345_25
25.1 vado    _ _ _ _ _ 0:root Antec=Yes|AntecPosit=Canto10-345_2-4|Form=Diff
(...)
```

Example (89) illustrates a case in which the antecedent is not recoverable from the text, but must be inferred from world knowledge. As in example (88), the *form* field is not populated.

89) *Inf.* III, v. 13

*Ed elli \_ a me, come persona accorta:*

‘And he to me, as one who comprehends’

```
# sent_id = Canto3-74
1     Ed      (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_1
2     elli   (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_2
2.1  _        (...)   Antec=No|AntecPosit=wk
3     a      (...)   Canto=3|Verso=13|UniqueNumber=Canto3-74_3
4     me     (...)   Canto=3|SpaceAfter=No|Verso=13|UniqueNumber=Canto3-74_4
```

Finally, in cases where the reconstruction is uncertain, as in Example (90), only the syntactic fields are filled. Moreover, the identification of the antecedent includes a forward slash (/) to indicate multiple possible antecedents.

90) *Inf.* XVII, vv. 10-12

*La faccia sua era faccia d'uom giusto, tanto benigna avea di fuor la pelle, e d'un serpente  
\_\_ tutto l'altro fusto*

‘The face was as the face of a just man, Its semblance outwardly was so benign, And of a  
serpent \_\_ all the trunk beside.’

```
# sent_id = Canto17-593
(...)
4      era      (...)      Canto=17|Verso=10|UniqueNumber=Canto17-593_4
(...)
12     avea     (...)      Canto=17|Verso=11|UniqueNumber=Canto17-593_12
(...)
21     serpente (...)      Canto=17|Verso=12|UniqueNumber=Canto17-593_21
21.1  _         (...)      Antec=Yes|AntecPosit=Canto17-593_4/12
22     tutto    (...)      Canto=17|Verso=12|UniqueNumber=Canto17-593_22
```

### 4.3 Summary

In Chapter 4, the treatment of ellipsis is first addressed from a theoretical perspective (Section 4.1), and subsequently from a practical standpoint (Section 4.2). The process of handling ellipsis is divided into three main steps: detection, identification, and reconstruction. For each of these steps, both a theoretical discussion and a practical description are provided. This methodology will be shared with the Universal Dependencies community to gather feedback and enable its application to other languages, thereby contributing to a richer cross-linguistic analysis of ellipsis.

In the following section, I will carry out an analysis of predicate ellipsis, along with object ellipsis, in the *Divine Comedy*, adopting the technique described in this Chapter.

## 5) Analysis: predicate ellipses and object omission

In this Chapter, it is presented both a quantitative and a qualitative analysis of the instances of ellipses (predicate and object) extracted from the data, as outlined in [Chapter 4](#). Specifically, [Section 5.1](#) addresses predicate ellipsis, while [Section 5.2](#) examines object omission. The quantitative analysis relies on the annotation of the treebank, whereas the qualitative analysis consists of the manual examination of all instances of ellipsis retrieved through the queries, followed by a manual clustering according to the patterns described in each subsection. The purpose of these analyses is to provide a data-driven classification of the types of predicates ellipsis and object omission identified in the text under investigation, the *Divine Comedy*, with the dual aim of comparing them with existing literature on the phenomenon and shedding further light on ellipsis.

It is important to emphasize once again that the analysis of predicate and object omission is specifically tailored to the *Divine Comedy*. This choice is primarily motivated by the lack of other annotated Old Italian texts that would allow for a comparable investigation (see Sections 1.2 and 1.4.5), although the development of such resources is envisaged for future work. Consequently, the analysis necessarily reflects the characteristics of the text type under consideration. As discussed in Section 1.1, the *Divine Comedy* is a poetic work and is therefore subject to metrical patterns and rhyme constraints, which may at times give rise to structural choices that would appear artificial outside a poetic context. In order to determine whether the types of ellipsis identified in the *Divine Comedy* also occur in non-literary texts, where metrical structure and the use of rhetorical figures do not impose comparable constraints, further investigation is required. Such analyses are therefore left to future work.

## 5.1 Predicate ellipses

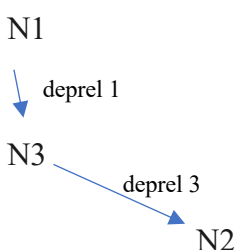
The analysis of predicate ellipsis in the *Divine Comedy* aims to account for all instances of predicate ellipsis by automatically retrieving them from the *Italian-Old* treebank. Accordingly, it considers both cases annotated with the `orphan` dependency relation and those marked through promotion (see Sections 3.4.3.1 and 3.4.3.2). Thanks to the automatic extraction technique described in [Section 4.2.1.1](#), and with the support of the information encoded in the treebank, it was possible to extract the following features, which served as the basis for the subsequent clustering:

- The effective presence of ellipsis, identified through its sentence position and its location within the text, thereby providing the textual context in which the ellipsis occurs;
- The syntactic position of the ellipsis, inferred from the syntactic position of the parent of the elliptical node.

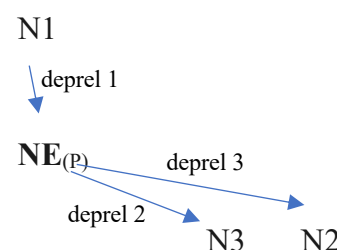
All the features listed above were extracted from the basic version of the *Italian-Old* treebank, as the implementation of enhanced annotations, following the guidelines proposed in Section 4.2.3, is still in progress and under discussion within the community (Corbetta et al. 2025d).

Once the instances of ellipsis were extracted through the queries described in Section 4.2.1.1, the sentences containing them were retrieved by means of the `text` and the `sent-id` fields provided for each syntactic tree (see Section 2.1.2.2.4). The `text` field records the sentence in which the ellipsis is located, while the `sent-id` specifies the identifier of the sentence containing the ellipsis, thereby pinpointing its exact position within the treebank. For a more fine-grained localization, the `Misc` field, available for each node in the tree (see Section 1.4.6), enables the precise identification of the *Canto* and verse where the ellipsis occurs. This information provides an accurate textual reference that goes beyond the treebank itself and ensures a reliable system for tracking and locating ellipsis, grounded in the internal structure of the work.

With respect to the syntactic position of the ellipsis, since elliptical nodes are not represented in the trees under basic annotation (refer to Section 3.4.3), the position was inferred from the parent node of the elided element. This node is significant, as it marks the syntactic information of the elided element itself. To clarify this point, the following notation is adopted (reported in Images 62 and 63):



**Image 62.** non-reconstructed predicate ellipsis



**Image 63.** reconstructed predicate ellipsis

Image 63 illustrates an elliptical construction in which predicate ellipsis is explicitly represented by the node  $NE_{(P)}$  and displaying the syntactic relations, indicated through arrows and labeled *deprel 1*, *deprel 2*, and *deprel 3*.  $N1$  is the parent node of  $NE_{(P)}$  and  $N2$  and  $N3$  are its children.

In contrast, under standard conditions (namely, in basic UD annotation), ellipsis is not explicitly represented in the syntactic tree (Section 3.4.3). In this case, the syntactic structure corresponds to Image 62, where no  $NE_{(P)}$  node is present. As thoroughly discussed in Section 4.2.1.1 and summarized here, the syntactic information that should belong to the elided node,  $NE_{(P)}$ , (specifically, *deprel 1*), is in Image 1 assigned instead to  $N3$ , which is promoted to assume the syntactic relation otherwise linked to  $NE_{(P)}$ . Despite this promotion, the node that ultimately conveys the syntactic information of the elided element,  $NE_{(P)}$  in Image 63 and  $N3$  in Image 62, remains dependent on  $N1$ , which is always the parent node. The syntactic relation of  $N1$  can be retrieved using querying tools such as Udapi (Popel et al. 2017). This makes it possible to recover the syntactic relation needed for analyzing the syntactic position of elliptical constructions, which forms the basis of the quantitative analysis presented in Section 5.1.2.

As already explained in Section 4.2.1.1, the queries differ depending on whether the ellipsis is resolved through promotion or *orphan* strategies. Furthermore, an additional distinction was introduced for cases involving coordination, in line with the discussion in Section 4.2.1.1.2.2. The full set of queries is provided in the Appendix. This information underpins the analysis and classification of predicate ellipsis along the syntactic dimension, with the aim of identifying recurrent patterns and gaining further insight into its manifestation and characterization.

### 5.1.1 General overview

This Section provides a general overview of predicate ellipsis in the *Divine Comedy*, illustrating the range of scenarios in which it can occur. These include cases involving an antecedent, a postcedent, and instances recoverable through world knowledge. The purpose of this section is to offer background information relevant to the potential NLP treatment of predicate ellipsis, as described in Chapter 4, by highlighting the scenarios that should be taken into account.

In the *Divine Comedy*, predicate ellipsis is attested both with antecedents and postcedents, as shown in Examples (91) and (92), respectively. In the examples, the dash marks the ellipsis site<sup>192</sup>.

(91) *Purg.* XII, vv. 113-114

---

<sup>192</sup> As mentioned in Section 4.1.1, determining the position of the elided site is not straightforward. The criteria adopted here follow the proposal based on the principle of parallelism (see Section 4.2.3.1).

*ché quivi per canti s'entra, e là giù \_\_\_ per lamenti feroci.*

‘For here it is with song one enters; down there, (it is) \_\_\_ with savage lamentations.’

(92) *Inf.* XXIX, vv. 67-68

*Qual \_\_\_ sopra 'l ventre e qual sopra le spalle l'un de l'altro giacea*

‘This \_\_\_ on the belly, that upon the back One of the other **lay**’

In Example (91), the antecedent is marked in bold in the text (*s'entra* ‘one enters’). The coordinated clause (*e là giù per lamenti feroci* ‘and down there, (it is) with savage lamentations’) is syntactically analyzed as containing an omitted verb, identical to the verb of the main clause, *s'entra* ‘one enters’. From a syntactic perspective, the coordinated clause is clearly elliptical: the two adverbs *là* ‘there’ and *giù* ‘down’ specify the location of the action and therefore depend on the elided verb, while the prepositional phrase *per lamenti feroci* ‘with savage lamentations’ functions as an oblique, also depending on the omitted verb.

A similar scenario is presented in Example (92), where, however, the ellipsis occurs in the main clause (*Qual sopra 'l ventre* ‘This on the belly’), while the verb *giacea* ‘lay(s)’ is overtly expressed in the coordinated clause (*e qual sopra le spalle l'un de l'altro giacea* ‘that upon the back One of the other **lay**’). This is therefore a case in which the ellipsis involves a postcedent.

Examples (91)-(92) illustrate cases in which the elided material is identical to the antecedent, thereby conforming to the identity principle (Section 3.1.4). However, this is not always the case. Such mismatches are particularly attested in comparative constructions, although these are not the only syntactic contexts in which they occur. Variation in the elided material with respect to its antecedent or postcedent may involve gender (Example 93), tense (Example 94), or even person (Example 95).

(93) *Purg.* XXXIII, vv. 79-81

*Sì come cera \_\_\_ da suggello, che la figura impressa non trasmuta, segnato è or da voi lo mio cervello.*

‘As by a signet \_\_\_ (is) the wax Which does not change the figure stamped upon it, My brain **is now imprinted** by yourself.’

In Example (93), the comparative clause (*Sì come cera da suggello* ‘As by a signet (is) the wax’) is elliptical, as the verb is not overtly expressed. The elided material can be reconstructed from the postcedent *segnato è* ‘is imprinted’. However, unlike the previous instances (Examples (91)–(93)), the reconstruction is not fully identical: the past participle *segnat-o* ‘imprinted-MASC.SG’ in the

reconstructed material should exhibit feminine agreement to match the subject *cera* ‘wax’, which is feminine in Italian.

(94) *Purg.* XXVI, vv. 134-135

*disparve per lo foco, come \_\_\_ per l' acqua il pesce andando al fondo.*

‘he **vanished** in the fire as fish \_\_\_ in water going to the bottom.’

In Example (94), the ellipsis occurs within a comparative structure, where the reconstruction, based on the antecedent *disparve* ‘vanished’, expressed in the past tense, would nonetheless require a shift in tense, namely to the present. This temporal shift appears to be characteristic of comparative constructions, which generalize the action by placing it in an indefinite present context. Ellipsis in comparative is not directly addressed in this study, as it warrants a dedicated, in-depth analysis.

(95) *Inf.* XIX, v. 114

*se non ch'elli \_\_\_ uno, e voi ne orate cento?*

‘that he \_\_\_ one, and ye a hundred **worship**?’

In Example (95), ellipsis occurs in a clause whose subject is the singular pronoun *elli* ‘he’, while its postcedent *orate* ‘worship’ shows plural agreement, consistent with its plural subject *voi* ‘ye’. Consequently, the reconstructed material must account for this shift in person agreement, which results in a change in the verbal form, namely *ora* ‘worships’.

All the mismatches observed in Examples (93)-(95) demonstrate that, in certain elliptical constructions, morphosyntactic identity between the elided material and its antecedent or postcedent is not strictly maintained. This underscores the importance of human verification in the reconstruction process.

Finally, another issue, also noted in Section (4.2.3), concerns cases where the antecedent is absent and must therefore be reconstructed through world knowledge, as illustrated in Example (96):

(96) *Par.* XXX, vv. 25-27

*ché, come \_\_\_ sole in viso che più trema, così lo rimembrar del dolce riso la mente mia da me medesmo scema.*

‘for like \_\_\_ the sun that strikes the frailest eyes, so (does) the memory of her sweet smile deprive me of the use of my own mind.’

In Example (96), Dante draws a comparison between the effect of the sun on weakened sight and the memory of Beatrice’s smile, which overwhelms the poet’s mind and renders him unable to speak of it. In this case, no antecedent is available to reconstruct the ellipsis in the comparative clause; rather, the ellipsis is filled by the reader with a generic verb such as *fare* (‘to do/make’), despite the fact that no trace of it is present in the text itself.

To conclude, this section has provided an overview of the main aspects of predicate ellipsis in the *Divine Comedy*, illustrating a range of possible scenarios in which the phenomenon occurs, with the aim of identifying issues that should be carefully addressed in NLP tasks dealing with ellipsis. More specifically, the aspects that emerged from the analysis of Old Italian and that require particular attention are:

- the position of the antecedent, which may either precede or follow the ellipsis site;
- cases of non-identical reconstruction with respect to person, tense, or gender;
- instances where world knowledge is necessary to recover the antecedent, requiring understanding beyond the phrasal and textual levels;

At the current stage, these difficulties necessitate careful human pre-processing. Although this study is limited to a single text and language, thus leaving open cases that do not occur in this specific corpus or linguistic context, it nonetheless has the merit of providing a detailed overview based on real data. Further studies on predicate ellipsis in other languages are advisable, as they will shed light on additional issues<sup>193</sup>.

### 5.1.2 Syntactic context of the predicate ellipsis

As mentioned in Section 5.1, the syntactic information of the elided predicate was automatically extracted (queries are provided in Appendix). The data used for this analysis, split by *Cantica*, are provided on GitHub<sup>194</sup>. Since the retrieval procedures for `orphan` and `promotion` differ (see Sections 4.2.1), their results are presented separately in Table 7 and Table 8, showing the number of occurrences of predicate ellipsis for the respective dependency relations, divided into `orphan` (Table 7) and `promotion` (Table 8) categories. In both tables, the results are further broken down according to the *Cantica* (*Inferno*, *Purgatorio*, *Paradiso*), with the overall count reported in the column labeled *Comedy*. The column labeled *Total* indicates the total number of predicate ellipses for each *Cantica* and for the whole *Comedy*. Finally, Table 9 provides a comprehensive summary of all predicate ellipses, irrespective of the orphan/promotion distinction.

---

<sup>193</sup> A preliminary investigation has already been carried out on Latin (Albertelli et al. 2025).

<sup>194</sup> <https://github.com/ClaudiaCorbe/Divine-Comedy-of-Dante-Alighieri>.

<b>orphan</b>			
<i>Inferno</i>	<i>Purgatorio</i>	<i>Paradiso</i>	<i>Comedy</i>
66 root	37 root	38 conj	141 root
33 conj	33 conj	29 advcl:cmp	95 conj
18 advcl:cmp	17 advcl:cmp	16 root	51 advcl:cmp
18 parataxis	6 parataxis	3 parataxis	27 parataxis
6 advcl	4 advcl	2 acl:relcl	11 acl:relcl
6 acl:relcl	3 acl:relcl		10 advcl
2 ccomp:reported	1 csubj		2 ccomp:reported
			1 csubj
<b>Total: 149</b>	<b>Total: 101</b>	<b>Total: 88</b>	<b>Total: 338</b>

**Table 7.** Distribution of predicate ellipsis by syntactic dependency relation (**orphan**)

<b>promotion</b>			
<i>Inferno</i>	<i>Purgatorio</i>	<i>Paradiso</i>	<i>Comedy</i>
79 advcl:cmp	64 conj (clausal)	88 advcl:cmp	226 conj (clausal)
75 conj (clausal)	59 advcl:cmp	87 conj (clausal)	226 advcl:cmp
32 root	53 ccomp:reported	32 root	103 root
22 advcl	39 root	19 ccomp:reported	92 ccomp:reported
20 ccomp:reported	28 parataxis	16 parataxis	53 parataxis
10 acl:relcl	16 advcl	9 advcl	47 advcl
9 parataxis	3 ccomp	3 ccomp	15 acl:relcl
8 ccomp	2 acl:relcl	3 acl:relcl	14 ccomp
3 acl	1 advcl:relcl	2 csubj	3 csubj
2 advcl:relcl			3 advcl:relcl
1 csubj			3 acl
<b>Total: 261</b>	<b>Total: 265</b>	<b>Total: 259</b>	<b>Total: 785</b>

**Table 8.** Distribution of predicate ellipsis by syntactic dependency relation (**promotion**)

<b>Total predicate ellipsis</b>			
<i>Inferno</i>	<i>Purgatorio</i>	<i>Paradiso</i>	<i>Comedy</i>
108 conj (clausal)	97 conj (clausal)	125 conj (clausal)	330 conj (clausal)
98 root	76 advcl:cmp	117 advcl:cmp	290 advcl:cmp
97 advcl:cmp	76 root	48 root	222 root
28 advcl	53 ccomp:reported	19 ccomp:reported	94 ccomp:reported
27 parataxis	34 parataxis	19 parataxis	80 parataxis
22 ccomp:reported	20 advcl	9 advcl	57 advcl
16 acl:relcl	5 acl:relcl	5 acl:relcl	26 acl:relcl
8 ccomp	3 ccomp	3 ccomp	14 ccomp
3 acl	1 csubj	2 csubj	4 csubj
2 advcl:relcl	1 advcl:relcl		3 acl
1 csubj			3 advcl:relcl
<b>Total: 410</b>	<b>Total: 366</b>	<b>Total: 347</b>	<b>Total: 1123</b>

**Table 9.** Total distribution of predicate ellipsis by syntactic dependency relation (**Total**)

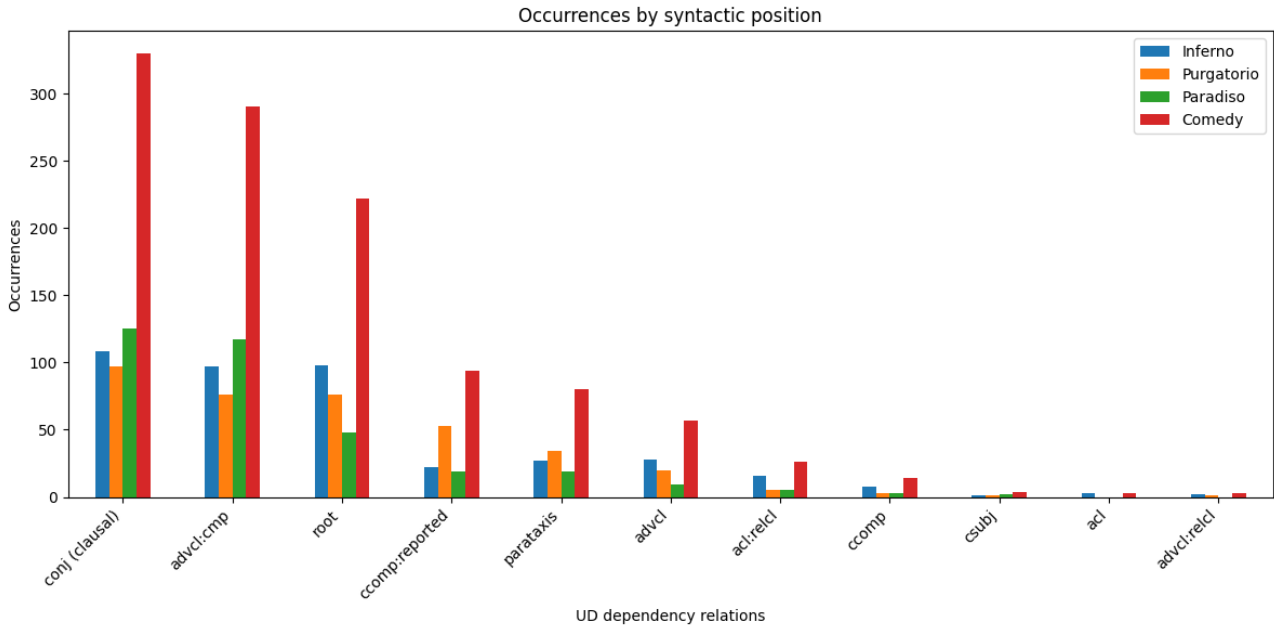
From an initial comparison of Tables 7 and 8, predicate ellipses annotated via promotion are more prevalent than those annotated via `orphan` relations (785 vs. 338 instances, respectively). A qualitative analysis of these cases will be presented in Section 5.1.3; for now, these tables underscore the importance of considering promotion when accounting for ellipsis.

**Table 7. (orphan):** Focusing on the `orphan` data, *Inferno* shows the largest number of `orphan` constructions (149), followed by *Purgatorio* (101) and *Paradiso* (88), the latter exhibiting comparatively few such structures. Within the `orphan` strategy, `root`, `conj`, and `advcl:cmp` dominate (141, 95, 51), together accounting for roughly 85% of orphan ellipses. A noteworthy pattern, discussed in Section 5.1.4, is the unusually high number of `root`-position orphans in *Inferno* (66), compared with *Purgatorio* (37) and *Paradiso* (16).

**Table 8. (promotion):** Turning to promotion, promoted structures are particularly frequent in clausal coordination (`conj`; 226) and in comparative clauses (`advcl:cmp`; 226), the latter largely reflecting annotation practices (see Section 2.2.2), followed by root position (`root`; 103) and reported speech (`ccomp:reported`; 92). Across the *Cantiche*, the distribution is more homogeneous than for `orphan`: *Purgatorio* shows the most promotion-based ellipses (265), followed by *Inferno* (261) and *Paradiso* (259).

**Table 9. (aggregate):** In the aggregate counts, the three most frequent `deprels` remain stable across *Cantiche*: clausal coordination (`conj`; 330), comparative clauses (`advcl:cmp`; 290), and root position (`root`; 222). Together they account for about three quarters of all cases (approximately 75%). Another notable finding is that, across the *Cantiche*, *Inferno* exhibits the highest number of ellipses (410), followed by *Purgatorio* (366) and *Paradiso* (347).

To better visualize the patterns summarized in Tables 7–9, Image 64 compares the number of occurrences by dependency relation across the three *Cantiche* (*Inferno*, *Purgatorio*, *Paradiso*) and juxtaposes them with the overall totals for the *Comedy*.



**Image 64.** Comparison of predicate ellipsis divided by deprel and *Cantiche*

Considering the data presented above, the syntactic environments in which predicate ellipsis most frequently occurs are clausal coordination, comparative structures, and main-clause (*root*) position. Together, these contexts account for about 75% of all cases, though they are not the only loci of ellipsis. A second, smaller block comprises *parataxis* (often used in annotation to reflect strong editorial punctuation), direct speech (*ccomp:reported*), and adverbial subordination (*advcl*), amounting to roughly one fifth of the data (approximately the 21%). Finally, a small residual set consists of relative clauses (*acl:relcl*), complement clauses (*ccomp*), subject clauses (*csubj*), and nominal and adverbial modifiers (*acl* and *advcl:relcl* with a relative interpretation), which collectively contribute less than 4%.

### 5.1.2.1 Quantifying the Impact of Predicate Ellipsis in and across the Comedy

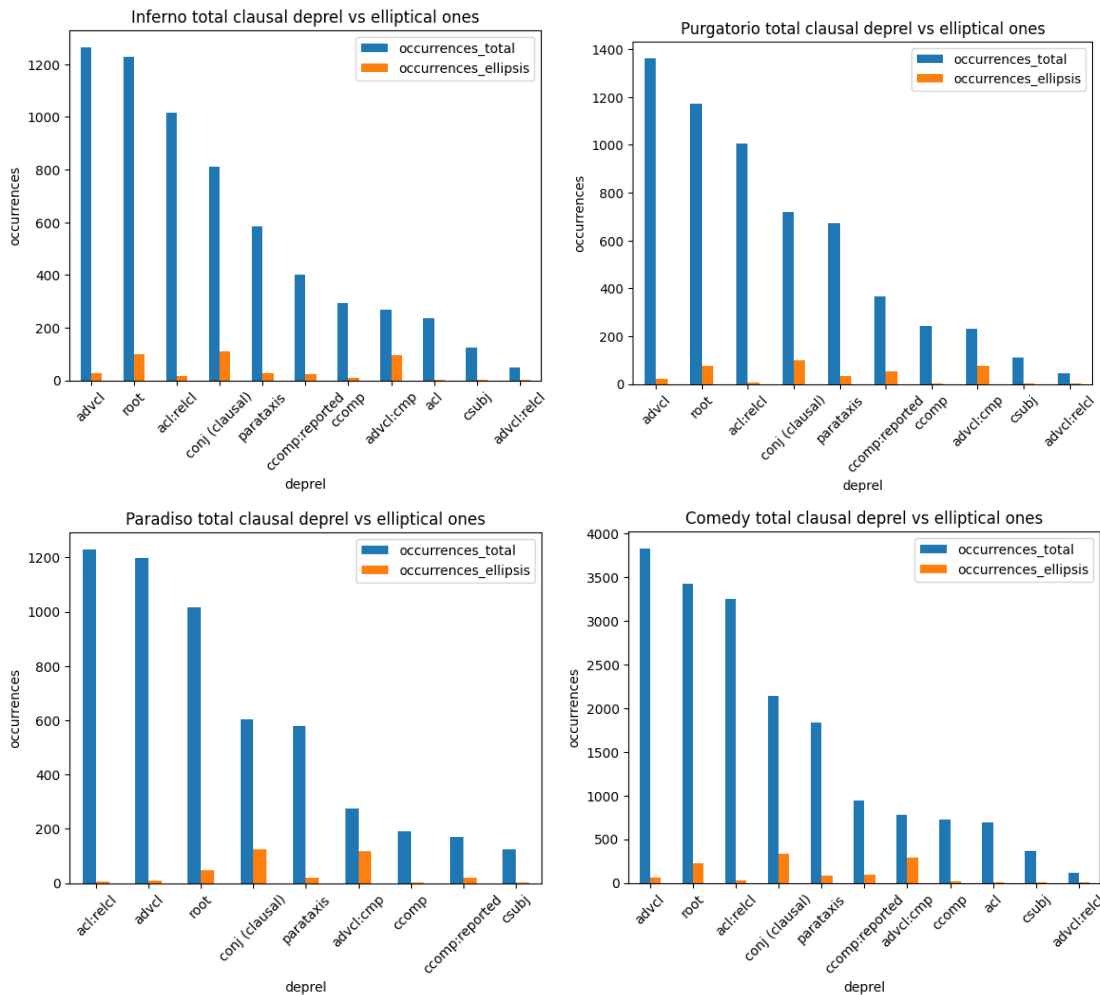
To quantify the impact of predicate ellipsis, the totals in Table 9 is normalized (overall counts of predicate ellipses) by the corresponding totals of clausal deprel occurrences reported in Table 10. The clausal totals were computed with the queries listed in the Appendix. As in the other extractions, coordination was retrieved separately so as to distinguish clausal from phrasal coordination.

Total Clausal dependency relations							
<i>Inferno</i>		<i>Purgatorio</i>		<i>Paradiso</i>		<i>Comedy</i>	
1266	advcl	1364	advcl	1231	acl:relcl	3829	advcl
1228	root	1174	root	1199	advcl	3419	root
1015	acl:relcl	1004	acl:relcl	1017	root	3250	acl:relcl
810	conj (clausal)	719	conj (clausal)	605	conj (clausal)	2134	conj (clausal)

586	parataxis	671	parataxis	578	parataxis	1835	parataxis
400	ccomp:reported	368	ccomp:reported	275	advcl:cmp	936	ccomp:reported
292	ccomp	244	ccomp	233	acl	773	advcl:cmp
269	advcl:cmp	229	advcl:cmp	189	ccomp	725	ccomp
235	acl	222	acl	168	ccomp:reported	690	acl
124	csubj	112	csubj	125	csubj	361	csubj
48	advcl:relel	45	advcl:relel	26	advcl:relel	119	advcl:relel
				3	csubj:pass	3	csubj:pass

**Table 10.** Total distribution of clausal dependency relation

To aid interpretation, Image 65 visually compares the clausal totals in Table 10 with the ellipsis counts in Table 9. Across dependency relations, the gap between total occurrences (blue) and elliptical instances (orange) is substantial. Panels in Image 65 report the results for *Inferno*, *Purgatorio*, *Paradiso*, and the *Comedy* as a whole, respectively.



**Image 65.** Comparison of total clausal dependency relations vs. predicate-ellipsis instances

For each *Cantica*, the **relative frequency** of ellipsis by deprel is computed, defined as the number of elliptical instances with that deprel divided by the total clausal occurrences of the same deprel.

Numerators are taken from Table 9 (ellipsis counts) and denominators from Table 10 (total clausal occurrences); the calculation is performed separately for each *Cantica* and for the Comedy as a whole. Formally, for *deprel d* in *Cantica C*:

$$\text{RelFreq}(d,C) = \frac{\text{ellipsis with deprel } d \text{ in } C}{\text{total clausal occurrences of } d \text{ in } C}$$

The data are reported in Table 11 as percentages (100 x RelFreq(*d,C*)). Table 12 gives the difference in percentage points (p.p.) for each *deprel* across the three *Cantiche*, defined as the arithmetic difference between the largest and smallest percentage, thereby quantifying the cross-*Cantica* range (max–min). Image 66 provides a graphical representation of the relative frequency of each *deprel* across the *Cantiche*.

Relative frequency							
<i>Inferno</i>		<i>Purgatorio</i>		<i>Paradiso</i>		<i>Comedy</i>	
<i>deprel</i>	Fr_rel %	<i>deprel</i>	Fr_rel %	<i>deprel</i>	Fr_rel %	<i>deprel</i>	Fr_rel %
advcl:cmp	36,059	advcl:cmp	33,188	advcl:cmp	42,545	advcl:cmp	37,516
conj	13,333	ccomp:reported	14,402	conj	20,661	conj	15,464
root	7,980	conj	13,491	ccomp:reported	11,310	ccomp:reported	10,043
ccomp:reported	5,500	root	6,474	root	4,720	root	6,493
parataxis	4,608	parataxis	5,067	parataxis	3,287	parataxis	4,360
advcl:relcl	4,167	advcl:relcl	2,222	csubj	1,600	advcl:relcl	2,521
ccomp	2,740	csubj	1,466	ccomp	1,587	ccomp	1,931
advcl	2,212	advcl	1,230	advcl	0,751	advcl	1,489
acl:relcl	1,576	ccomp	0,893	acl:relcl	0,406	csubj	1,108
acl	1,277	acl:relcl	0,498	acl	0,000	acl:relcl	0,800
csubj	0,806	acl	0,000	advcl:relcl	0,000	acl	0,435
				csubj:pass	0,000	csubj:pass	0,000

**Table 11.** Relative frequency of predicate ellipsis by dependency relation

Percentage points	
<i>deprel</i>	p.p.
advcl:cmp	9.357
ccomp:reported	8.902
conj	7.328
advcl:relcl	4.167
root	3.260
ccomp	1.847
parataxis	1.780
advcl	1.461
acl:relcl	1.170
acl	1.277
csubj	0.794

**Table 12.** Percentage points for each *deprel* across the three *Cantiche*

Relative frequency of predicate ellipsis by deprel

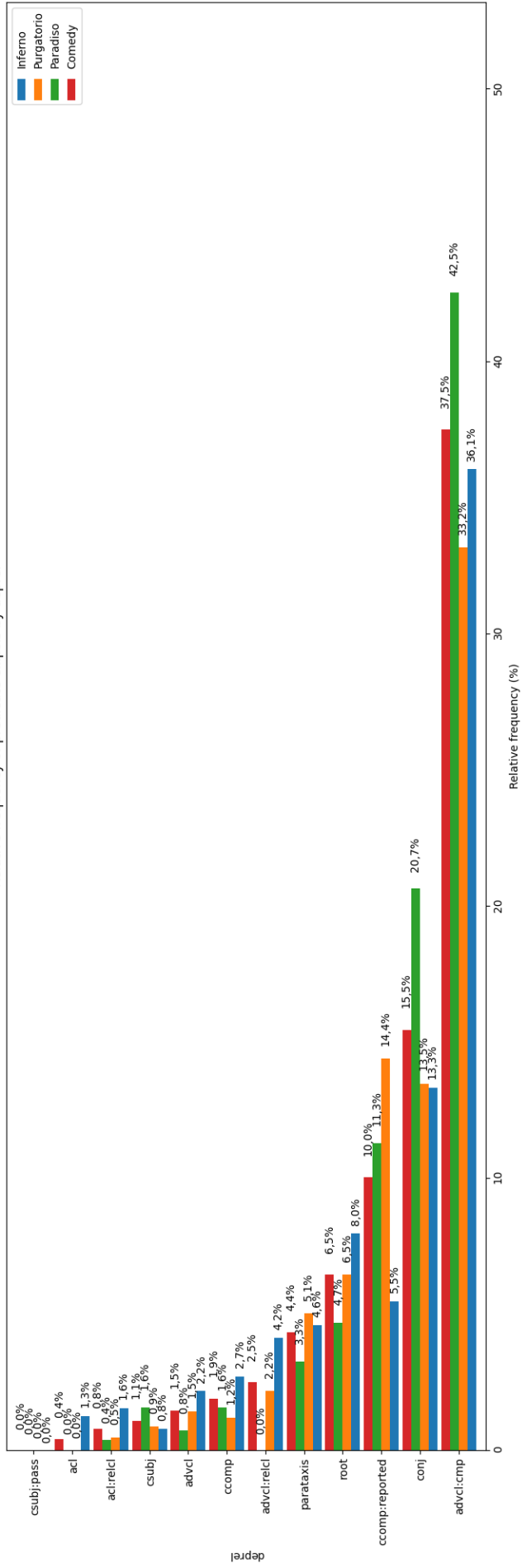


Image 66. Relative frequency by deprel

Considering the ellipsis rate normalized by the total number of clausal occurrences for each dependency relation, the highest proportion is observed in comparative clauses (`advcl : cmp`), which lead in all three *Cantiche*. This is followed by clausal coordination (`conj`) and reported speech (`ccomp : reported`), at approximately 37.5%, 15.4% and 10%, respectively (*Comedy* totals). Next come root position (`root`; ~6.5%) and parataxis (`parataxis`, ~4.3%). All remaining relations form a residual cluster, each contributing under 3%, namely `advcl : relcl`, `ccomp`, `advcl`, `csubj`, `acl : relcl`, and `acl`.

With respect to the cross-*Cantica* range, reported in Table 12 as differences in percentage points, three relations stand out: comparatives, reported speech and clausal coordination. Comparatives (`advcl : cmp`) is the most frequent relation and shows the largest range, 9.357 p.p., scoring from *Paradiso* 42.55% to *Inferno* 36.06% and *Purgatorio* 33.19%. Reported speech (`ccomp : reported`) also show a non-trivial spread, around 8.902 p.p., speaking in *Purgatorio* (14.40%), followed by *Paradiso* (11.31%), and *Inferno* (5.50%). Finally, clausal coordination (`conj`) also varies markedly, with 20.66% in *Paradiso* and lower values in *Purgatorio* (13.49%) and *Inferno* (13.33%), showing a range of 7.328 p.p.. All other relations exhibit cross-*Cantica* ranges below 5 percentage points.

To conclude this Section, the quantitative analysis of the syntactic contexts of ellipsis has shown that ellipses annotated through promotion are more numerous than orphan structures (785 vs. 338). Consequently, an analysis of ellipsis restricted to orphans would lead to underestimation. Both in raw counts (Tables 7-9) and normalized rates (Tables 11 and 12), predicate ellipsis is mainly concentrated in clausal coordination (`conj`), comparative clauses (`advcl : cmp`), and root position (`root`), which together account for about 75% of the total cases.

At the level of the individual *Cantiche*, distinct profiles emerge. *Inferno* exhibits the highest absolute number of ellipses (410); noteworthy is also the high share of `root`-position orphans. *Purgatorio* (366 number of ellipses) shows a strong presence of `ccomp : reported`, which ranks second in relative frequency within this *Cantica* (whereas it ranks third in the other two). *Paradiso* (347 number of ellipses) is characterized by a predominance of `advcl : cmp` and `conj`; with comparative adverbial clauses representing the most frequent relation in this *Cantica*.

### 5.1.3 Classification of predicate ellipsis type

To complement the quantitative analysis presented in Section 5.1.2, a manual inspection of the data was conducted. Among the retrieved instances of predicate ellipsis, several recurring patterns emerged. These were classified and are described in the following subsections, with each subtype specified and exemplified. The clusters are organized as follows: [Section 5.1.3.1](#) gapping; [Section 5.1.3.2](#) elliptical narrative framing of reported speech; Section 5.1.3.3 exclamations; Section 5.1.3.4 foreign languages; Section 5.1.3.5 presentative and locative clauses; Section 5.1.3.6 listing and stylistic repetition; Section 5.1.3.7 truncations and fragments; Section 5.1.3.8 sluicing; Section 5.1.3.9 stripping or bare argument ellipsis; Section 5.1.3.10 ellipsis of the copular *essere* (to be) and Section 5.1.3.11 comparative clauses. Each of these categories is described individually in its respective section.

#### 5.1.3.1 Gapping

One of the largest clusters consists of instances of gapping, a phenomenon involving the omission of the verb in the second conjunct of a coordination (see Section 3.2.1.1). Since gapping “can leave two and only two constituents behind” (Kuno 1976: 318), the `orphan` relation proved to be decisive in extracting such cases, as opposed to promotion cases. This is because `orphan` is marked when two or more phrases survive the ellipsis, thus making it easier to cluster such cases. For this reason, the analysis of gapping in this section is limited to `orphan` structures.

In the *Comedy*, gapping can occur in both forward and backward forms, that is, with the verb appearing before the ellipsis (forward) or after it (backward). This pattern is also attested in other languages, such as Latin (Pinkster 2021 585-586; 610-611)<sup>195</sup> and Ancient Greek (Gaeta & Luraghi 2001), among others. Gapping is also attested in Modern Italian (Marello 2010: 262), with forward gapping appearing to be more pervasive than the backward type, although in-depth studies on the phenomenon remain limited.

As mentioned, both type of gapping, forward and backward, is attested in all the *Cantiche*, as shown in Table 13:

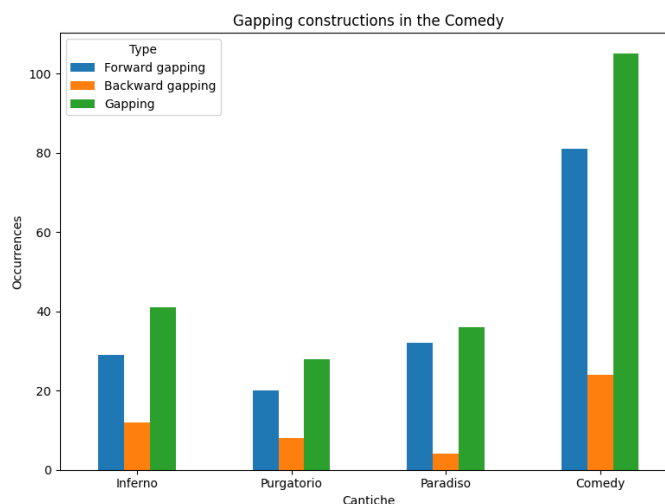
	<i>Inferno</i>	<i>Purgatorio</i>	<i>Paradiso</i>	<i>Comedy</i>
<b>Forward gapping</b>	29	20	32	81
<b>Backward gapping</b>	12	8	4	24
<b>Gapping</b>	41	28	36	105

**Table 13.** Gapping constructions in the *Comedy*.

---

<sup>195</sup> It should be pointed out that Pinkster prefers the terms “reduction” and “conjunction reduction” to “ellipsis” (Pinkster 2021: 721).

As shown in Table 13 and in Image 67, forward gapping is more frequently attested than backward gapping. Out of a total of 105 gapping constructions, only 24 (22.9%) are backward, whereas the remaining 81 occurrences (77.1%) are forward constructions.



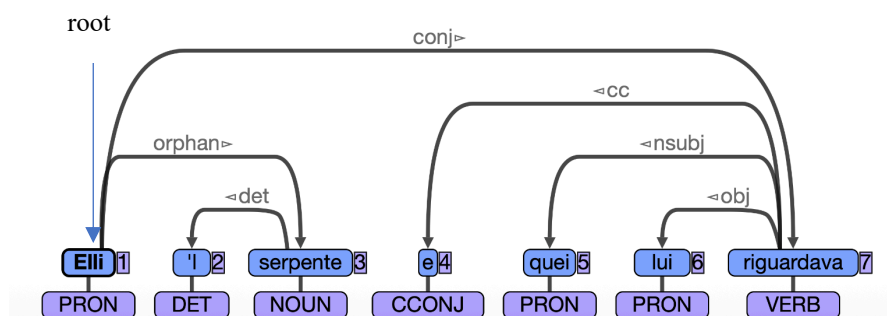
**Image 67.** Gapping constructions in the *Comedy*

Concerning the syntactic position of gapping, forward gapping always occurs in coordinated clauses (see Table 14), but this is not the case for backward gapping, which raises questions about its annotation. According to the UD guidelines, coordination is head-initial (i.e., left-headed)<sup>196</sup>. This implies that the `conj` `deprel` can only go from left to right, never in the opposite direction. This constraint, however, strictly penalizes the annotation of backward gapping, undermining its analysis, since it no longer appears as a coordinated structure, as shown in Example (97) and Image 68.

(97) *Inf.* XXV v. 91

*Elli 'l serpente e quei lui riguardava;*

‘He at the serpent gazed, and it at him;’ (Lit.) He at the serpent, and it at him gazed;



**Image 68.** Head-initial coordination in backward gapping

<sup>196</sup> <https://universaldependencies.org/v2/coordination.html>



Total 81	Total 24
----------	----------

**Table 14.** Syntactic position of gapping

In the following Sections, examples of gapping are reported. More precisely, Section 5.1.3.2.1 discusses forward gapping, section 5.1.3.1.2 backward gapping and Section 5.1.3.1.3 instances of bi-directional and nested gapping.

#### 5.1.3.1.1 Forward Gapping:

Examples (98)-(100) report instances of forward gapping:

(98) *Inf.* XIII, vv. 101-102

*l'Arpie, pascendo poi de le sue foglie, fanno dolore, e al dolor fenestra.*

‘and then the Harpies, feeding on its leaves, cause pain **and for that pain** (provide) a vent.’

Example (98) is taken from *Inferno* XIII, in the forest of suicides. The verses describe part of the punishment in this circle the damned souls of suicides are transformed into plants, and the Harpies, mythological monsters, feed on them by tearing off their leaves (*pascendo poi de le sue foglie* ‘feeding on its leaves’ v. 101). In so doing, they cause pain to the damned soul (*fanno dolore* ‘cause pain’ v. 102) and create an opening through which the pain can escape (*e al dolor fenestra* ‘and for that pain a vent’ v. 102). In (98), the clause undergoing gapping is shown in bold (*e al dolor fenestra* ‘and for that pain (provide) a vent.’), whose verbal antecedent occurs in the previous clause and corresponds to the verb *fanno* ‘cause’ (lit. ‘they make’), thus representing an instance of forward gapping.

(99) *Purg.* XXVII, vv. 112-113

*le tenebre fuggian da tutti lati, e 'l sonno mio con esse*

‘The darkness fled away on every side, **and slumber with it**’

In Example (99), an extract from *Purgatorio* XXVII, the dawn is described as the moment when the darkness of night disperses (*le tenebre fuggian da tutti lati*, ‘the darkness fled away on every side,’ v. 112), and the poet’s sleep fades, vanishing together with the shadows of the night and awakening him (*e il sonno mio con esse*, ‘and my slumber with it,’ v. 113). The verbal antecedent of this instance of forward gapping is the verb *fuggian* ‘fled away’, which is omitted in the gapped clause.

(100) *Par.* XIV, vv. 40-41

*La sua chiarezza séguita l' ardore; l' ardor la visione*

‘Its brightness takes its measure from our ardor, **our ardor from our vision**’

Finally, Example (100) is taken from *Paradiso* XIV, and is part of the description of the radiance of the blessed souls. As Dante is told, the blessed souls will shine for eternity because their brightness is caused by love and ardor (*La sua chiarezza séguita l' ardore* ‘Its brightness takes its measure from our ardor’ v. 40), and, in turn, this ardor is caused by the degree of their vision of God (*l' ardor la visione* ‘our ardor from our vision’ v. 41). The verbal antecedent in this instance is the verb in the main clause, *séguita* ‘takes its measure’.

It is interesting to note that, in Example (100), a syntactic parallelism is maintained between the non-elided sentence containing the antecedent and the gapped one, namely, by preserving a similar structure in both, namely, SVO and S(V)O (*la chiarezza* (S) – *séguita* (V) - *l' ardore* (O); *l' ardor* (S) - \_ - *la visione* (O)).

By contrast, the other two examples, (98) and (99), differ in their syntactic patterns between the non-elided sentence and the gapped one. More specifically, in Example (99), the two phrases in the non-elided clause are the subject (*le tenebre* ‘the darkness’) and an oblique referring to a location (*da tutti lati* ‘on every side’), whereas the two remnants in the elided sentence are the subject (*il sonno mio* ‘slumber’) and an oblique (*con esse* ‘with it’), which conveys companionship rather than location.

Similarly, in Example (98), the parallelism is broken: to a pattern SVO (*l'arpie-fanno-dolore* ‘the Harpies – cause-pain’) in the non-elided sentence corresponds an (S)Obl(V)O pattern (*l'arpie*) - *al dolore* - (*fanno*) – *fenestra* ‘(the Harpies) – for that pain – (provide) – a vent’). Specifically, in the gapped sentence, the subject and the verb are omitted, and what was the object in the non-elided sentence (*dolore* ‘pain’) is here repeated but encoded as an oblique with the preposition *a* ‘to’. Even if the syntactic structure does not replicate the exact pattern of the non-elided sentence, from a lexical–semantic point of view it is worth noting that the expression *fare finestra a qualcuno/qualcosa* ‘to provide a vent for someone/something’ can be considered a multiword expression, thereby compacting the semantic content into a single unit, even though its syntactic realization remain different.

Another interesting point is that, due to the specific poetic genre of the text, forward gapping occurs in constructions that also display rhetorical figures, such as anadiplosis (Mortara Garavelli 1997: 191), as in Examples (100) and (98). While the scheme in (100) is a classical instance of anadiplosis, with the repeated element appearing in the final and initial position in the two segments (...X; X...), Example (98) represents a specific variant in which the preposition *a* ‘to’ is inserted to encode the new syntactic role of the repeated element. The use of anadiplosis contributes to enhancing the

evocative quality of the verse, recalling, in addition to the symmetry created by the ellipsis (which implies the reconstruction of the same verb in both clauses), a repetition that falls on a remnant (*ardor* in (100) and *dolor* in (98)) and that, in one case (Example (98)), also entails a syntactic shift that further challenges the reader.

Concerning the syntactic position of gapping, it is important to note that while the gapped clause always occurs in a coordinated clause (necessarily coordinated with the clause containing the antecedent), the syntactic position of the clause including the antecedent may vary.

Examples (98)-(100) display instances of forward gapping in which the clause containing the antecedent is the main clause. However, the clause including the antecedent can also be an embedded clause, such as a causal clause in Example (101), a consecutive clause, as in (102), or a coordinated clause, as in (103):

(101) *Inf.* XXV, vv. 4-7

*perch'una li s'avvolse allora al collo (...); e un'altra a le braccia*

‘For one entwined itself about his neck (...); **And** (round) **his arms another**’

Example (101) is taken from Canto XXV of *Inferno*, which describes how, as punishment for an obscene gesture made by a damned soul, Vanni Fucci, two serpents coil themselves around the culprit, one entwining his neck and the other his arms. The verbal antecedent *s'avvolse* ‘entwined itself’ occurs in a clause functioning as a causal clause.<sup>197</sup>

(102) *Par.* XXIV, vv. 14-15

*si giran sì, che 'l primo a chi pon mente quieto pare, e l' ultimo che voli*

‘Revolve so that the first to the beholder Motionless seems, and **the last one to fly**’

In Example (102), the antecedent *pare* ‘seems’ appears in a consecutive clause. The verses describe the brightness of the circles of the blessed, whose lights shine at different speeds, as in the transmission mechanisms of ancient clocks (Leonardi Chiavacci 2012: 431). For this reason, the first circle appears slower to the observer (*'l primo a chi pon mente quieto pare* ‘the first to the beholder Motionless seems’), while the last seems to fly because of its speed (*e l' ultimo che voli* ‘and the last one to fly’).

(103) *Inf.* XXV, vv. 109-111

---

<sup>197</sup> For the classification of subordinate clauses, see DanteSearch (Section 2.1.2.1) and Gigli (2004).

*Togliea la coda fessa la figura che si perdeva là, e la sua pelle si faceva molle, e quella di là dura.*

‘He with the cloven tail assumed the figure, The other one was losing, and his skin Became elastic, and the other’s hard.’

Finally, from the same *Canto* as Example (101), Example (103) represents an instance of forward gapping in which the antecedent appears in a coordinate clause. The verses describe the metamorphosis of a damned soul into a serpent and, conversely, the transformation of the serpent into the damned soul, as in an exchange of bodies. The serpent’s skin is depicted as becoming soft (*e la sua pelle si faceva molle* ‘and his skin became elastic’), while, conversely, that of the man becomes hard (*e quella di là dura* ‘and the other’s hard’), resembling the serpent’s skin.

Finally, forward gapping has been shown to involve more than one gapping structure at time, as in Example (104), where there are four ellipsis sites in total (marked here with an underscore).

(104) *Par. XXVIII, vv. 28-30*

*e questo era d' un altro circumcinto, e quel \_ dal terzo, e 'l terzo poi \_ dal quarto, dal quinto\_ il quarto, e poi dal sesto \_ il quinto.*

‘And this was by another circumcinct, **That by a third, the third then by a fourth, By a fifth the fourth, and then by a sixth the fifth**’

Example (104) describes the light of God (vv. 16-17 *lume acuto* ‘light so acute’) around which the angelic intelligences revolve. The description of the latter unfolds in these verses in a pressing and rapid manner, as if to simulate the movement of the eye from one fire to another, each interlinked with the next (Leonardi Chiavacci 2012: 505). The verbal antecedent *era circumcinto* ‘was circumcinct’ is followed by four verbal ellipses, which depict the various concentric circles with a movement outward from the focal point. Interestingly, in (104) the antecedent is not contiguous; rather, it is interrupted by a phrase (*d'un altro* ‘by another’), thus exhibiting hyperbaton<sup>198</sup> and showing that an antecedent can also be non-contiguous.

#### 5.1.3.1.2 Backward Gapping

Moving to another type of gapping, namely cases in which the antecedent is in fact a postcedent, that is, it occurs after the ellipsis site, Examples (105)-(108) illustrate instances of backward gapping:

---

<sup>198</sup> Hyperbaton is a rhetorical and syntactic figure involving a marked deviation from the canonical word order of a sentence, whereby elements that are normally adjacent are separated by intervening material.

(105) *Inf. XXV, v. 91*

***Elli 'l serpente e quei lui riguardava***

He at the serpent (gazed), and it at him;

Example (105) is again taken from the *Inferno XXV*, a Canto in which the description is highly dynamic, as also evidenced by the quantity of ellipses in the narration of the scenes (see also Examples (101) and (103)). The verse describes the exchange of glances between a serpent and the damned soul, who are about to undergo the metamorphosis described in Example (103). The elided verb in *elli 'l serpente* 'he at the serpent' is *riguardava* 'gazed', which is a postcedent.

(106) *Purg. XXVII, v. 108*

***lei lo vedere, e me l' ovrare appaga***

'**Her, seeing, and me, doing satisfies.**'

Also in Example (106), the elided verb has a postcedent, namely the verb *appaga* 'satisfies'. The verse is taken from the words of Leah, a biblical figure, who describes herself as fulfilled by action, while depicting her sister Rachel as fulfilled by contemplation.

(107) *Par. II, v. 22*

***Beatrice in suso, e io in lei guardava***

'**Beatrice (gazed) upward. I watched her.**'

Finally, Example (107) describes the scene of the ascent to the Heaven of the Moon. The transition from one Heaven to another is conveyed through the device of the gaze: while Dante looks into Beatrice's eyes and she looks at the Heaven, the shift takes place. As in (105) and (106), also in (107) there is a postcedent, namely, *guardava* 'watched'.

In all Examples (105)-(107), the remnants are parallel to their corresponding phrases in the non-elided clause, thereby maintaining a parallelism that is also reflected in their syntactic function. More precisely, the syntactic scheme is SO-SOV in (105); OS-OSV in (106), and SObl-SObIV in (107). However, although parallelism seem to be more often preserved in backward gapping structures than in forward gapping (Gaeta & Luraghi 2001: 104), as illustrated in Examples (105)-(107), this is not always the case, as shown in Example (108):

(108) *Inf.* VIII, vv. 118-119

*Li occhi a la terra e le ciglia avea rase d'ogne baldanza*

‘His eyes (turned) to the ground, his brows deprived of every confidence’

In Example (108), the syntactic parallelism found in (105)-(107) is absent. Even though one remnant maintains parallelism with its counterpart in the non-elided clause (namely, *li occhi* ‘his eyes’ and *le ciglia* ‘his brows’, both referring to a part of the body described and both functioning as syntactic objects in their respective clause), the second remnant, *a la terra* ‘to the ground’ and its counterpart *rase d'ogne baldanza* ‘deprived of every confidence’, does not display a syntactically parallel pattern. More precisely, the scene described in these verses depicts Virgil’s bitterness after being refused entry to the city of Dis, a disappointment conveyed through his downcast eyes and a gaze devoid of boldness. Although both clauses constitute a single descriptive unit that physically depicting Virgil’s expression of bitterness, the syntactic parallelism is not preserved.

Finally, as for forward gapping, also backward gapping can occur both in the main clause, as shown in (105), (107), and (108), but also in embedded clauses of different types, such as the exceptive clause in (109), and the relative clause in (110) and (111):

(109) *Inf.* XIX, vv. 113-114

*e che altro è da voi a l'idolatre, se non ch'elli uno, e voi ne orate cento?*

‘how are you different from idolaters, **save that they** (worship) **one** and you a hundred?’

Example (109) presents an instance of backward gapping in an exceptive clause *se non ch'elli uno* ‘save that they (worship) one’, where the verbal postcedent is *orate* ‘worship’. The context is that of an accusation of idolatry toward coins and money, compared to the biblical worship of the golden calf.

Similarly to (109), also in (110) and (111), backward gapping occurs in embedded clauses, here in relative ones:

(110) *Purg.* VII, vv. 98-99

*resse la terra dove l'acqua nasce **che Molta in Albia**, e Albia in mar ne porta*

‘Governed the region where the water springs, **The Moldau** (bears) **the Elbe**, and Elbe the sea.’

(111) *Para* XXXI, vv. 37-38

*io, che al divino da l' umano, a l' eterno dal tempo era venuto.*

'I who to the divine (had) from the human, From time unto eternity, had come'

In (110) the verses describe Ottokar II, King of Bohemia, portrayed as the one who ruled the lands where the river (*l'acqua* 'the water') that the Moldau bears into the Elbe and that the Elbe bears into the sea rises; here, the postcedent is the verb *porta* 'bears'. Similarly, (111) presents a relative clause with a postcedent, *era venuto* 'had come'. In these verses, the poet marvels at having passed from the human to the divine realm (*al divino da l'umano* 'to the divine from the human'), that is, from the temporal world to the eternal one (*a l'eterno dal tempo* 'from time unto eternity').

To conclude, unlike forward gapping in (97), no instances of multiple backward gapping are attested.

#### 5.1.3.1.3 Bi-directional gapping and nested gapping

Among the instances of gapping in the *Comedy*, there are cases of so-called bi-directional gapping (Gaeta & Luraghi 2001:107), that is, instances in which both forward and backward gapping are simultaneously attested, as in (112), where the backward gapping is marked in bold, whereas the forward gapping are underlined.

(112) *Par. XI, vv. 4-7*

***Chi dietro a iura e chi ad amforismi sen giva, e chi seguendo sacerdozio, e chi regnar per forza o per sofismi, e chi rubare e chi civil negozio***

**'One after laws and one to aphorisms Was going, and one following the priesthood, And one to reign by force or sophistry, And one in theft, and one in state affairs'**

Example (112) expresses the poet's commiseration for mankind's striving after earthly goods. The long list is structured through various bi-directional ellipses: following a backward gapping *chi \_ dietro a iura* 'one after laws', whose postcedent is *sen giva* 'was going', come several instances of forward gapping, where the antecedent is always *sen giva* 'was going': *e chi \_ seguendo sacerdozio, e chi \_ regnar per forza o per sofismi, e chi \_ rubare e chi \_ civil negozio* 'and one following the priesthood, and one to reign by force or sophistry, and one in theft, and one in state affairs'.

Moreover, Example (112) not only presents bi-directional gapping, but also an instance of nested gapping: the initially elided material involves only the verb *sen giva* 'was going', which is missing in all the gapped structures; at a later point, however, a further verb is elided, namely, *seguendo* 'following', which constitutes an instance of forward gapping within forward gapping itself, thus displaying two elided verbs: *e chi \_ [G1] seguendo sacerdozio, e chi \_ [G1] \_ [G2] regnar per forza o per*

*sofismi, e chi* <sub>[G1]</sub> <sub>[G2]</sub> *rubare e chi* <sub>[G1]</sub> <sub>[G2]</sub> *civil negozio* ‘and one <sub>[G1]</sub> following the priesthood, and one <sub>[G1]</sub> <sub>[G2]</sub> to reign by force or sophistry, and one <sub>[G1]</sub> <sub>[G2]</sub> in theft, and one <sub>[G1]</sub> <sub>[G2]</sub> in state affairs’.

Finally, as already discussed in Sections 5.1.3.2.1.1 and 5.1.3.2.1.2, it is worth noting that, while backward gapping shows less freedom in pattern variation, here displaying perfect parallelism with a S-Obl (*chi – dietro a iura* ‘one – after laws’) and S-Obl-V pattern (*chi – ad amforismi – sen giva* ‘one – to aphorisms – was going’), forward gapping is more open to modifying the parallelism, as in this example, which features the following syntactic patterns: S-Obl-V (*chi – ad amforismi – sen giva* ‘one – to aphorisms – was going’); S-Advcl (*chi – seguendo sacerdozio* ‘one – following the priesthood’); S-O-Obl (*chi regnar per forza o per sofismi* ‘and one to reign by force or sophistry’); S-O (*chi rubare* ‘one in theft’); S-O (*chi civil negozio* ‘one in state affairs’).

### 5.1.3.2 Elliptical narrative framing of reported speech

One of the emerging clusters has been defined as elliptical narrative framing of reported speech, as the omitted verbs, consistently belonging to the semantic domain of *verba dicendi* (such as ‘to say’, ‘to reply’, and similar), serve to introduce reported speech. This type of predicate ellipsis, frequently attested across all three *Cantiche*, appears to function as a narrative device aimed at enhancing the fluency and vividness of the dialogue between the characters, thereby simulating rapid conversational exchanges. It occurs both with orphan and promotion annotation system. Examples (113)-(115) reports instances of ellipsis in narrative framing of reported speech for each *Cantica*:

(113) *Inf.* III, vv. 10-12

*Queste parole di colore oscuro vid'io scritte al sommo d'una porta; per ch'io: «Maestro, il senso lor m'è duro».*

‘These words—their aspect was obscure—I read inscribed above a gateway, **and I** (said): “Master, their meaning is difficult for me.”’

Example (113) describes the view from the gates of Hell, whose inscription (*queste parole di colore oscuro* ‘these words—their aspect was obscure’ v. 10) proves difficult for Dante to understand (*il senso lor m'è duro* ‘their meaning is difficult for me’ v. 12). He turns to his guide to ask for clarification. The introduction to Dante’s direct speech is structured by the poet through verbal ellipsis: the *verbum dicendi* is omitted, leaving only the subject/speaker *io* ‘I’ explicitly expressed. Notably, the absence of an explicit addressee is compensated for by the vocative (*Maestro* ‘Master’) at the beginning of the reported speech.

(114) *Purg.* III, vv. 94-100

«*Sanza vostra domanda io vi confesso che questo è corpo uman che voi vedete; (...)».* **Così**  
**'l maestro;**

“Without your asking, I shall tell you plainly that you are looking at a human body; (...)”.  
These (were) my master’s words.

Example (114) illustrates another instance of ellipsis of the verb introducing reported speech. In this case, the narrative hinge follows the reported speech (refer to [Section 5.2.3.1](#) for an analysis of the use of the deictic *così* ‘so’) and once again omits the *verbum dicendi*. The words spoken by Virgil concerning Dante’s human nature (*questo è corpo uman che voi vedete* ‘you are looking at a human body’ v. 95) reach the reader before the speaker is identified. The narrator does not make it necessary to specify the verb ‘to say’, thus leaving it unexpressed, especially given that a second utterance follows only a few verses later, this time explicitly introduced (vv. 100–101: *e quella gente degna «Tornate», disse*, ‘That worthy band **replied**: “Come back’’).

(115) *Par.* XXVI, vv. 82-84

**E la mia donna:** «*Dentro da quei rai vagheggia il suo fattor l’anima prima che la prima virtù creasse mai*».

**My lady** (answered): “In those rays there gazes with love for his Creator the first soul ever created by the Primal Force.”

Example (115) once again illustrates the omission of the verb to say. In this specific context, no *verbum dicendi* appears in close proximity to the verses in question. Rather, the contextualization of speech is triggered by the preceding narrative passage, in which the author recounts his question regarding a light (vv. 80-81: *domandai d’un quarto lume ch’io vidi tra noi*. ‘I asked about the fourth light that I saw among us.’). Beatrice’s response follows immediately (Example 115, vv. 82-84) and, as the reader can infer from both the direct speech and the surrounding context, functions as a natural continuation of the narrative path.

As evidenced by Examples (113)-(115), the patterns in which this narrative strategy is employed vary and contribute to the dynamism of the text. The omitted verb may precede (as in Examples (113) and (114)) or follow (as in Example (115)) the reported speech. It may be preceded or followed by an explicit *verbum dicendi* (as in 114) or may simply emerge from the narrative flow without any overt

marker (as in 115). Moreover, this mechanism is not confined to a single turn of speech but can be iterated across multiple turns, making the interaction even more dynamic and engaging, as shown in *Inf. X*, as reported in Example (116):

(116) *Inf. X*, vv. 4-21

«*O virtù somma (...)*», **cominciai**, «*(...)*». ***E quelli a me***: «*Tutti (...) tu mi taci*». ***E io***: «*Buon duca (...)*».

“O power supreme (...)”, I **began**, “(...)”. **And he to me**: “They all (...) thou keepest silent.”

**And I**: “Good Leader, (...)”.

This exchange is part of an extended dialogue between Dante and his guide Virgil as they venture among the open tombs of the heretics condemned in this circle. The narrative framing of their dialogue progressively diminishes: an initial utterance is introduced with a first-person verb announcing Dante’s speech (*cominciai* ‘I began’), followed by a brief transitional description marking Virgil’s reply (*E quelli a me* ‘And he to me’). At this point, the narrator omits the verb of saying, reducing the narration to the mere identification of the speaker (*e quelli* ‘and he’) and the addressee (*a me* ‘to me’). Following Virgil’s speech, Dante’s response is rendered even more elliptically than the previous turn, being introduced simply by *e io* ‘and I’. By this stage, the reader is fully attuned to the rhythm of the exchange. This progressive reduction in narrative mediation serves to accelerate the interaction and preserve the immediacy and liveliness of the dialogue.

### 5.1.3.3 Exclamation

Among the extracted clauses lacking a verb, one identifiable cluster consists of exclamations. The term exclamation is used here as synonym of interjection, both terms also attested in Serianni (Serianni, 1991: 367). However, the term exclamation is adopted rather than interjection, as the latter refers to a specific part of speech used in UD<sup>199</sup>. To avoid potential overlap or ambiguity with UD’s POS label, exclamation is preferred throughout this analysis. As for the elliptical narrative framing of reported speech, also this cluster is attested in all the three *Cantiche*.

Exclamations include a wide range of clauses displaying different purposes. The most common type of exclamations are sentences that are characterized by the presence of expressions conveying various types of emotional reactions, such as *Ah!*, *Oimè!*, and similar forms, all marked with the upos tag INTJ (interjection). Examples of this type of exclamations are reported in Examples (117)-(7):

---

<sup>199</sup> Refer to: <https://universaldependencies.org/u/pos/INTJ.html>

(117) *Inf.* XII, vv. 49-51

*Oh cieca cupidigia e ira folle, che sì ci sproni ne la vita corta, e ne l'eterna poi sì mal c'immolle!*

'O blind cupidity and insane anger, which goad us on so much in our short life, then steep us in such grief eternally!'

(118) *Purg.* II, v. 79

*Ohi ombre vane, fuor che ne l' aspetto!*

'O shades—in all except appearance—empty!'

(119) *Par.* XXIII, v. 34

*Oh Beatrice, dolce guida e cara!*

'O Beatrice, sweet guide and dear!'

All the instances in Examples (117)-(119) begin with an interjection (*Oh, Ohi, Oh*) and feature a nominal root, *cupidigia* 'cupidity', *ombre* 'shades' and *Beatrice*, respectively.

In addition to clauses introduced by interjections composed of vowel sounds, the most frequently attested type of exclamation, the corpus also includes other forms of exclamatory expressions. These comprise phrases drawn from ecclesiastical psalms or prayers (Examples (120) and (121)), which are particularly frequent in *Paradiso* and *Purgatorio*; exhortations (Example (122)); incipits of songs or sonnets (Example 123); isolated exclamations uttered by characters (Examples (124) and (125)); and admonitions (Example 126).

(120) *Par.* XXVI, vv. 68-69

*e la mia donna dicea con li altri: «Santo, santo, santo!».*

'and my lady said with the others: "Holy, holy, holy!"'

(121) *Par.* III, vv. 121-122

*Così parlammi, e poi cominciò 'Ave, Maria' cantando*

This said, she then began to sing "Ave Maria"

(122) *Purg.* VIII, v. 65

*gridando: «Sù, Currado!*

called (...): "Up, Currado!"

(123) *Purg.* XXIV, vv. 50-51  
*cominciando* «**Donne ch' avete intelletto d' amore**»  
'beginning: 'Ladies who have intelligence of love.'

(124) *Purg.* XXVI v. 40  
*la nova gente:* «**Soddoma e Gomorra**»  
'the new group shouting: "Sodom and Gomorrah"'

(125) *Purg.* XXIV v. 37  
*El mormorava; e non so che* «**Gentucca**»  
'He murmured; something like "Gentucca" was'

(126) *Inf.* III, v. 84  
*gridando:* «**Guai a voi, anime prave!**»  
'was shouting: "Woe to you, corrupted souls!'

All the Examples reported in (117)-(126) illustrate clauses in which the predicate is not only absent, but also syntactically unnecessary. I will come back to this point in the conclusion section.

#### 5.1.3.4 Foreign language

Other instances retrieved by the query include clauses in foreign languages. Since words in foreign languages are not morpho-syntactically analyzed<sup>200</sup>, they are annotated with the universal part-of-speech tag X<sup>201</sup>. As a result, these clauses were retrieved as instances of verb-less clauses. However, since their retrieval is not motivated by linguistic analysis but rather results from a practical limitation in annotation, this cluster will be only briefly addressed and will not be treated as genuine instances of predicate ellipsis.

Should future work involve the recruitment of annotators specialized in the specific foreign languages present in the *Comedy*, such as Latin, additional cases may be subject to proper analysis. Nonetheless, since the foreign languages also include instances of fictional and invented languages, these cases will, by necessity, remain unannotated. For the sake of clarification, examples of clauses in foreign

---

<sup>200</sup> An exception is represented by the sonnet in Provençal by Arnaut Daniel (*Purg.* XXVI, vv. 140-147), which, due to its length, has been annotated by a specialist in the language.

<sup>201</sup> Refer to: <https://universaldependencies.org/u/pos/X.html>

languages are provided, both in a natural language, Latin (Example 127), and in a fictional one (Example 128):

(127) *Par.* XIII, v. 100

*non si est dare primum motum esse* (lat.)

(This sentence was left untranslated by the translators.)

(128) *Inf.* VII, vv. 1-2

«*Pape Satàn, pape Satàn aleppe!*», cominciò Pluto con la voce chioccia

“Pape Satan, Pape Satan aleppe!” so Plutus, with his grating voice, began.’

It is worth noting that, at times, the exclamation and foreign language categories converge, as certain psalms and prayers are reproduced in the original language (Latin), as shown in Example (129):

(129) *Purg.* II, v. 46

«*In exitu Israel de Aegypto*» cantavan tutti insieme

“In exitu Israel de Aegypto!” They chanted all together in one voice’

#### 5.1.3.5 Presentative and locative clauses

This cluster includes presentative and locative clauses, namely, utterances whose primary function is to introduce or announce an event, or to identify and specify a position or location. What these two types share is that both are headed by an adverb, which serves as the syntactic root of the clause. Included in this cluster are, for instance, clauses introduced by the adverb *ecco* ‘there’, which is classified as a presentative adverb (Serianni 1991: 509), as well as *quivi* ‘here’, adverb used to indicate location (Serianni 1991: 501). Examples (130)-(135) illustrate instances of presentative and locative clauses without verbs. While presentative clauses lacking a verb are attested across all three *Cantiche* (Examples (130)-(132)), bare locative constructions without a verb are less frequent: there is only one instance in *Inferno* (Example 133), with no attestations of this kind found in *Purgatorio* and *Paradiso*.

(130) *Inf.* XIII, vv. 115-116

*Ed ecco due da la sinistra costa, nudi e graffiati*

‘And **there** upon the left were two who, scratched and naked’

(131) *Purg.* II v. 29

*Ecco l' angel di Dio*

‘behold the angel of God’

(132) *Par.* XXIII, vv. 19-20

*E Beatrice disse: «Ecco le schiere del triunfo di Cristo*

‘And Beatrice said: “There you see the troops of the triumphant Christ’

(133) *Inf.* V, vv. 34-35

*Quando giungon davanti a la ruina, **quivi** le strida, il compianto, il lamento;*

‘When they come up against the ruined slope, then **there** are cries and wailing and lament’

On the specificity of the adverb *ecco* ‘there’ and its sharing properties with both interjections and verbs, see De Cesare (2019: 56-57).

It is worth noting that *ecco* most frequently occurs without a verbal head, contributing to a high proportion of non-verbal clauses in which this adverb functions independently. Indeed, out of 31 occurrences of the lemma *ecco*, only 14 are annotated as depending on a verbal head, while 17 appear without one and therefore are included in this type of presentative clause.<sup>202</sup>

By contrast, *quivi* ‘there/here’ is more likely to co-occur with an overt verb: *quivi* appears 94 times. Of these, 84 instances of *quivi* are annotated as depending on a verbal head. The remaining occurrences, where the adverb does not depend on a verb, correspond to cases in which the adverb co-occurs with the verb *essere* ‘to be’, which is annotated as a copula and therefore does not function as a syntactic head.<sup>203</sup>

Here, as already mentioned in Exclamation (5.1.3.2.2), no predicate could be implied and reconstructed.

#### 5.1.3.6 Listing and stylistic repetition

---

<sup>202</sup> The following GrewMatch queries were used to investigate the syntactic behavior of the lemma *ecco*:

(1) To retrieve all occurrences of the lemma *ecco*: pattern { X [lemma="ecco"] };

(2) To retrieve occurrences where *ecco* depends on a verbal head: pattern { X [lemma="ecco"] }, pattern { Y -> X }, pattern { Y [upos="VERB"] };

(3) To retrieve occurrences where *ecco* does not depend on a verbal head: pattern { X [lemma="ecco"] }, pattern { Y -> X }, without { Y [upos="VERB"] }.

Similar queries were used to investigate also *quivi* and *quinci*.

<sup>203</sup> Only one occurrence does not include the verb *essere* and corresponds to a correlational pattern.

Another cluster that emerged from the analysis consists of cases of listing and stylistic repetition, as illustrated in Examples (134)-(136). Listing (Examples (134) and (135)) involves the enumeration and presentation of a sequence of elements, typically without the presence of a verb.

(134) *Purg.* XVI, vv. 121-126

*Ben v' èn tre vecchi ancora in cui rampogna l' antica età la nova, e par lor tardo che Dio a miglior vita li ripogna: **Currado da Palazzo e 'l buon Gherardo e Guido da Castel**, che mei si noma, francescamente, il semplice Lombardo.*

‘True, three old men are there, in whom old times reprove the new; and they find God is slow in summoning them to a better life: **Currado da Palazzo, good Gherardo, and Guido da Castel**, whom it is better to call, as do the French, the candid Lombard.’

Example (134) presents a listing of individuals. The verse is taken from a broader reproach concerning the Church’s desire to retain both temporal and spiritual power (Leonardi Chiavacchi 2012: 291). The disordered state resulting from this is exemplified in Lombardy, a region where only three Elders are said to have remained virtuous. Following the introduction of these three Elders and a description of their qualities, their names are listed without the use of a verb.

Example (135) also presents an instance of listing, though in a more structured form, with multiple elements that contribute to marking the rhythm of the narration.

(135) *Inf.* XXII, vv. 4-12

*Io vidi già cavalier muover campo, e cominciare stormo e far lor mostra, e talvolta partir per loro scampo; corridor vidi per la terra vostra, o Aretini, e vidi gir gualdane, fedir torneamenti e correr giostra; **quando con trombe, e quando con campane, con tamburi e con cenni di castella, e con cose nostrali e con istrane**; né già con sì diversa cennamella cavalier vidi muover né pedoni, né nave a segno di terra o di stella.*

‘Before this I’ve seen horsemen start to march and open the assault and muster ranks and seen them, too, at times beat their retreat; and on your land, o Aretines, I’ve seen rangers and raiding parties galloping, the clash of tournaments, the rush of jousts, **now (done) with trumpets, now with bells, and now with drums, and now with signs from castle walls, with native things and with imported ware**; but never yet have I seen horsemen or seen infantry or ship that sails by signal of land or star move to so strange a bugle!’

The verses in Example (136) are the opening verses of *Inferno Canto XXII*, which link back to the conclusion of the previous *canto*, where a devil gives the signal to set the group escorting the two poets in motion (*Inf. XXI*, v.139 *ed elli avea del cul fatto trombetta* ‘And he had made a trumpet of his ass.’).

In these opening verses, the author, ironically, compares the devil’s gesture to real-life experiences. The lengthy incipit (vv. 1–9) evokes scenes of war and peace, introducing the reader to a military world through brief yet vivid descriptions capable of conjuring expansive imagery. However, the informative cue of this extended period (spanning 12 verses) appears only in vv. 10–12, where the initial affirmative construction (v. 1 *vidi* ‘I saw’), used to describe familiar scenes, is overturned by a negation (vv. 10-11 *né già (...) vidi* ‘nor did I ever see’). This marks a shift from reality to unreality, thereby emphasizing the exceptional nature of the devil’s gesture. In the extended depiction of scenes from military life, certain episodes are introduced elliptically, both to avoid redundancy, already underscored through the repeated use of *vidi* (occurring 4 times in the period), and to streamline the narration while preserving the brisk rhythm of the verse. Syntactically, the absence of the verb in verses 7-9 results in clauses that are elliptical of the predicate and presents a list of conventional signals, which will later be contrasted, highlighting its parodic effect, with the devil’s signal.

Finally, Example (137) reports an instance of stylistic repetition:

(136) *Par. XXX*, vv. 38-42

*Noi siamo usciti fore del maggior corpo al ciel ch' è pura luce: luce intellettual, piena d' amore; amor di vero ben, pien di letizia; letizia che trascende ogni dolzore.*

‘From matter’s largest sphere, we now have reached the heaven of pure light, **light of the intellect, light filled with love, love of true good, love filled with happiness, a happiness surpassing every sweetness.**’

In this passage, part of Beatrice’s reported speech describing the Empyrean, the author employs the rhetorical figure of anadiplosis, that is, the repetition of the final word of one segment at the beginning of the next (Mortara Garavelli 1997: 191). In this case, the repetition occurs three times: *luce: luce*, *amore: amor*, and *letizia: letizia*. These repetitions occur between syntactically isolated segments that lack an explicit verb.

#### 5.1.3.7 Truncations and fragments

This cluster includes truncation, that is, sentences explicitly left incomplete through the use of ellipsis marks (namely, "..."), as well as fragments, that is parts of sentences, such as short answers to questions or utterances lacking an overt antecedent (as defined in [Section 3.2.3.2](#)).

Truncations, for which examples are reported in (137) and (138), are not attested in *Paradiso*.

(137) *Inf.* IX vv. 7-8

*«Pur a noi converrà vincer la punga», cominciò el, «se non... Tal ne s'offerse.*

“We have to win this battle,” he began, “**if not**. . . But one so great had offered help.’

(138) *Purg.* VII, v. 12

*dicendo, «Ella è... non è... »*

‘saying, “**It is...is not**,”’

Fragments have been retrieved both as responses to questions, whether negative (Examples 139 and 140) or affirmative (Example 141), as well as in replies that are not strictly polar, as in Example (142):

(139) *Purg.* XVI, vv. 97-99

*Le leggi son, ma chi pon mano ad esse? Nullo, però che 'l pastor che procede, rugumar può*

‘The laws exist, but who applies them now? No one—the shepherd who precedes his flock can chew the cud’

(140) *Par.* VIII, vv. 118-120

*«E puot' elli esser, se giù non si vive diversamente per diversi offici? Non, se 'l maestro vostro ben vi scrive».*

‘Can there be citizens if men below are not diverse, with diverse duties? **No**, if what your master writes is accurate.’

(141) *Par.* VIII, v. 117

*«Sì», rispuos' io; «e qui ragion non cheggio».*

‘**Yes**, I replied, and here I need no proof.’

(142) *Inf.* XXVIII, v. 109

*E io li aggiunsi: «E morte di tua schiatta»*

‘I added: “—**and** (brought) **death to your own kinsmen**” ’

#### 5.1.3.8 *Sluicing*

Among predicate ellipses, instances of the so-called sluicing have also emerged (refer to Section 3.2.3.3).

(143) *Par. XX, vv. 88-89*

*«Io veggio che tu credi queste cose perch' io le dico, ma non vedi **come**;*

‘ “I see that thou believest in these things Because I say them, but thou seest not **how**;

In (143) is reported the opening of the speech delivered by the eagle to the poet, in which it tells Dante that it understands that he believes what he is told, but also suspects that he does not understand how these things actually occur (*ma non vedi come* ‘but thou seest not how’). In this case, *come* ‘how’ functions as the syntactic head of an elided clause with a completive value: *ma non vedi come [accadono]* ‘but you do not see how [they happen]’.

Structures like the one reported in (144) are however more akin to appear in coordinated structures such as the one represented in (145):

(144) *Purg. XXVIII, vv. 49-51*

*Tu mi fai rimembrar **dove e qual** era Proserpina nel tempo che perdette la madre lei, ed ella primavera».*

‘Thou makest me remember **where and what** Proserpina that moment was when lost Her mother her, and she herself the Spring.” ’

(145) *Purg II, v. 95*

*se quei che leva **quando e cui** li piace*

if he who takes up **whom—and when**—he pleases

Examples (144) and (145) appears to be challenging as it seems to be an instance of sluicing followed by a coordinated clause containing the elided material, therefore classifying it as an instance of postcedent.

#### 5.1.3.9 *Stripping/bare argument ellipsis*

Among elliptical structures, instances of stripping, or bare argument ellipsis, also emerged (see Section 3.2.3.1). In the *Comedy*, stripping is more frequently attested in negative polarity, defined in the literature as not-stripping or negative stripping (Lee & Kim 2021), than in its simple positive form. As already shown in Lee & Kim’s (2021) corpus-based study on English, negative stripping is more likely to appear in specific registers, among which fiction, as is the case with the *Comedy*.

The patterns attested in (not-)stripping are varied, and their schemata are listed in Table 15 and described in the following Sections.

	<b>Antecedent</b> (V + _)	<b>Postcedent</b> (_ + V)
<b>a) Stripping</b>	i) V + <i>e</i>	-
	ii) <i>non</i> V + <i>ma</i>	-
<b>b) Not-stripping</b>	i) V + <i>e/ma non</i>	i) <i>non</i> + <i>ma</i> V
	ii) <i>non</i> V + <i>né</i>	
	iii) <i>non</i> V + <i>se non</i>	
<b>c) Nested stripping</b>	i) V + <i>non/né</i> + <i>né</i>	-
	ii) V + <i>non</i> + <i>ma</i>	-

**Table 15.** (Not-)stripping patterns

#### 5.1.3.9.1 (Positive) Stripping

**Stripping** in its positive polarity form (a) appears to be attested only with an antecedent pattern (V + \_). It can occur in affirmative sentences (a.i), as in Example (146), and in negative sentences (a.ii), as in Example (147).

- (146) *Purg.* XXII, v. 99  
*dimmi se son dannati, e in qual vico*  
‘tell me if they are damned, **and in what quarter**’

Example (146) records a request made by Statius to Virgil concerning the presence of other Latin poets in the underworld. In (146), *e in qual vico* ‘and in what quarter’ is a coordinate clause of the affirmative completive *se son dannati* ‘if they are damned’ in which only the oblique phrase is attested.

- (147) *Par.* XI, vv. 52-54  
*Però chi d’esso loco fa parole, non dica Ascesi, ché direbbe corto, ma Oriente*  
‘Therefore let him who speaketh of that place, Say not Ascesi, for he would say little, **But Orient**’

The verses in (147) are taken from the description of the birthplace of Saint Francis. In these lines, it is stated that referring to the city of his birth by the term *Ascesi*, the Old Tuscan form of *Assisi* (Chiavacci Leonardi 2012: 200), would be reductive (*ché direbbe corto* ‘for he would say little’), and that it would be more appropriate to use the term *Oriente*. A negative main clause, *non dica Ascesi* ‘say not Ascesi’, is followed by an ellipsis classified as stripping, since only a single phrase is present. It should be noted that the function is clearly contrastive.

#### 5.1.3.9.2 Not-Stripping

Moving on to instances of not-stripping, the patterns prove to be richer. Not-stripping is attested both with an antecedent (V + \_) and with a postcedent (\_ + V). As shown in Table 15, patterns with an antecedent are more numerous and detailed than those with a postcedent.

##### 5.1.3.9.2.1 Not-Stripping with antecedent in affirmative clause

Examples (148)-(150) present instances of not-stripping with antecedent, where the main clause is affirmative, following pattern (b.i) reported in Table 15. Pattern (b.i), V + *e/ma non*, involves the coordination, either with *e* ‘and’ or the adversative *ma* ‘but’, of two clauses.

(148) *Purg.* XIV, vv. 140-141

*e allor, per ristriagnermi al poeta, in destro feci, e non innanzi, il passo*

‘and then, to draw more near the poet, I moved to my right **instead** (of moving) **forward**.’

Example (148) describes a case of punished envy, specifically the episode from classical mythology, drawn from Ovid, of Aglauros’s metamorphosis into a statue as a result of the jealousy over Mercury’s love for her sister Herse (Chiavacci Leonardi 2012: 257). Aglauros says that she moved towards Virgil, shifting to the right without taking a step forward. In this example, an affirmative clause (*in destro feci (...) il passo* ‘I moved to my right’) governs an ellipsis defined as not-stripping, consisting of the negation *non* ‘not’ and the adverb *innanzi* ‘forward’. Here too, the specification of the direction of the movement and the use of not-stripping serve a contrastive function, further reinforced by the postposition of the object of the affirmative clause, *il passo* ‘the step’, after the not-stripping.

Patterns such as V + *ma non* in (b.i) are particular instances that display a peculiar characteristic that needs to be examined more closely in further studies. Two examples are given in (149) and (150):

(149) *Inf.* XIII, vv. 103-105

*Come l'altre verrem per nostre spoglie, **ma non però** ch'alcuna sen rivesta, ché non è giusto aver ciò ch'om si toglie.*

‘Like other souls, we shall seek out the flesh that we have left, **but none of us shall wear it**, it is not right for any man to have what he himself has cast aside.’

(150) *Purg.* VIII, vv. 49-51

*Temp' era già che l' aere s' annerava, **ma non sì** che tra li occhi suoi e ' miei non dichiarisse ciò che pria serrava.*

‘The hour had now arrived when air grows dark, but not so (dark) that it deprived my eyes and his of what—before—they were denied.’

Example (149) explains that on the Day of Judgment, when souls take back their bodies, the suicides, although they go to recover their remains (*verrem per le nostre spoglie* ‘we shall seek out the flesh that we have left’), will not clothe themselves with them (*ma non però ch'alcuna sen rivesta* ‘but none of us shall wear it’); having willingly deprived themselves of their bodies, they will not have them returned. Example (150) is a narrative description of the arrival of darkness; despite the obscurity, the gaze of the two poets opens onto the space that receives them.

Both Examples (149) and (150) make use of the contrastive value of stripping, thereby enhancing the narrative effect. In (149), the peculiarity is that the negation *non* ‘not’ precedes the conjunction *però che*, thereby negating the whole clause rather than the predicate alone. One might hypothesize the presence of an omitted verb *essere* ‘to be’, although this remains only a hypothesis. This type of construction warrants further investigation, particularly with regard to its diachronic development from Latin. In (150), by contrast, the ellipsis involves not only the verb but also the term of comparison *ma non sì (nera)* ‘but not so (dark)’; such structures, involving comparative deletion, will be discussed in Section 5.1.3.2.10.1.

#### 5.1.3.9.2.2 Not-Stripping with antecedent in negative clause

Now let’s move on the description of not-stripping patterns with an antecedent, there are also cases in which the clause containing the verb is negative, displaying both patterns *non V + né* and *non V + se non*, as in Examples (151)-(152):

(151) *Inf.* XXXII, vv. 7-9

*ché non è impresa da pigliare a gabbo discriver fondo a tutto l'universo, né da lingua che chiami mamma o babbo.*

‘For ’tis no enterprise to take in jest, To sketch the bottom of all the universe, **Nor for a tongue that cries Mamma and Babbo.**’

(152) *Inf.* III, vv. 7-8

Dinanzi a me non fuor cose create **se non etterne**, e io eterno duro.

‘BEFORE ME **NOTHING BUT ETERNAL** THINGS WERE MADE, AND I ENDURE ETERNALLY.’

Example (151) contains an excerpt from the invocation to the Muses preceding the narrative passage in the *Canto* concerning the most sinful souls. In these specific verses, the poet underscores the difficulty of describing the bottom of the universe, and even more so if one does not master the language. In this context, the not-stripping has an emphatic function.

Example (152) reproduces part of the inscription on the gate of Hell, which states that before its construction only eternal things were created, and that it will endure for eternity. The sententious formulation makes use of not-stripping, combined with a nominal ellipsis (*se non [cose] etterne*, ‘nothing but eternal [things]’). Once again, the formulation serves a narrative emphasis.

#### 5.1.3.9.2.3 Not-Stripping with postcedent

Continuing with not-stripping with a postcedent, an instance is illustrated in Example (153):

(153) *Purg.* XXIII, vv. 113-114

vedi che **non pur io**, ma questa gente tutta rimira là dove 'l sol veli

See that **not only I**, but all these people Are gazing there, where thou dost veil the sun.”

Example (153) records a soul’s request that Dante reveal his condition, as it is of interest not only to that soul but also to the other souls in that place. Once again, the use of not-stripping serves an emphatic function. Another example of not-stripping with a postcedent will be addresses in ellipsis of the copular *essere* ‘to be’, in 5.1.3.2.9.

#### 5.1.3.9.3 Nested (Not-) Stripping

Finally, as for gapping (see Section 5.1.3.2.1.3) also for (not-)stripping nested instances are attested, as shown in patterns in (c.i) V + *non* + *né* in Example (154) and (c.ii) V + *non* + *ma* in Example (155), both involving more than one ellipsis:

(154) *Inf.* XXVII, vv. 85-87

*Lo principe d'i novi Farisei, avendo guerra presso a Laterano, e non con Saracin né con Giudei*

‘The Leader of the modern Pharisees Having a war near unto Lateran, **And not with Saracens nor with the Jews**’

(155) *Par.* XIV, vv. 61-66

*Tanto mi parver sùbiti e accorti e l' uno e l' altro coro a dicer «Amme!», che ben mostrar disio d' i corpi morti: forse non pur per lor, ma per le mamme, per li padri e per li altri che fuor cari anzi che fosser sempiterne fiamme.*

‘One and the other choir seemed to me so quick and keen to say “Amen” that they showed clearly how they longed for their dead bodies— **not only for themselves, perhaps, but for their mothers, fathers, and for others dear** to them before they were eternal flames.’

Example (154) describes, in an ironic manner, the rivalry between Pope Boniface VIII and the Colonna family (Chiavacci Leonardi 2012: 476), neither of whom is explicitly named in the passage but both indicated through periphrasis (*Lo principe d'i novi Farisei* ‘The Leader of the modern Pharisees’ for the Pope, *presso a Laterano* ‘near unto Lateran’ for the Colonna). To emphasize the absurdity of this conflict, it is ironically pointed out that the Pope had not waged war against religious enemies, namely, the Saracens and the Jews, but rather against the Colonna themselves, making use of two successive instances of not-stripping to stress the point.

Example (155) describes the souls’ desire to regain their bodies. The poet’s sensitivity, here introduced into the narrative context by a *forse* ‘perhaps’ emerges in the following verses, expressed first through a not-stripping (in bold in the Example), followed by an adversative stripping that reinforces the value of the concept expressed (underlined). According to the poet, the souls’ longing to have their bodies back is not motivated by personal desire (*non pur per loro* ‘not only for themselves’), but for the sake of their loved ones, their mothers and fathers (*ma per le mamme, per li padri e per li altri che fuor cari anzi che fosser sempiterne fiamme* ‘but for their mothers, fathers, and for others dear to them before they were eternal flames’).

In both Examples (154)-(155), it is evident that the use of cumulative and nested not-stripping serves the narrative rhythm, focusing the contrast and intensifying the force of the message, ironically in (154) and emotionally in (155).

#### 5.1.3.10 *Ellipsis of the copular essere (to be)*

Another consistent cluster of ellipses involves the omission of the copular *essere* ‘to be’, which, having leaf status in UD annotation, is not marked with `orphan` but becomes evident when checking for promotion mismatches. The omission of the copula can involve both a postcedent or and antecedent and it occurs in gapping structures, namely when coordination is involved. Examples of antecedent is reported in (156), whereas postcedent in (157):

(156) *Inf.* XXIII, vv. 64-65

*Di fuor dorate son, sì ch'elli abbaglia; ma dentro tutte piombo, e gravi tanto*

‘Outside, these cloaks were gilded and they dazzled; **but inside they (were) all of lead, so heavy**’

(157) *Inf.* I, v. 67

*Rispuosemi: «Non omo, omo già fui*

‘He answered me: “**Not man**; I once was man’

Example (156) describes the cloaks worn by the hypocrites, outwardly covered in gold and inwardly lined with heavy lead. In this case, the verb *essere* ‘to be’ is omitted in both coordinated clauses (*dentro tutte piombo* ‘inside they (were) all of lead’; *gravi tanto* ‘so heavy’), but is expressed in the main clause *son* ‘were’. By contrast, in (35) the verb *essere* is postposed, also aided by a chiasmic structure that reinforces the concept. The verse is spoken by Virgil, who emphasises that he is now only a soul and no longer a man, although he once was a man in life. A similar construction is used in the following *Canto*, *Inf.* II, v. 32 (*Io non Enea, io non Paulo sono* ‘For I (am) not Aeneas, am not Paul’), but here spoken by the poet himself, who, while denying the identity, in fact draws a skilful comparison. Example (157) can also be clustered as an instance of not-stripping with a postcedent (Section 5.1.3.9.2.3). Omission of *essere* ‘to be’ can also occur in answer to question, as in (158):

(158) *Purg.* XXIII, vv. 13-15

*«O dolce padre, che è quel ch' i' odo?», comincia' io; ed elli: «Ombre che vanno forse di lor dover solvendo il nodo».*

‘ “O my sweet Father, what is this I hear?” Began I; and he answered: “**Shades that go Perhaps the knot unloosing of their debt.**” ’

In (158), asks his guide Virgil about the origin of the sound he hears. Virgil’s reply reveals that the laments and songs heard come from the shades who are damned there. The answer is given without introducing the verb *essere* ‘to be’, which, although present in Dante’s question (*che è* ‘what is’) is absent from the reply, which contains only the nominal element *ombre* ‘shades’.

Finally, Example (159) presents an instance in which the verb *essere* ‘to be’ is omitted; it is not textually present elsewhere, but is instead inferred from world knowledge.

(159) *Par.* VI, vv. 133-135

*Quattro figlie ebbe, e ciascuna reina, Ramondo Beringhiere, e ciò li fece Romeo, persona umile e peregrina.*

‘Four daughters, **and each one of them a queen**, Had Raymond Berenger, and this for him Did Romeo, a poor man and a pilgrim’

The verses describe Raymond Berenger and his daughters, each of whom was a queen. As is evident from the example itself, the verb is supplied by the reader.

#### 5.1.3.11 *Comparative clause*

Another consistent cluster includes instances of ellipsis in comparative clauses<sup>204</sup>. Given the richness of comparative structures in the *Comedy*, the number of elliptical structures in comparative clauses is not surprising. Comparative structures, and more specifically elliptical constructions within them, would merit a dedicated study, is unfortunately not undertaken here due to the specific focus such a study would require. In this regard, reference may be made to the analyses of comparatives in Old Italian by Pelo (2012: 441–465) and Belletti (2010: 1135–1143), the latter also addressing comparative ellipsis (ibid.: 1141-1142). The present section serves only to illustrate some cases of ellipsis in comparatives, without any claim to exhaustiveness in describing the strategies adopted by the author; this work is left for future investigation. Following the description in Section 3.2.3.4, the present Section is divided into a description of comparative deletion (Section 5.1.3.11.1) and comparative ellipsis (Section 5.1.3.11.2).

---

<sup>204</sup> As already noted in Section 2.2.2, in this study cases that can be classified as comparative complement are treated uniformly as instances of comparative ellipsis (see the discussion in Section 2.2.2 for the relevant arguments).

#### 5.1.3.11.1 Comparative deletion

Comparative deletion (Section 3.2.3.4.1) occurs when the gradable property within the comparative clause is omitted. In Section 5.1.3.9.2.1, an example of comparative deletion has already been shown in Example (149), and it is also exemplified in Example (160):

(160) *Purg.* IV, vv. 19-24

*Maggiore aperta molte volte impruna con una forcatella di sue spine l' uom de la villa  
quando l' uva imbruna, **che non era la calla** onde salìne lo duca mio, e io appresso, soli,  
come da noi la schiera si partìne.*

‘A greater opening oft times hedges up With but a little forkful of his thorns The villager,  
what time the grape im browns, **Than was the passage**—way through which ascended Only  
my Leader and myself behind him, After that company departed from us.’

Example (160) depicts a descriptive scene in which the two poets must pass through a cleft in the rock that will lead them to the ascent of Mount Purgatory. To describe the size of the opening in the rock, the narrator creates a comparison with an opening that a peasant might close with a forkful of thorns to prevent the entry of thieves (Chiavacci Leonardi 2012: 66–67). In the comparative clause (*che non era la calle* ‘than was the passage’), the element of comparison is omitted, leaving only the verb *era* ‘was’ (e.g., *che non era [grande / aperta] la calla*, ‘than was [large/open] the passage’). However, it is more common to find deletion of both the comparison element and the verb itself. Such instances, referred to as comparative ellipsis, are described in the following Section 5.1.3.11.2.

#### 5.1.3.11.2 Comparative ellipsis

As with all elliptical constructions, ellipsis in comparatives may involve multiple remnants, thus being explicitly annotated with the `deprel orphan`, or a single remnant, in which case it results in promotion and can therefore only be retrieved through targeted queries (Chapter 4). As specified in Section 2.2.2, cases of comparatives with a single remnant have also been annotated as adverbial clause of the subtype comparative (`advcl : cmp`), for consistency with the annotation of cases where comparatives have multiple remnants. An alternative would be to adopt the `obl : cmp deprel`, namely marking oblique with the subtype comparative, to distinguish phrasal comparative cases. However, there are studies showing that phrasal comparatives are in fact instances of deletion and could therefore be treated as clauses (Lechner 2004). As mentioned before, comparative ellipsis deserves more in-depth investigation, which is planned for future works.

The following Sections distinguish instances of comparative ellipsis with more than one remnant (Section 5.1.3.11.2.1), namely those marked with `orphan`, from comparative ellipsis with a single remnant (Section 5.1.3.11.2.2), which are therefore retrieved through promotion.

#### 5.1.3.11.2.1 Comparative ellipsis with more than one remnant

As for other types of ellipsis, also comparative ellipsis with more than one remnant may occur both with a postcedent or an antecedent. Instances of comparative ellipsis with an antecedent appear more frequently than postcedent ones. Examples (161)-(163) illustrate instances of comparative ellipsis with an antecedent:

(161) *Inf.* XVII, vv. 135-136

*e, discarcate le nostre persone, si dileguò come da corda cocca.*

‘And once our weight was lifted from his back, he vanished **like an arrow from a bow.**’

(162) *Purg.* XXVI, vv. 134-135

*Poi, (...), disparve per lo foco, come per l' acqua il pesce andando al fondo.*

‘he disappeared into the fire, just **as a fish, through water, plunges toward the bottom.**’

(163) *Par.* V, vv. 65-66

*siate fedeli, e a ciò far non bieci, come Ieptè a la sua prima mancia;*

‘Be faithful and not blind in doing that, **As Jephthah (was) in his first offering**’

In all three Examples (161-163), the poet constructs a comparison between a narrated event and an image that figuratively enhances the effectiveness of the scene. While in (161) and (162) the images derive from familiar scenes, such as the release of an arrow or the plunging of a fish, in (163) the comparison actively engages the reader’s knowledge and the biblical, theological, and philosophical background to which the citation refers. More precisely, in Example (161), the comparison is drawn between the swift disappearance of the monster Geryon (*si dileguò* ‘(he) vanished’), who has carried the two poets to the entrance of the Malebolge, and the release of an arrow from a bow (*come da corda cocca* ‘like an arrow from a bow’). Example (162), by contrast, compares the vanishing into the fire of a soul, specifically Guinizzelli, a contemporary poet whom the narrator regards as his father (*padre mio e de li altri miei miglior che mai rime d’amor usar dolci e leggiadre* ‘the father of me and of the others—those, my betters—who ever used sweet, gracious rhymes of love’, vv. 97–99) and as the founder of the *Dolce Stil Novo*, to which Dante himself belongs, to the plunging of a fish into

deep waters. In contrast to the two readily accessible images in (161) and (162), the comparison in Example (163) requires a more elaborate reading, appealing to the reader's prior knowledge. Jephthah was a judge of Israel who vowed to God that, should he return victorious from battle against the Ammonites, he would sacrifice the first person he saw coming out of his house (Chiavacci Leonardi 2012: 88). The first person Jephthah encountered upon his victorious return was his daughter, whom he consequently sacrificed. The comparison refers to the decision to make such a vow, and appears in a context in which the author warns against making it unjustly (*e a ciò far non biece* 'and not blind in doing that' = the vow), as Jephthah did.

Concerning comparative ellipsis with a postcedent, Examples (164)–(166) illustrate one instance from each *Cantica*:

(164) *Inf.* XXXI, vv. 144-145

*né, sì chinato, li fece dimora, e come albero in nave si levò.*

'nor did he, so bent over, stay there long, but, **like a mast above a ship**, he rose.'

(165) *Purg.* XXXII, vv. 136-139

*Quel che rimase, come da gramigna vivace terra, da la piuma, offerta forse con intenzion sana e benigna, si ricoperse*

'That which remained behind, **even as with grass A fertile region**, with the feathers, offered Perhaps with pure intention and benign, Reclothed itself'

(166) *Par.* XIX, vv. 58-60

*Però ne la giustizia sempiterna la vista che riceve il vostro mondo, com' occhio per lo mare, entro s' interna*

therefore, the vision that your world receives can penetrate into Eternal Justice no more **than eye (can penetrate) the sea;**

Examples (164)–(166) present cases of ellipsis in comparatives, in which the verb of the main clause follows the comparative clause. Example (164) describes another movement assisted by an infernal character, specifically Antaeus, one of the classical giants (Chiavacci Leonardi 2012: 544), who gently sets the two poets down at the bottom of Hell and then rises again, as towering as the mast of a ship. Example (165) occurs in the description of the holy chariot (Chiavacci Leonardi 2012: 587), which is covered with eagle feathers as a benevolent sign. The covering of the chariot is compared to that made by couch grass over fertile soil. Finally, Example (166) addresses the impossibility for

humankind to comprehend the infinity of God, describing such an attempt as akin to that of the human eye trying to penetrate the depths of the sea.

In all the previous Examples (161)-(166), the ellipsis, and therefore the effort to reconstruct the verbal meaning, occurs in the comparative clause. Example (167), however, shows a case in which the omitted verb is in the main clause, while it is present in the comparative clause:

(167) *Purg.* XXII, vv. 133-134

*e come abete in alto si digrada di ramo in ramo, così quello in giuso,*

‘**And even as a fir—tree tapers upward From bough to bough**, so downwardly (did) that’

In (167), the encounter with a tree along the path is described; this tree constitutes the punishment for the gluttonous souls. Rich in fruit, it is compared to a fir tree, which by its very nature becomes wider as its branches descend from the top of the crown. In this case, the ellipsis occurs in the main clause (*così quello in giuso* ‘so downwardly (did) that’). This type of ellipsis differs from gapping (Section 5.1.3.2.1), which occurs in coordination, although in this case as well the remnants in the main clause are more than one.

Finally, Example (168) illustrates an instance of comparative ellipsis with gapping:

(168) *Inf.* XXIV, vv. 50-51

*cotal vestigio in terra di sé lascia, qual fummo in aere e in acqua la schiuma.*

‘Such vestige leaveth of himself on earth. **As smoke in air or in the water foam.**’

The example is taken from Virgil’s exhortation to Dante to continue along the path without lingering, stating that one who idles leaves on earth a trace of himself similar to that which smoke leaves in the air and foam leaves in the water, thus alluding to the evanescence caused by inertia. In this example, the comparative clause (*qual fummo in aere* ‘As smoke in air’) is elliptical and is in turn coordinated with another elliptical comparative clause (*e in acqua la schiuma* ‘or in the water foam’), thus constituting a case of gapping.

#### 5.1.3.11.2.2 *Comparative ellipsis with a single remnant*

As with comparative ellipsis involving multiple remnants, cases with a single remnant are attested in the data, occurring with both an antecedent and a postcedent.

Examples (169)-(171) present instances with an antecedent. Examples (169) and (171) are cases of comparative clauses of equality, the former employing the adverb *come* ‘like’ as the introducing

element of the comparison, also named ‘comparative operator’ in Pelo (2012: 443), and the latter using *quale*, no longer in use in Modern Italian except in correlation with *tale* (*tale...quale* ‘like’). Example (170) reports a comparative of majority expressed through *più che* ‘more than’.

(169) *Inf.* III, v. 136

*e caddi come l'uom cui sonno piglia.*

‘and like a man whom sleep has seized, I fell.’

(170) *Purg.* IX, vv. 97-99

*Era il secondo tinto più che perso, d' una petrina ruvida e arsiccia, crepata per lo lungo e per traverso.*

‘The second, tinted of deeper hue than perse, Was of a calcined and uneven stone, Cracked all asunder lengthwise and across.’

(171) *Par.* IX, vv. 68-69

*mi si fece in vista qual fin balasso in che lo sol percuota.*

‘then appeared before my eyes like a pure ruby struck by the sun’s rays.’

Example (169) is the closing part of *Canto* III of the *Inferno*, in which the poet describes his fainting, a strategy used to depict the entrance into the realm of the dead and also employed elsewhere in the poem to (not) describe the transition from one place to another. The poet’s fainting is compared to that of a man falling asleep. In Example (170), the comparative of majority is used to describe one of the steps at the gate of Purgatory: the second step is said to be blacker than a deep purple. Finally, Example (171) describes Folquet of Marseille, a Provençal troubadour, who appears to the poet as radiant as a ruby struck by sunlight.

Moving to instances of postcedent, only cases of comparatives of equality are attested, as shown in Examples (172)–(174):

(172) *Inf.* I, vv. 22-27

*E come quei che con lena affannata, uscito fuor del pelago a la riva, si volge a l'acqua perigliosa e guata, così l'animo mio, ch'ancor fuggiva, si volse a retro a rimirar lo passo che non lasciò già mai persona viva.*

**‘And just as he who, with exhausted breath, having escaped from sea to shore, turns back to watch the dangerous waters he has quit, so (did) my spirit, still a fugitive, turn back to look intently at the pass that never has let any man survive.’**

(173) *Purg.* XXV, vv. 10-15

*E quale il cicognin che leva l' ala per voglia di volare, e non s' attenda d' abbandonar lo nido, e giù la cala; tal era io con voglia accesa e spenta di dimandar, venendo infino a l' atto che fa colui ch' a dicer s' argomenta.*

**‘And as the little stork that lifts its wing With a desire to fly, and does not venture To leave the nest, and lets it downward droop, Even such was I, with the desire of asking Kindled and quenched, unto the motion coming He makes who doth address himself to speak.’**

(174) *Par.* V, vv 91-93

*e sì come saetta che nel segno percuote pria che sia la corda queta, così corremmo nel secondo regno.*

**‘and even as an arrow that has struck the mark before the bow-cord comes to rest, so did we race to reach the second realm.’**

Example (172) presents the first simile of the poem, in which the poet likens his own state of mind to that of a shipwreck survivor who, having reached the shore, turns to look back at the waters that had threatened him. In Example (173), the poet again employs a simile, comparing his own state of mind to that of a young stork which, though eager to fly, holds back out of fear, mirroring Dante’s desire, later repressed, to speak. Finally, Example (174) describes the movement of Beatrice and Dante in the Second Heaven, portraying it as swift as that of an arrow that reaches its target before the bowstring has ceased vibrating. For each instance of comparison with a postcedent, a resumptive element is always present, namely, *così* ‘so’ in (172) and (174), and *tal* ‘such’ in (173). These resumptive elements appear to serve a rhythmic function, supporting the parallelism of the correlated elements (Pelo 2012: 454). The apparent obligatoriness of such elements in cases of postcedents, and more specifically when the standard of comparison precedes the compare, constitutes an indicator worthy of further investigation, extending the sample to include sentences in which the verb is not omitted.

#### 5.1.4 Merging qualitative and quantitative analysis: contextualizing predicate ellipses

The automatic extraction of all instances of clauses without a predicate (see 4.2.1.1 for the notion of predicate adopted here) was analyzed in two ways: Section 5.1.2 presents data on the quantification and diversification of such structures according to their syntactic position, with the aim of observing the syntactic distribution of these constructions and their structural impact; Section 5.1.3, by contrast, provides a qualitative analysis, based on manual inspection, to investigate the typology of the extracted data. The quantitative analysis initially provided evidence for a concentration of ellipses in comparative and coordination contexts, followed by direct speech and main clauses (see Section 5.1.2.1). The qualitative analysis, in turn, offered an insightful overview of the sample under scrutiny, highlighting a range of different constructions.

In this section, quantitative and qualitative data are combined to provide a more fine-grained account of the phenomenon and to clarify the percentages reported in Section 5.1.2.1. The focus is placed on the dependency relations (syntactic positions) that show the highest number of occurrences, namely main clause position (*root*), parataxis, reported speech (*ccomp:reported*), and coordination (*conj*). Comparative clauses are not considered here, as they constitute a separate block. Tables 16–19 provide an overview of the most frequent dependency relations intersected with the typology of constructions.

<b>Category</b>	<b>Total occurrences in <i>Comedy</i></b>
Elliptical narrative framing of reported speech	168
Exclamation	21
Gapping	10
Presentative and Locative	9
Stripping	4
Fragment and Truncation	3
Copular “essere”	2
Listing and stylistic repetition	1
Foreign Language	1
Other	3
<b>Total</b>	<b>222</b>

**Table 16.** *root* analysis

<b>Category</b>	<b>Total occurrences in <i>Comedy</i></b>
Listing and stylistic repetition	21
Elliptical narrative framing of reported speech	15
Gapping	13
Copular “essere”	11
Fragment and Truncation	6
Stripping	4
Exclamation	3

Presentative and Locative	3
Foreign Language	2
Other	2
<b>Total</b>	<b>80</b>

**Table 17.** *parataxis* analysis

<b>Category</b>	<b>Total occurrences in <i>Comedy</i></b>
Exclamation	42
Foreign Language	33
Fragment and Truncation	10
Presentative and Locative	5
Copular “essere”	2
Other	2
<b>Total</b>	<b>94</b>

**Table 18.** *ccomp:reported* analysis

<b>Category</b>	<b>Total occurrences in <i>Comedy</i></b>
Gapping (forward)	89
Copular “essere”	71
Stripping	26
Exclamation	22
Listing and stylistic repetition	17
Presentative and Locative	7
Fragment and Truncation	4
Foreign language	1
Elliptical narrative framing of reported speech	1
False positive	92
<b>Total</b>	<b>330 (238)</b>

**Table 19.** *conj* analysis

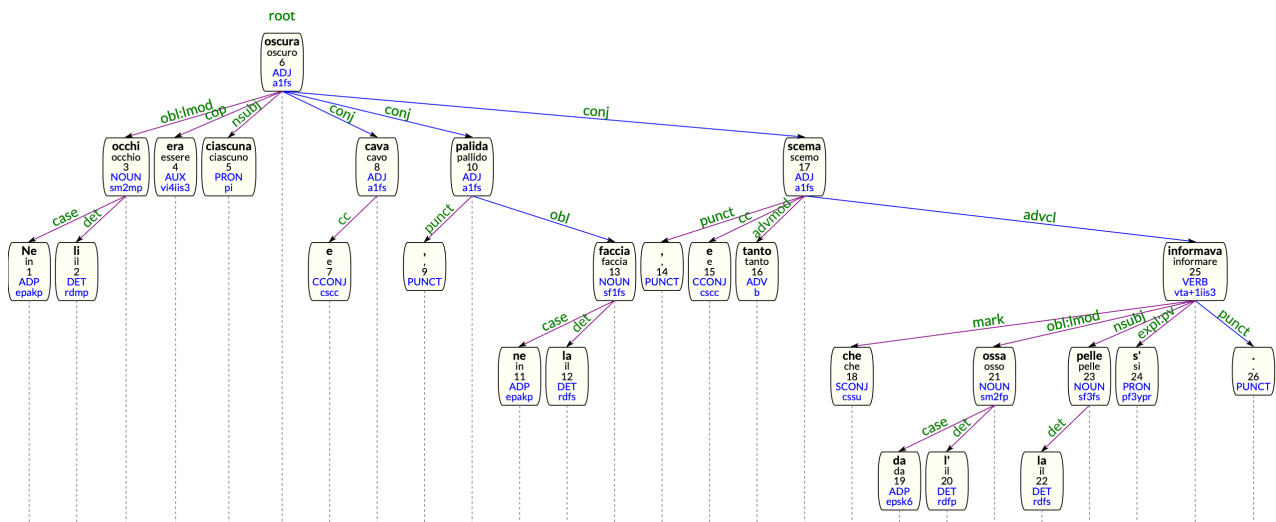
Instances of false positives in the analysis of *conj* are mainly due to the annotation system of the copula *essere* ‘to be’. In UD, the copula is not treated as a content word and is therefore not assigned the function of head, but rather that of dependent<sup>205</sup>; instead, another element is promoted to head, thereby assuming the syntactic position of the clause head. Consequently, in cases of coordination between two adjectives with a shared copula, one of the adjectives is promoted to clause head. Since the *conj* relation is not specific to either clausal or phrasal coordination (see Section 4.2.1.1.2.2), such instances are retrieved by the query, resulting in false positives. To illustrate, Example (175) and Image 163 report a case of false positive due to annotation, followed by cases that do involve omission of the copula.

<sup>205</sup> <https://universaldependencies.org/u/dep/cop.html>

(175) *Purg.* XXIII, vv. 22-24

*Ne li occhi era ciascuna oscura e cava, pallida ne la faccia, e tanto scema che da l'ossa la pelle s'informava.*

'Each in his eyes was dark and cavernous, Pallid in face, and so emaciated That from the bones the skin did shape itself.'

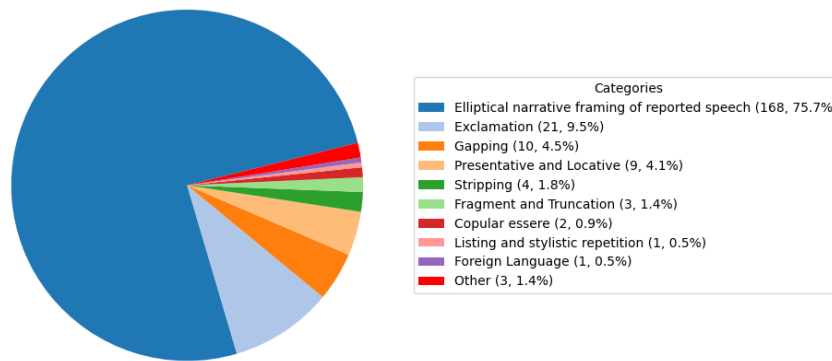


**Image 163.** Syntactic tree of *Purg.* XXIII, vv. 22-24

In Example (175), *oscura* ‘dark’ and *cava* ‘hollow’ are simply coordinated adjectives, both sharing the copula *era* ‘was’. However, due to the annotation, as shown in Image 163, *oscura* ‘dark’ is promoted to the head of the clause, and as a result *cava* ‘hollow’ is mistakenly extracted as an independent clause as well, thus representing a false positive. Interestingly, in the same example, *pallida* ‘pallid’ and *scema* ‘emaciated’ are instead interpreted as cases of copular omission rather than as coordinated elements of *oscura* ‘dark’ (as was the case for *cava* ‘hollow’). This interpretation is motivated by the presence of additional dependents: in the first case, *pallida* ‘pallid’ is modified by a locative phrase (*ne la faccia* ‘in face’), while in the latter *scema* ‘emaciated’ is accompanied by an intensifier *tanto* ‘so’ and occurs at the end of a description structured into three clauses, pattern of a descriptive tricolon (three syntactically parallel structures).

Returning to the discussion of syntactic positions, it is worth noting that the main clause position (`root`), as reported in Table 16 and illustrated in the pie chart in Image 164, displays a high proportion of elliptical narrative framing of reported speech (75.7% of the total), followed by exclamative structures (9.5%), gapping (4.5%), and presentative and locative constructions (4.1%), while all the remaining categories do not exceed 2%.

Distribution in root position (Comedy)

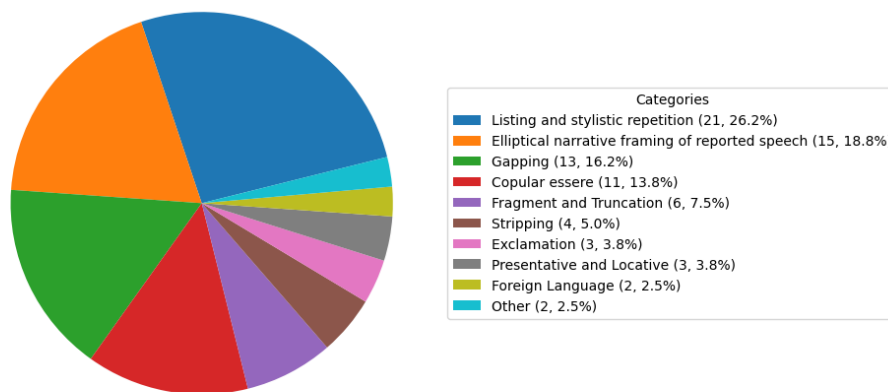


**Image 164.** Pie chart of categories in main clause position

Elliptical narrative framings of reported speech (Section 5.1.3.2) are ellipses shaped by the narrative need to introduce dialogic exchanges in a natural and fluent manner within the text. These exchanges are frequently introduced in the main clause, which accounts for the high number of ellipses occurring in root position.

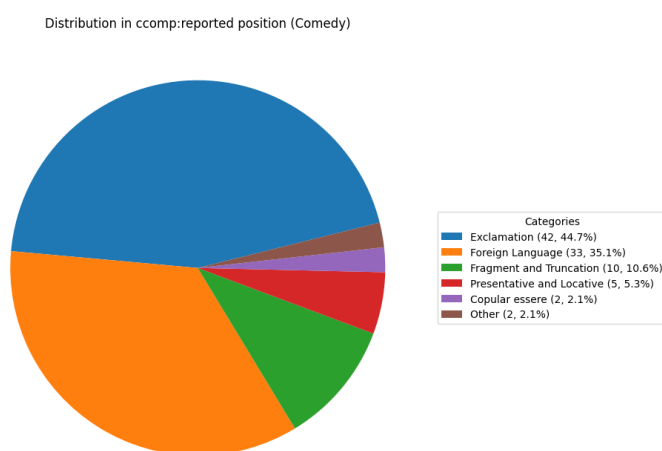
Concerning Table 17, which reports ellipses in parataxis position, the pie chart in Image 165 shows a less polarized distribution compared to the categories in main clause, reflecting the hybrid nature of the parataxis label, which is more closely related to the punctuation system adopted by the editor. The largest groups correspond to listing and stylistic repetition (26.2%), followed by cases of elliptical narrative framings of reported speech (18.8%), gapping (16.2%), and copular *essere* (13.8%), while all other categories account for less than 10% of occurrences.

Distribution in parataxis position (Comedy)



**Image 165.** Pie chart of categories in parataxis position

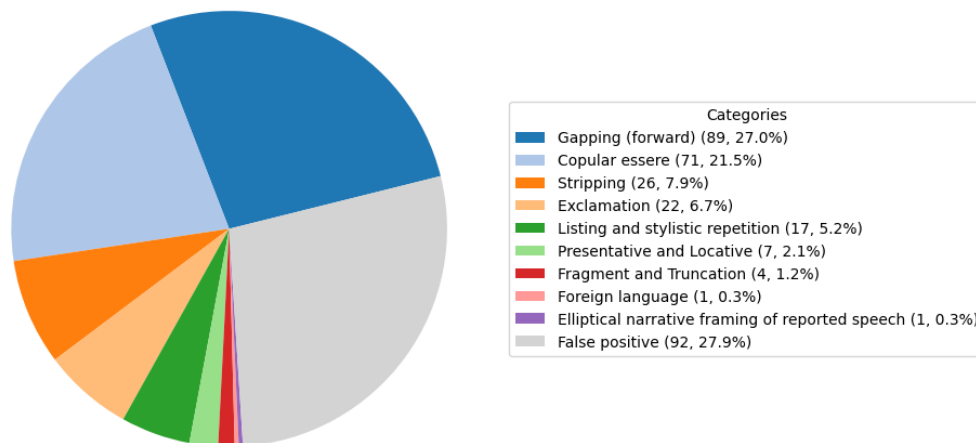
In reported speech (Table 18 and Image 166), the distribution is more polarized, with exclamations (44.7%) and foreign language insertions (35.1%) as the most frequent categories, followed by fragment and truncation (10.6%) and presentative and locative structures (5.3%). With the exception of the category of foreign languages, these are structures typical of spoken language, namely exclamations and fragments, which are strongly associated with the textual typology of direct speech. It is also noteworthy that no instances of elliptical narrative framings of reported speech are attested. These represent ellipses characteristic of narrative structure and are therefore not expected in direct discourse. Likewise, no cases of gapping or stripping are found. The absence of these latter constructions is significant in establishing that they are more typical of narrative framing.



**Image 166.** Pie chart of categories in reported speech position

Finally, the coordination analysis (Table 19 and Image 167) revealed issues with false positives arising from the annotation system, again underscoring the importance of distinguishing the `conj` relation at the clausal and phrasal levels (see Section 4.2.1.1.2.1). Despite the presence of false positives (27.9%), the main category attested is gapping (27%), specifically forward gapping, followed by cases of ellipsis of the copular *essere* (21.5%) and stripping (7.9%).

Distribution in clausal conj position (Comedy)



**Image 167.** Pie chart of categories in clausal coordination position

What emerges from this closer examination of the main cluster, combining quantitative and qualitative perspectives, is a fine-grained complexity that intertwines purely syntactic analysis, focusing on syntactic position of ellipsis, with the specialization of individual constructions. This includes cases in which the syntactic position is decisive, such as forward gapping, which occurs only in coordination, as well as cases that are more flexible in their syntactic position but are instead constrained by other factors, such as textual type and stylistic influences, as in instances of listing or stylistic repetition. This reinforces the idea and the necessity of contextualizing ellipsis as a cross-domain phenomenon (Kehler 2000) and, in light of the plurality of heterogeneous types that have emerged, the need to “adjust” the notion of ellipsis, which will be introduced in the following section.

### 5.1.5 What should be considered as predicate ellipsis?

Given the analysis presented in Section 5.1, can all the retrieved constructions be regarded as instances of predicate ellipsis?

Within the grid outlined above, cases involving foreign languages can be excluded, since their extraction is not genuinely related to predicate omission but rather to annotation gaps. Another debatable group includes exclamatives, presentatives, locatives, listings, and stylistic uses, for which the traditional notion of predicate (as discussed in 4.2.1.1.1) is challenged: here, the lack of a verb or copular element does not signal omission but rather a genuine absence, as such elements are neither syntactically nor semantically required for sentence completion. These cases, therefore, push forward the discussion on the very notion of predication. Other cases reveal a continuum. At one end lies the most explicit instance, gapping, where even from a dependency-syntactic perspective the ellipsis is

acknowledged: otherwise, the remnants would remain syntactically unattached in the syntactic tree. For this reason, the orphan relation is adopted, which makes the presence of an omitted verb overt. At the other end are structures such as stripping, sluicing, fragments, and similar constructions which, while receiving extensive analysis in constituency syntax, are discarded in dependency syntax. This is because they do not require the insertion of empty nodes, as no remnants need to be attached. As already noted in Chapter 3, the concept of ellipsis in syntax is closely tied to the notions of syntactic structure and empty nodes, and therefore varies depending on the framework adopted: on the one hand, frameworks that admit empty nodes; on the other, those that aim to adhere as closely as possible to the overt surface realization of the sentence. In light of these differences, the very notion of ellipsis proves to be highly framework-dependent, and thus both polarized and polarizing. Within the scientific community, however, this has not led to a shared usage or a unified mapping, but rather to a fragmented conceptualization shaped by the assumptions of each theoretical model, thereby fueling debate over the very essence of the phenomenon. In light of what has been stated above, an insoluble issue across different theories, it emerges that the focus of ellipsis should shift from being conceived as a mere empty node in the syntactic structure to being understood as a genuine absence in the text, as its very etymology suggests. By its intrinsic nature, absence should not be considered in and of itself, since a lack is not functional per se, but only in relation to what is present, which it highlights through its very absence. That is why, despite the structural debate in which the term ellipsis has paid the price of being strongly theory-dependent, a more neutral and predicate-less-oriented approach makes it possible to bring to light a variety of constructions that fall under the notion of predicate ellipsis, understood as absence in the text, thereby highlighting what remains.

## 5.2 Object omission

After automatically extracting the data (see [Section 4.2](#)), a manual analysis was carried out on 735 instances of object omission to classify the omitted objects according to several parameters. Specifically, the parameters under investigation are:

- i) whether the referent of the omitted object is generic or specific;
- ii) if the referent is specific, determining whether it is textual or non-textual;
- iii) if the referent is textual, its position, i.e., whether it functions as an antecedent (in the strict sense, preceding the omission) or as a postcedent (following the omission).

In addition to these manually annotated parameters, further information about the verb hosting the omitted object was automatically extracted from the treebank, namely its dependency relation, lemma, and PoS. Table 20 illustrates the distribution of omitted objects, grouped according to the parameters outlined above, across the *Cantiche*.

	<i>Inferno</i>	<i>Purgatorio</i>	<i>Paradiso</i>	<i>Commedia</i>
<b>Generic referent</b>	108	152	124	385
<b>Specific without textual referent</b>	21	7	28	56
<b>Specific with textual referent</b>	81	105	88	274
<b>Non-canonical pattern (other)</b>	5	7	8	20
<b>Total</b>	215	271	249	735

**Table 20.** cluster of object omission

The distinction underlying the first parameter, namely whether the referent is generic or specific, draws on the classification proposed for Modern Italian by Lo Duca (2000) and Cennamo (2015 2017)<sup>206</sup>, building on Fillmore (1986) (refer to [Section 3.2.2.2](#)). Within this framework, instances of object omission were annotated as generic whenever they could be replaced by a generic object, i.e., one paraphrasable as “something” (Fillmore 1986: 96). These cases are comparable to examples such as the one discussed by Fillmore (1986: 96) and reported in (176):

- (176) When my tongue was paralyzed I couldn't eat or drink.

---

<sup>206</sup> In her analysis, Cennamo (2017) also considers lexico-semantic properties of the verb (see Section 3.2.2.2). This dimension has not been addressed in the present study and remains a potential avenue for future research.

In this example, the two actions, eating and drinking, do not target specific items; rather, they are interpreted generically ('I couldn't eat or drink anything').

By contrast, when substitution with "something" is not possible and a more precise reconstruction of the referent is required, the omitted object is classified as specific. This label applies to all instances in which the omitted object refers to a particular entity, either as a contextual referent that is explicitly retrievable from the text (specific with referent), or as one that is inferable from world knowledge (specific without textual referent).

The latter case applies when the omitted object is present in the mind of the reader/listener, yet no explicit antecedent is provided in the text. These instances correspond to what Fillmore (1986: 97) terms "definite null complements", where the complement is defined as "definite omissible", that is, even if the referent "is not present in the sentence, its identity must be recoverable from the context" (Fillmore 1986: 97). For example, a sentence such as (177) is typically uttered in a context where the addressee is known or assumed as shared knowledge (Fillmore 1986: 98).

(177) I contributed five dollars.

Conversely, the label "specific with textual referent" applies to instances of object omission in which a referent is explicitly present in the text, having been mentioned either prior to the omission (as an antecedent) or afterwards (as a postcedent). This category is particularly relevant from a diachronic perspective, as it reveals variation in the constraints governing object omissibility over time (see [Section 3.3.1.2](#)). To clarify, the following example of object omission, already presented in [Section 3.3.1](#), illustrates an instance with a textual referent typical of Old Italian texts:

(178) *veggendolo il duca, gli fae lo maggio onore del mondo, e molto l'abbraccia e bacia teneramente*  
'seeing him the duke, does him the greatest honor in the world, and much embraces **him** and kisses (him) tenderly'

As shown in the Example (178), the verb *bacia* 'kisses' lacks an overt object. However, the referent of the omitted object can be retrieved from the masculine pronoun *l(o)* 'him', which appears before the preceding verb *abbraccia* 'embraces'. The referent is the same in both cases, referring to the duke. The following Sections provide examples of each type of object omission listed in Table 20, with data drawn from each *Cantica*. [Section 5.2.1](#) and [5.2.2](#) present examples of generic omitted objects and specific omitted objects without referent; [Section 5.2.3](#) illustrates instances of specific omitted

object with a textual referent along with all its subtypes; finally, [Section 5.2.4](#) presents instances that do not fit into the aforementioned categories, referred to here as non-canonical patterns.

### 5.2.1 Generic omitted objects

As mentioned in [Section 5.2](#), the label generic object is assigned to instances in which the omission can be paraphrased as “something”, thereby referring to a non-specific entity. This type of omission is attested in all the three *Cantiche* (see to Table 20) and appears to be the most frequently occurring among the various types of object omission. The following Examples (179–181) illustrate instances of generic object omission drawn from each of the three *Cantiche*.

(179) *Inf. I, vv. 55-56*

*E qual è quei che volontieri **acquista**, e giugne 'l tempo che perder lo face*

‘Even as he who glories while he **gains** will when the time has come to tally loss’

In Example (179), *acquista* ‘gains’ features a generic object that can be replaced by the generic term “something” leaving the nature of the gain open to the reader’s interpretation. The verse appears within a simile comparing Dante’s emotions to those of a miser or a gambler who gains something unspecified, presumably coins or material goods. Similarly, Example (180), which is taken from the *Purgatorio*, reflects an instance of generic omission with the verb *sente* ‘to feel’:

(180) *Purg. XXV, vv. 74-75*

*e fassi un' alma sola, che vive e **sente** e sé in sé rigira.*

‘and becomes one soul that lives and **feels** and has self-consciousness.’

These verses are taken from the description of human generation; more specifically, they depict how the *motor primo* ‘the First Mover’, that is, God, breathes into the fetus the intellectual soul (*spirito novo*, ‘new spirit’, v. 72), which, together with the fetus’s potential (*ciò che trova attivo quivi*, ‘all that is active in the fetus’, v. 73), forms a single soul that lives, feels, and is self-aware (*che vive e sente e sé in sé rigira* ‘soul that lives and feels and has self-consciousness’). In this example, *sente* ‘feels’ lacks an overt object and does not refer to a specific entity, rather, the verb refers more broadly to the full range of perceptions associated with life itself.

Finally, Example (181) also illustrates a case of generic object omission. This instance is taken from Saint Bernard’s prayer to the Virgin Mary.

(181) *Par. XXXIII, vv. 16-18*

*La tua benignità non pur soccorre a chi domanda, ma molte fiata liberamente al dimandar precorre.*

‘Your loving-kindness does not only answer the one who **asks**, but it is often ready to answer freely long before the asking.’

In this example (181), the kindness of the Virgin is said to assist those who ask, without specifying what is being asked for. In all the instances illustrated in Examples (179-181) the omitted object is generic, as it can be replaced by an unspecified “something”.

### *5.2.2 Specific omitted objects without textual referent*

Regarding specific omitted objects without textual referent, this category includes cases in which the omitted object has a specific referent that is defined in the mind of the reader or listener. Crucially, the omission cannot be adequately replaced by a generic term such as “something” (as for generic omitted objects, see [Section 5.2.1](#)), as the referent belongs to a particular set or category of entities. Unlike specific omitted objects with textual referent, the referent is not retrieved from an explicit antecedent in the text but is instead inferred from lexical cues or the broader discourse context. Examples of specific omitted objects without a textual referent are reported in Examples (182-184).

(182) *Inf. XV, vv. 121-124*

*Poi si rivolse, e parve di coloro che corrono a Verona il drappo verde per la campagna; e parve di costoro quelli che **vince**, non colui che **perde**.*

‘Then he turned round, and seemed to be of those who at Verona run for the Green Mantle across the plain; and seemed to be among them the one who **wins**, and not the one who **loses**.’

Example (182) is taken from the final verses of *Canto XV*, which is dedicated to Dante’s teacher, Brunetto Latini. In these lines, the poet describes Brunetto’s departure from the scene, drawing a comparison between Brunetto and the runners in a race held in Verona, referred to as *il drappo verde* (‘the Green Mantle’), in which the participants ran naked, just like the souls condemned in Brunetto’s circle. Here, Brunetto is defined as *quelli che vince* (‘the one who wins’), rather than as someone who loses, in a direct, though implicit, reference to the race itself. This formulation implies two specific omitted objects for the verbs *vince* ‘wins’ and *perde* ‘loses’, respectively.

(183) *Purg.* XIII, vv. 82-84

*da l' altra parte m' eran le divote ombre, che per l' orribile costura premevan sì, che bagnavan le gote.*

‘and to my other side were the devout shades; through their eyes, sewn so atrociously, those spirits **forced** (the tears) that bathed their cheeks.’

Moving to *Purgatorio*, Example (183) is taken from a passage describing the punishment of the envious souls. The scene is centered on the *contrapasso*<sup>207</sup> of vision, and from the very first mention of these souls, the poet engages with the theme of sight (Chiavacci Leonardi 2022: 376). The envious are initially introduced to the reader as disembodied voices: *verso noi volar furon sentiti, non però visti, spirito parlando* ‘when we heard spirits as they flew toward us, though they could not be seen—spirits pronouncing’ (vv. 25-26). Their punishment is made explicit in vv. 70-71: *a tutti un fil di ferro i cigli fora e cusce* ‘for iron wire pierces and sews up the lids of all those shades’. Their eyelids are sewn shut as a sign of penance, and the stitching is so tight that it causes them to weep, with tears rolling down their cheeks, as described in Example (5). In this example, the verb *premevan* (‘pressed out’ or ‘forced out’) appears without an expressed direct object. As in Example (182), and in contrast to the generic omissions discussed in [Section 5.2.1](#), the omitted object is not generic and cannot be replaced by a generic “something”. Rather, it refers to a specific object that is not explicitly stated but can be inferred from the context, namely the eyelids.

(184) *Par.* XXVII, vv. 61-63

*Ma l' alta provedenza, che con Scipio difese a Roma la gloria del mondo, soccorrà tosto, sì com' io concipio;*

‘But the high Providence, that with Scipio At Rome the glory of the world defended, Will speedily bring **aid**, as I conceive;’

Finally, turning to the last *Cantica*, Example (184) is taken from *Paradiso*, specifically from *Canto XXVII*, which is structured around a double register reflecting the contrast between the sorrow of the earthly world and the peace of Heaven (Chiavacci Leonardi 2022: 735). Verses 61-63 begin with the strong adversative conjunction *ma* ‘but’, spoken by Saint Peter, which shifts the focus from a denunciation of the earthly world, where corrupt ecclesiastical shepherds are compared to rapacious wolves (*In vesta di pastor lupi rapaci si veggion di qua sù per tutti i paschi* ‘In garb of shepherds the rapacious wolves are seen from here above o’er all the pastures!’, vv.55-56) to a renewed hope in the

---

<sup>207</sup> For an overview of the notion of *contrappasso*, see Faibisoff (2019).

higher providence of God (*l' alta provedenza* 'the high Providence' v. 129). Within this context, it becomes clear that the referent of the verb *soccorrà* 'bring aid' is not a generic "something", but rather a specific entity: the Church, previously depicted as corrupted and in need of salvation. Also in this Example, the omission, therefore, does not involve a generic object but a specific, contextually inferable referent that remains unexpressed in the text.

All the examples described in (182-184) are instances of what Fillmore defines as "definite null complement", that is, cases in which the omitted complement is not generic (as in Examples (179-181), where the reader is not required to infer a specific entity), but instead refers to a specific referent that, although not textually explicit, must be inferred from the surrounding context.

Some instances are more explicit and therefore easier to reconstruct, as in the aforementioned Examples (182-184), where the context naturally guides the reader/listener to complete the scene and identify the omitted object as the race, the eyelids, and the Church, respectively. By contrast, other instances of specific omitted objects without a textual referent are more opaque and require a greater interpretive effort from the reader to recover the missing element, almost as if the author were challenging the reader's attention and interpretive cooperation. A similar instance is provided in Example (185):

(185) *Par. XVIII, v. 130*

*Ma tu che sol per cancellare scrivi*

'But you who only **write** to then erase'

These verses are specifically addressed to Jacques Duèse di Cahors, Pope Giovanni XXII, whose pontificate lasted from 1316 to 1334 (Chiavacci Leonardi 2022: 517). In this passage, he is directly addressed, though not explicitly named, through the accusatory use of the personal pronoun *tu* 'you'. Giovanni XXII is accused of writing 'to then erase' (*per cancellare scrivi*). Drawing on historical knowledge of Giovanni XXII's papacy, a careful reader/listener can infer that this verse is far from incidental; rather, it encodes a specific allusion. Several commentators<sup>208</sup> interpret this passage as referring to the Pope's alleged tendency to issue excommunications only to revoke them in exchange for money. Under this interpretation, the omitted object of *scrivi* 'you write' would be the excommunications themselves.

---

<sup>208</sup> Among the commentators who support this interpretation are also figures close to Dante's time, such as Jacopo della Lana, who writes: "Cioè tu cherico o pastore che solo scomunichi per toller poi denari e cancellare tale ragione.", and L'Ottimo Commento: "tu che scomunichi e condanni, solo per avere danari, acciò che ricomunichi e cancelli" (both: [https://dante.dartmouth.edu/search\\_view.php?cmd](https://dante.dartmouth.edu/search_view.php?cmd)).

Therefore, while on the one hand the object of what is being written is clearly not a generic ‘something’, on the other hand its referent, absent from the text, is difficult to infer solely from the immediate context. Rather, it requires recourse to world knowledge, specifically concerning the historical figure of Giovanni XXII.

To conclude this section, an example is provided of a verb marked as transitive in DanteSearch, yet whose transitivity status is somewhat borderline, an instance that highlights the expressive flexibility of poetic language. More specifically, the verb under examination is *varcare* ‘to cross’ which appears a total of 8 times throughout the *Comedy*, used both with an explicit object (3 occurrences) and without one (5 occurrences). In cases where a direct object is present, the verb conveys the meaning of *oltrepassare* ‘to cross’, and the object clearly identifies the entity being crossed, as illustrated in Example (186):

(186) *Purg. X, v. 53*

*io varcai Virgilio*

‘I went past Virgil’

On the other hand, in instances where *varcare* occurs without a direct object, its meaning tends to shift toward *procedere* ‘to proceed’, with no object implied:

(187) *Purg. XII, v. 4*

*Lascia lui e varca*

‘Leave him and **hurry on**’

The second usage corresponds to what is often referred to as “uso assoluto” (Brambilla Ageno 1964), where the verb stands alone without an explicit complement. Determining the intended meaning of the verb in such cases depends heavily on the context. However, the context is not always explicit or concrete enough to fully clarify the intended meaning, particularly in poetic language, where structure and lexical choice are consciously crafted to evoke imagery and ambiguity. As a result, it can be difficult to determine whether the verb aligns more closely with the transitive sense (‘to cross [something]’) or with the intransitive sense (‘to move ahead’). It is also possible, though unverifiable, that such ambiguity was deliberately or skillfully employed by the author as a rhetorical device to engage the reader more actively in interpretation. An instance of this kind of ambiguity is illustrated in Example (188):

(188) *Par.* II, vv. 1-4

*O voi che siete in picciotta barca, desiderosi d'ascoltar, seguiti dietro al mio legno che cantando **varca**, tornate a riveder li vostri liti:*

'O YE, who in some pretty little boat, Eager to listen, have been following Behind my ship, that singing **sails along**, Turn back to look again upon your shores'

This passage forms the prologue to *Canto* II of *Paradiso*, where the poet directly addresses the readers, warning them of the boldness of the journey ahead, namely, the poetic treatment of the sublime realm of Paradise. This warning to the readers (Chiavacci Leonardi 2022: 30) is conveyed through the metaphor of a sea voyage: the little bark (*picciotta barca*) of the readers, who have followed the ship of poetry thus far, is now advised to return to familiar shores. In Example (9), the verb *varca* lends itself to both of the interpretations discussed above. On one hand, it may be read as *oltrepassare* 'to cross something', with the direct object left implicit; on the other, it may be understood as *procedere* 'to proceed', aligning with the *uso assoluto*.

Both readings are grammatically and semantically valid, and the choice between them depends on interpretation, an ambiguity which, in the absence of explicit authorial guidance, is ultimately left to the reader. Even if a definitive and univocal interpretation is not possible, and indeed, this openness is part of the text's beauty, as it invites the reader to take an active role in its interpretation, it is worth noting that commentators closer to Dante's time, and therefore more familiar to Old Italian usage, tend to support the first reading (i.e., use with an omitted but implied object). In the following examples (189–191), the interpretations<sup>209</sup> offered by L'Ottimo Commento, Benvenuto da Imola, and Francesco Buti, respectively, are presented with regard to this specific verse:

(189) L'Ottimo Commento (1333)

*valica il profondo e largo mare dell'Universo*

'he **crosses** the deep and vast sea of the Universe'

(190) Benvenuto da Imola

*che **varca**, idest, transit et intrat novam et profundam aquam,*

'who **crosses**, that is, passes through and enters a new and deep water'

(191) Francesco da Buti (1385-95)

***varca**; cioè passa lo mare*

---

<sup>209</sup> The interpretations are taken from: [https://dante.dartmouth.edu/search\\_view.php?cmd](https://dante.dartmouth.edu/search_view.php?cmd)

‘**varca**; that is, he crosses the sea’

All the commentators appear to interpret the use of *varcare* as transitive. As previously noted, it is not possible to determine with certainty which interpretation is correct, since both are grammatically acceptable and both use of the verb are attested elsewhere in the *Comedy*. Whether this instance represents an “uso assoluto” (Brambilla Ageno 1964) or a transitive verb with an omitted object whose referent is not textually expressed, it nonetheless illustrates how poetic language engages the reader’s imagination. By encouraging imagery and leaving interpretive space, the text grants the reader the freedom to fill in potential gaps, if, indeed, any gap is perceived at all.

### 5.2.3 *Specific omitted objects with referent*

As mentioned in [Section 5.2](#) and discussed in detail in [Section 3.3.1.2](#), specific omitted objects with textual referent have been the subject of scholarly investigation in the context of Old Italian. Some instances of this type reflect a diachronic shift from Latin to Italian (Luraghi 1998), and similar patterns have been observed across Romance languages more broadly (Giacalone Ramat 1990; Miller 1992). Before turning to a close analysis of this phenomenon in the *Comedy*, an overview of the types of specific omitted objects is provided with referents that have been identified in the text. Since arriving at a homogeneous classification proved difficult, given the variety of forms this phenomenon may take, the resulting categorization is based on types of behavior. The categories identified in the data are as follows: deictic *così*-constructions ([Section 5.2.3.1](#)), comparative constructions ([Section 5.2.3.2](#)), reduplication ([Section 5.2.3.3](#)), shared dependent structures ([Section 5.2.3.4](#)), lexically-driven patterns ([Section 5.2.3.5](#)), and finally, specific cases of non-sentential antecedents ([Section 5.2.3.6](#)). Within the classes of share dependent structures and lexically-driven patterns, further subclusters were necessary. Specifically, for the former, the following distinctions were made: shared dependent with a postcedent ([Section 5.2.3.4.1](#)), share dependent with an antecedent ([Section 5.2.3.4.2](#)) and finally share dependent in Modern Italian ([Section 5.2.3.4.3](#)). As for the latter, it includes a subclass defined as narrative/diegetic reference ([Section 5.2.3.5.1](#)).

#### 5.2.3.1 The deictic “*così*”

A significant portion of omitted objects with a specific and textual referent involves the use of the adverb *così* “so”, a deictic element that refers back to an element which would function as the object of the clause containing *così*, but which in fact appears in a different clause. This structure frequently occurs with *verba dicendi*, such as *dire* ‘to say’ and related verbs, which typically select for propositional complements. The most common functions of the adverb *così* include its use as a

referential device pointing to a preceding utterance, typically summarized by the adverb itself, which is most often placed at the beginning of a clause or sentence<sup>210</sup>. The following Examples (192-194) illustrate additional instances of this phenomenon.

(192) *Inf.* II, vv. 139-141

*Or va, ch'un sol volere è d'ambedue: tu duca, tu signore, e tu maestro». Così li dissi;*

‘Now go; a single will fills both of us: you are my guide, my governor, my master.’ **These were my words** to him;’ [lit. ‘(...). So to him I said;’]

(193) *Purg.* XXXII, vv. 43-47

«Beato se', grifon, che non discindi col becco d'esto legno dolce al gusto, poscia che mal si torce il ventre quindi». **Così** dintorno a l' albero robusto **gridaron** li altri;

‘“Blessed are you, whose beak does not, o griffin, pluck the sweet—tasting fruit that is forbidden and then afflicts the belly that has eaten!” **So**, round the robust tree, the others **shouted**;’

(194) *Par.* XXV, vv. 46-48

*di' come se ne 'nfiora la mente tua, e di onde a te venne». Così seguì 'l secondo lume ancora.*

‘o tell what hope is, tell how it has blossomed within your mind, and from what source it came to you’—**so did** the second flame **continue**.’

In all these instances, the verbs, belonging to the semantic field of saying, are separated from the statements that would normally function as their complement clauses. This separation is clearly a consequence of editorial choices that structure the dialogue or exchange as two separate syntactic units, leading to the treatment of the reported speech as an independent clause rather than as a complement clause (i.e., *ccomp:reported*) of the verb to say. In light of these examples (192-194), it becomes even more apparent how the editor’s decisions influence the shape of the text and its subsequent linguistic analysis. Thus, these cases of apparent object omission are not genuine syntactic omissions, but rather the outcome of editorial decisions in the structuring of the text. Moreover, there are also instances in which the adverb *così* is used in a proleptic function, implying that its referent appears after the deictic adverb:

(195) *Inf.* XXII, vv. 81-86

---

<sup>210</sup> [https://www.treccani.it/enciclopedia/cosi\\_\(Enciclopedia-Dantesca\)/](https://www.treccani.it/enciclopedia/cosi_(Enciclopedia-Dantesca)/)

*Fu frate Gomita, (...) e fé sì lor, che ciascun se ne loda. Danar si tolse, e lasciolti di piano, sì com'e' dice;*

‘Fra Gomita of Gallura, (...) but **handled** them in ways that pleased them all. He took their gold and smoothly let them off, as he himself says; [lit. ‘and made so them that’]

Example (195) is taken from *Inferno Canto XXII*, which describes the punishment of the grafters. Among these sinners, condemned to immersion in boiling pitch, is Frate Gomita, a friar guilty of bribery and corruption (Chiavacci Leonardi 2022: 386). His portrayal is brief and concise and, in line with the comic tone that characterizes this *Canto* (Chiavacci Leonardi 2022: 392), is delivered by the damned soul himself, who recounts his guilt alongside that of the other sinners in the pitch. In Example (195), the verb *fé* ‘handled’ (lit. ‘made’) lacks an explicit direct object in the sentence. It is immediately followed by the adverb *sì* ‘so’ and the clause introducer *che* ‘that’, without clearly stating what Frate Gomita actually did. The content of his actions is instead provided in the following sentence: *Danar si tolse, e lasciolti di piano*, ‘he took the money and let them go’ (referring to the enemies of Nino Visconti). In this instance, the direct object is not truly omitted but rather postponed, creating a more vivid and dynamic narrative rhythm through the delayed delivery of key information<sup>211</sup>.

### 5.2.3.2 Comparative structures

Another instance of the omission of specific textual elements is found in comparative clauses. Comparative clauses, such as the one shown in (196), are composed of a ‘first term of comparison’ and a ‘second term of comparison’ (Belletti 2001: 833), the latter of which can also be a verbal phrase (Belletti 2001: 840):

(196) *Ti aiuteranno più di quanto tu (non) pensi*

They will help you more than you (don’t) think.

As also shown in (196), the object of the verb in the comparative clause (*pensi* ‘think’), that is, the second term of comparison, is omitted but can be easily inferred from the main clause, the ‘first term of comparison’, on which the comparative clause depends. In (196), the omitted object of *pensi* ‘think’ is implicitly something like *che ti possano aiutare* ‘that they can help you’, therefore resuming and encapsulating the meaning expressed by the main clause itself (see Section 3.2.3.4.1). To this

---

<sup>211</sup> This use of *fare* ‘make’ followed by an adverb is also observed in Salvi et al. (2010: 123-189).

end, the observations are restricted to cases of similar object or complement clause omission in the *Comedy*. Cases of object/complement clause omission are not rare in the *Cantiche*, extending to various type of verbs. Below, one example from each *Cantica* is provided:

(197) *Inf.* XXIX, vv. 31-35

*O duca mio, la violenta morte (...) fece lui disdegnoso; ond'el sen gio senza parlarmi, sì com'io **estimo***

‘My guide, it was his death by violence (...) that made him so disdainful now; and, I **suppose**, for this he left without a word to me’

(198) *Purg.* XXV, vv. 133-135

*indi donne gridavano e mariti che fuor casti come virtute e matrimonio **imponne**.*

‘and they praised aloud those wives and husbands who were chaste, as virtue and as matrimony **mandate**.’

(199) *Par.* XXI, vv. 70-72

*Ma l' alta carità (...) sorteggia qui sì come tu **osserva***

‘the deep charity (...) allots our actions here, as you **perceive**’

In all the Examples (197)-(199) reported below, the verbs in the second term of comparison appears without an object, which, in each case, is omitted but specific. These omissions do not involve a generic or vague referent, but rather refer back to what is stated in the first term of comparison. More specifically, in (197), the verb *estimo* ‘suppose’ implicitly refers to what is expressed in the main clause, namely, that the violent death which made the spirit disdainful is the cause of the soul’s departure. Similarly, in Example (198), the verb *imponne* ‘mandate’ refers to the content of the relative clause *che fuor casti* ‘who were chaste’, which again is part of the first term of comparison. Finally, in (199), the verb *osserva* ‘perceive’ once again implies what is stated in the main clause, namely, that deep charity determines the distribution of actions.

Finally, one mention to a specific subtype, is constituted by the instances with the verb *fare* “to make”.

(200) *Inf.* V, vv. 96

*mentre che 'l vento, **come fa**, ci tace.*

‘While silent is the wind, as it is now.’ [lit. ‘the wind, as makes, here/us make-silent’]

Example (200) is taken from one of the most famous *Canto* of the *Inferno*, the one about Paolo and Francesca, the two lovers condemned to Hell for the sin of lust. In this specific verse, Francesca, who is speaking to the poets, declares her intention to tell them her story, taking advantage of the fact that the wind, the storm that sweeps the lustful souls away, has momentarily ceased. In (200), the verb in the comparative clause (*come fa* ‘as it is now’, lit. ‘as makes’), namely *fa* ‘makes’, does not take an object or complement clause, which is instead omitted and implicit. The referent is specific, as it can be identified and is textually present, referring to the action of falling silent. The verb *fare* ‘to do/to make’ is used as a *verbo vicario* (Pelo 2012: 459) and is closely tied to the comparative structure (Consales et al. 2003: 47). In such cases, *fare* is employed to avoid repetition of the verb, but also of the object itself<sup>212</sup>.

### 5.2.3.3 Reduplication

Omission of an object with a specific textual referent also occurs in instances of reduplication. By reduplication is meant the repetition of the same verb in a subsequent position, as in Example (201), where the verb *look* appears twice:

(201) Look, look what is happening!

This mechanism can be associated with the rhetorical figure of epanalepsis or *geminatio* (Mortara Garavelli 1997 :189), in which a word or expression is repeated at the beginning, in the middle, or at the end of a clause. It falls under the broader category of rhetorical figures based on repetition, which are considered a constant and defining feature of poetic discourse (Mortara Garavelli 1997: 187). In the *Divine Comedy*, rhetorical figures of repetition, such as the aforementioned epanalepsis, as well as anaphora and enumeration, serve to capture and retain the reader’s attention (Longo 2013).

In this section, the more neutral term reduplication is adopted, as the focus is not on the rhetorical function of repetition (for which Longo’s study (2013) remains relevant, though it does not address the specific cases discussed here), but rather on instances of object omission within such structures. As clarified, the term reduplication here refers specifically to verbal reduplication.

Interestingly, when reduplication occurs with transitive verbs, one of the verb typically lacks an explicit object, which is expressed only once across the two identical occurrences. The verb with the omitted object may either precede or follow its counterpart in which the object is overtly realized.

---

<sup>212</sup> See also the entry *fare* in the *Enciclopedia Dantesca* by Delcorno (Delcorno 1970: 794–803).

Examples in which the referent appears as a postcedent are provided in (202-203), while Example (204) illustrates a case in which the referent functions as an antecedent.

- (202) *Inf.* XIV, v. 69  
*ed ebbe e par ch'elli abbia Dio in disdegno*  
'and held, and seems to hold God in disdain'

In Example (202), the reduplication involves the verb *avere* 'to have/hold' which is repeated in a different tense to stress the continuity of the action involved, more specifically the first verb is expressed in the past (*ebbe* 'held'), and it is followed by the verb in the present (*abbia* 'seems to hold'). The verse refers to Capaneus, a Greek hero who supported the war in Thebes in favor of Polynices over his brother Eteocles (*quei fu l'un d'I sette regi ch'assiser Tebe* 'One of the Seven Kings was he who Thebes besieged' XIV, vv. 68-69). He condemned in the *Inferno* for having defied Jupiter and, as a result, was struck down by lightning. His blasphemy is portrayed as both enduring and continuous, persisting in life and in death, and the repetition in the text skillfully underscores this sense of unbroken defiance. In this Example, not only is the object expressed in the second instance of the reduplication, but also the phrase *in disdegno* 'in disdain', which is essential for reconstructing the meaning of the verb<sup>213</sup>.

- (203) *Purg.* XXVII, v. 31  
*Pon giù omai, pon giù ogni temenza*  
'put down, by now put down, your every fear'

In example (203), we observe that the reduplication may also involve other elements dependent on the verb, though not the object itself, which is expressed only once. The verse conveys the reassurance that Virgil offers to Dante, who is afraid to pass through the flames that will lead him to the final stairway of the mountain of Purgatorio. Virgil exhorts him to place his trust in him, having already shown the poet evidence of support and reliability, and now urges him to abandon all fear: *pon giù ogni temenza*, 'by now put down, your every fear'. As shown in (203), the reduplication also includes the locative adverb *giù* 'down', extending beyond the verb itself. Similarly to example (202), the object *ogni temenza* 'your every fear' appears only in the second occurrence of the verb.

---

<sup>213</sup> The expression *avere in disdegno* ('to hold in disdain' or 'to look down upon') can be considered a multi-word expression, although it is not annotated as such in UD. For an overview of multi-word expressions in NLP, see the PARSEME initiative (<https://parsemefr.lis-lab.fr/parseme-st-guidelines/1.3/>).

(204) *Para, X v. 52*

*Ringrazia, ringrazia il Sol de li angeli,*

‘Give thanks, give thanks Unto the Sun of Angels’

Finally, Example (204) reports a case of **anadiplosis** (Mortara Garavelli 1997: 191) namely the repetition of a segment in the first part of the following one in a scheme like “(...) X/X...”. This example is taken from Dante’s ascent to the Heaven of the Sun, which he reaches thanks to Beatrice’s guidance. She invites him to give thanks to God (*il Sol de li angeli*, ‘the Sun of Angels’) for having granted him such a privilege. Once again, the object is expressed in only one of the two occurrences in the reduplication, namely, in the second instance.

Object omission in reduplication is not only attested with a postcedent object, as shown in Examples (202-204), but can also occur with an antecedent, as illustrated in (205):

(205) *Inf. X, vv. 31-32*

*Drizza la testa, drizza, e vedi a cui s’aperse a li occhi d’i Teban la terra;*

‘Lift, lift your head and see the one for whom the earth was opened while the Thebans watched’

In (205), the omitted object appears in the second position, thus causing a structure that is not acceptable in Modern Italian where the object clitic is indeed needed. This point will be further addressed in Section 5.2.3.4. In this Example, Virgil urges Dante to lift his head and look upon one of the seven kings of Thebes, Amphiaraus (Chiavacci Leonardi 2022: 350). The exhortation is repeated to heighten the “urgency of the moment”<sup>214</sup>.

Finally, to conclude this section on reduplication, it should be noted that there are also instances in which the object, interpreted either as specific or generic, may be omitted in both occurrences of the repeated verb:

(206) *Inf. XIV, v. 57*

*“Bon Vulcano, aiuta, aiuta!”*

‘O help, good Vulcan, help!’

---

<sup>214</sup> Malato (2022: 507): *L’iterazione esprime la concitazione del momento* ‘The repetition expresses the urgency of the moment.’

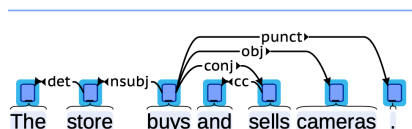
Example (206) presents Capaneus’s cry for help (see also Example 202). As Chiavacci Leonardi (2022: 251) notes, this form corresponds to the ancient form of a cry for help<sup>215</sup>, expressed through a conjugated verb in the second person singular without an explicit object, most likely implying a first person singular object pronoun. In modern usage, by contrast, the substantivized form (*aiuto*, ‘help’) is typically used, making the form in (206) ungrammatical without an explicit object (e.g., *aiutami* ‘help me’).

#### 5.2.3.4 Shared dependent

In this Section, instances of so-called **shared or conjoined dependent**<sup>216</sup> are presented, specifically, instances in which a dependent is shared by more than one element. In this context, the focus is on cases where a single object, shared by two coordinated verbs, functions as a common dependent, as illustrated in (207):

(207) The store buys and sells cameras.

In example (207), the two verbs ‘buys’ and ‘sells’ share the object ‘cameras’ (as well as the subject ‘store’, though the subject is not treated here as it is not the focus of this analysis). In basic dependency (refer to Section 1.4), the annotation is as follows (Image 168):



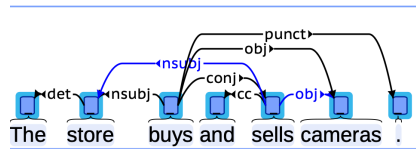
**Image 168:** conjoined verbs in basic dependencies<sup>217</sup>

The object (obj) ‘cameras’ is marked as depending on the first verb ‘buys’, while no syntactic connection to the second verb ‘sells’ is explicitly indicated. The syntactic dependency of the object with the second verb is, however, represented in enhanced dependency annotation (see Section 1.4.4), as illustrated in Image 169:

<sup>215</sup> Occurrences of *aiuta* ‘help’ without an explicit object are attested in the OVI Corpus, for instance in Lancellotto (volgarizzato, 14th c., Florentine). Refer to: [http://gattoweb.ovi.cnr.it/\(S\(y30lrumgrlzk4slce3qo5gjd\)\)/CatForm22.aspx](http://gattoweb.ovi.cnr.it/(S(y30lrumgrlzk4slce3qo5gjd))/CatForm22.aspx).

<sup>216</sup> The term is borrowed from the Universal Dependencies framework. Such instances are handled through enhanced annotation (see: <https://universaldependencies.org/u/dep/conj.html#shared-dependents-and-effective-parents-in-coordination> and <https://universaldependencies.org/u/overview/enhanced-syntax.html#conjoined-verbs-and-verb-phrases>.)

<sup>217</sup> The example and image are taken from: <https://universaldependencies.org/u/overview/enhanced-syntax.html#conjoined-verbs-and-verb-phrases>.



**Image 169:** conjoined verbs in enhanced dependencies

In this image, the dependency relations of both the object and the subject are marked not only on the first verb ‘buys’, but also on the second verb ‘sells’, as indicated by the blue arrows.

As already mentioned, *Italian-Old* treebank does not include enhanced dependency annotation, and therefore treats such cases as shown in Image 168. Whether one chooses to interpret them in terms of object omission or as instances of shared dependents, such constructions are nonetheless attested in the *Comedy*. These constructions are attested both with a postcedent object (Section 4.2.3.4.1), as in the example shown in Images 168-169, and with an antecedent object (Section 4.2.3.4.2), namely, instances in which the object appears near the first of the two verbs.

#### 5.2.3.4.1 Shared dependent with postcedent

Examples (208-210) display instances of postcedent objects.

(208) *Inf.* III, vv. 59-60

*vidi e conobbi l'ombra di colui che fece per viltade il gran rifiuto.*

‘I **saw** and **recognized** the **shade** of him who made, through cowardice, the great refusal.’

In example (208), the two verbs *vidi* ‘saw’ and *conobbi* ‘recognized’ share a common object: *l'ombra* ‘the shade’ of a soul whose identity remains debated among commentators (Chiavacci Leonardi 2022: 50). The spirit is described as having made a *gran rifiuto* ‘great refusal’ due to his cowardice, and for this reason is damned among the ranks of the pusillanimous souls. The two verbs sharing the object serve to both describe and introduce the character: he is first seen by the author and then recognized, indicating a logical sequence of actions that together present the figure to the reader.

(209) *Purg.* VIII, v. 10

*Ella giunse e levò ambo le palme*

‘It joined and lifted upward both its palms’

Example (209) describes the arrival of a soul who asks to be heard in preparation for prayer, which is sweetly pronounced in the following verses. The soul is depicted in the act of praying, by joining and raising its palms. As in example (209), the description unfolds through a sequence of actions, among which the gesture of prayer is conveyed by the two cojoined verbs *giunse* ‘joined’ and *levò* ‘lifted’ sharing the object *ambo le palme* ‘both its palms’, evoking the physical posture and image of a person in prayer.

(210) *Par. IX, vv. 127-130*

*La tua città (...) produce e spande il maladetto fiore*

‘Your city (...) produces and distributes the damned flower’

Finally, in example (210), the object *fiore* ‘flower’, which here represents the currency of Florence (*la tua città* ‘your city’), is shared by two verbs that introduce a sequence of events, *produce* ‘produces’ and *spande* ‘distributes’. Together, they describe the circulation of wealth as a corrupting force that leads the clergy astray.

Before turning to the discussion of antecedent instances, it is worth noting that postcedent objects may also occur when other elements intervene between the two coordinated verbs, disrupting their linear contiguity. An example is provided in (211):

(211) *Par. XIX, v.10*

*ch' io vidi e anche udi' parlar lo rostro*

‘For speak I saw, and likewise heard, the beak’

In this verse, the author describes to the reader how the group of radiant lights, souls of the blessed who had gathered to form an eagle, begins to speak by moving its beak. As in the previous examples (208-210), the verbs *vidi* ‘saw’ and *udi'* ‘heard’, which share the postcedent object *lo rostro* ‘the beak’, describe two sequential actions and serve to introduce the reader to the striking vision of the eagle. However, unlike in example (208-210), the two verbs here are separated by other elements, specifically, the adverb *anche* ‘likewise’, which reinforces the sense of veracity by emphasizing the involvement of two different senses in perceiving this vision.

All these instances of a shared object between two coordinated verbs, where the object is postcedent, are also attested and considered grammatical in Modern Italian. I report an example taken from ISDT treebank in (212):

(212) ISDT: isst\_tanl-148

*Invece nessuno può **sostituire** o **abbattere** il **capo** di un movimento.*

‘By contrast, no one can replace or overthrow the head of a movement.’

In example (212), as in examples (208-211), the shared object *capo* ‘head’ follows the second verb *abbattere* ‘overthrow’, which is coordinated with the first verb *sostituire* ‘replace’.

#### 5.2.3.4.2 Shared dependent with antecedent

The *Comedy* also contains instances in which two coordinated verbs share an object that is syntactically linked to the first verb, while the second verb neither express it overtly nor resumes it with a clitic, as would be expected in Modern Italian (a point that will be addressed again shortly). Some instances of shared dependent with antecedent are reported in the following Examples (213-215):

(213) *Inf.* V, vv. 68-69

*e più di mille ombre mostrommi e nominommi a dito*

‘and more than a thousand **shades** **did** he **name** and **point out** with his finger’

Example (213) is taken from Canto V, where the souls of the lustful are damned. Verses 67-68 strategically follow Virgil’s reported speech, in which he names to the poet the souls he recognizes among the damned. Since the crowd is so large, the author abbreviates the list by stating that the Guide pointed to and named more than a thousand shades among the lustful. In (213), the object *ombre* ‘shades’ is expressed only in combination with the first verb, whereas the second verb does not display it, nor an object clitic.

Interestingly, when examining the lectiones in Petrocchi (1994: 85), we find that some manuscripts report the variants *nominolle a dito* (Cha) and *nominolladito* (Vat) ‘name her (shades) with the finger’, both of which display an accusative clitic (*lle, ll* = the shades) in place of the dative one (*mi* ‘to me’). This serves as further evidence for the optionality of clitic omission, which may occur when the object referent shared between the verbs is the same, as in example (213), where the shared referent is *ombre* ‘the shades’. However, since the dative clitic *mi* ‘to me’ also refers to the same constant referent (*mi* ‘to me’ = to Dante), another possibility is to omit the dative clitic, option which is reflected in the two manuscript variants reported by Petrocchi.

(214) *Purg.* XIV vv. 1-3

«*Chi è costui che (...) apre li occhi a sua voglia e coverchia?*».

“‘Who is this man who (...) can, at will, open and shut his eyes?’”

Example (214) presents the opening of *Canto XIV*, which begins with the surprised question of a soul who inquires about the unusual living man, Dante, wandering through the mountain of Purgatory and blinking his eyes, unlike the other souls, whose punishment requires their eyelids to be sewn shut (see also Example X before). As in (213), the object *li occhi* ‘the eyes’ follows the first verb *apre* ‘can open’, whereas the second one, *coverchia* ‘shut’ is left without any object. Unlike in example (213), the critical apparatus of Petrocchi does not report any variant attesting a possible clitic pronoun. Finally, Example (214) reports another example of shared dependent with antecedent in the last *Cantica, Paradiso*.

(215) *Par.* XXVII v.73-75

*Lo viso mio seguiva i suoi sembianti, e seguì fin che 'l mezzo, per lo molto, li tolse il trapassar del più avanti.*

‘My sight **was following up** their **semblances**, and **followed** till the medium, by excess, the passing farther onward took from it’

In example (215), the author describes how he watched the rising of the spirits (*sembianti* ‘semblances’) for as long as it was possible for him to do so. As in examples (213-214), the object *sembianti* ‘semblances’ appears only with the first occurrence of the verb, but not with the second. Moreover, no variants are reported in this passage in Petrocchi’s edition (1994: 450). However, a noteworthy peculiarity emerges in this specific example. Whereas in all the previously discussed cases (213-215), the shared object not only refers to the same referent but also maintains the same syntactic function in both clauses, this example presents a case in which the shared object assumes different syntactic roles across the clauses. More precisely, the pronoun *li* (*li tolse* ‘took from it’), which refers to the same referent but fulfills a different syntactic role, is the element in question. In the phrase *li tolse* ‘took from it’, *li* functions as a dative pronoun. As a result, it is not omitted but it is overtly expressed.

Interestingly, the omission of the object does not involve only instances where the expressed object is in a nominal form (as in Examples 213-215), but the object can also appear in a pronominal form as shown in (216):

(216) *Purg.* XVI vv. 10-13

*Sì come cieco va dietro a sua guida per non smarrirsi e per non dar di cozzo in cosa che 'l molesti, o forse ancida m'andava io per l'aere amaro e sozzo*

‘E’en as a blind man goes behind his guide, lest he should wander, or should strike against aught that may **harm** or peradventure **kill him** So went I through the bitter and foul air’

Example (216) presents a comparison between Dante, who is crossing a space thick with smoke, where it is difficult to see, and who relies on Virgil’s guidance, and a blind man who follows his guide to avoid getting lost or bumping into something that might injure or kill him. In this example, the object, expressed only with the first verb *molesti* ‘harm’, appears in pronominal form, *l* ‘him’ (= the blind man). Petrocchi (1994: 262) records several variants in which the pronoun is also present with the second verb *ancida* ‘kill’, specifically *o chelluccida* ‘or that him kill’ (Laur) and *o forsi lancida* ‘or perhaps him kill’ (Mad).

Constructions such as those reported in examples (213-216) are not considered grammatical in Modern Italian, where an object clitic would typically be required in such contexts. Below, an example of Modern Italian in which two coordinated verbs share the same object in the same syntactic function is reported, but the clitic is explicitly expressed:

(217) ISDT - isst\_tanl-2196

*atleta che non nasconde lo sforzo ma lo rende tanto naturale*

‘An athlete who does not hide the effort but makes it seem completely natural.’

In example (217), the two verbs *nasconde* ‘hide’ and *rende* ‘makes seem’ both take *lo sforzo* ‘the effort’ as their object. It is expressed as a nominal dependent of the first verb and is resumed by the object clitic *lo* ‘it’, which depends on the second verb. Omitting the clitic in the second context, such as in (218), would result in an ungrammatical construction in Modern Italian:

(218) \**atleta che non nasconde lo sforzo ma rende tanto naturale*

‘An athlete who does not hide the effort but makes (it) seem completely natural.’

#### 5.2.3.4.3 Shared dependent in Modern Italian

As mentioned in Section 4.2.3.4.2, in Modern Italian, omitting the object clitic in the second verb is considered ungrammatical when the object of the first verb is a nominal (Salvi et al. 2001).

As extensively discussed in Chapter 4, conducting a corpus-based analysis to verify this claim requires an annotated corpus. However, unlike Italian-Old, Modern Italian treebanks do not include a transitivity/intransitivity tag, which makes it impossible to perform a detailed search of the kind conducted for *Italian-Old* (refer to Chapter 2).

Therefore, to extract a sample, albeit not exhaustive, given the annotation limitations, the following query was applied to retrieve contexts in which cases of shared dependents in Modern Italian may occur. The query applied is the following:

```
pattern {X1 [upos=VERB]; X2 [upos=CCONJ]; X3 [upos=VERB]; X1 < X2; X2 < X3}
pattern {X1 -[obj]-> Y}
```

### Query 3.

This query searches for a node X1 with part-of-speech tag VERB, followed by a coordinating conjunction (X2), and then by another verb (X3), with the linear order X1 < X2 < X3. Additionally, X1 must have a dependent node Y linked to it via the obj (object) relation.

Applying a similar query allows us to retrieve instances of potential shared dependents within a fixed pattern of “verb – coordinating conjunction – verb”, where at least the first verb has an object (without specifying its position). While such a query simulates the extraction of contexts involving shared dependents between two contiguous verbs, it fails to capture instances in which other elements interrupt the linearity between the two coordinated verbs, as in examples (211) and (213-216). Designing a more flexible query is not feasible without generating a large number of false positives<sup>218</sup>, thus requiring a thorough manual review of each instance. Such an analysis is left for future work.

For the purposes of the present study, the occurrences retrieved through Query 3 offer an interesting scenario. Table 21 reports the number of occurrences for treebank<sup>219</sup>, distinguishing between postcedent and antecedent:

	ISDT	MarkIT	PUD	ParTUT	ParlaMint	PoSTWITA	TWITTIRO	VIT	VALICO	Tot.
Postcedent	15	4	4	0	1	8	2	20	0	54
Antecedent	13	6	0	2	4	11	0	4	0	40
Total	28	10	4	2	5	19	2	24	0	94

**Table 21.** Modern Italian treebanks

More specifically, an examination of all the Modern Italian treebanks reveals that the most frequent instances involve a postcedent object (54 instances), that is, constructions like the one illustrated in example (219), which is repeated here for the sake of clarity.

<sup>218</sup> For instance, removing the requirement of strict linearity between the verbs and the conjunction yields 996 occurrences only in the ISDT treebank.

<sup>219</sup> Italian treebank correspond to UD version 2.16 (<http://hdl.handle.net/11234/1-5901>).

(219) ISDT: isst\_tanl-148

*Invece nessuno può **sostituire** o **abbattere** il **capo** di un movimento.*

‘By contrast, no one can replace or overthrow the head of a movement.’

In example (219), the two coordinated verbs *sostituire* ‘replace’ and *abbattere* ‘overthrow’ share a postponed object, *il capo* ‘the head’, thus forming a case of a shared dependent with a postcedent (see Section 5.2.3.4.1).

Interestingly, instances of shared dependents with an antecedent are not only fewer in number (40 instances), but also appear in highly specific contexts. Specifically, the antecedent object is always an object clitic (such as in Examples (220-221)) or a relative pronoun *che* ‘that’ (Examples 222-223), attested both with simple verbs (e.g., Example (220-222)) and with compound verbs, where the auxiliary ‘have’ is attested only once (e.g., Example (221)). Notably, clitic objects tend to occur in particular types of texts, such as bureaucratic documents (222) and tweets.

(220) ISDT- tut-536

*Il proprietario (1) può rivendicare la cosa da chiunque **la possiede** o **detiene***

‘The owner (1) may claim the item from anyone who **possesses** or **holds it**.’

Lit.

who her posseses or holds.

(221) ISDT- 2Parole\_5-113

*Poi **li hanno arrestati** e **rinchiusi** nei campi di concentramento*

‘Then they **arrested them** and **locked** them up in concentration camps.’

Lit. They them arrested and locked up

(222) Mark-IT: 361

*Questa è la domanda che bisogna farsi al giorno d'oggi: quello **che io studio** o **faccio**, ha poi un riscontro economico?*

‘This is the question one must ask today: does **what I study** or **do** have any economic relevance or impact?’

(223) ISDT: isst\_tanl-1474

*In fondo è la via **che hanno seguito** e **seguono** da anni gli Stati Uniti.*

‘After all, it is the path **that** the United States **has followed**—and **continues to follow**—for years.’

The query retrieved only one instance in which the antecedent object is a nominal phrase, and it occurs within a relative clause (see Example 224).

(224) VIT-6400

*Bloch si era rifatto alla vicenda di Ed Gein, arrestato nel 1957 nel Wisconsin per aver massacrato dozzine di attempate signore capitate nella sua farm, i cui cadaveri straziava e conservava nelle forme più disparate.*

‘Bloch had drawn on the case of Ed Gein, arrested in 1957 in Wisconsin for having massacred dozens of elderly ladies who happened to come across his farm, whose **corpses** he **mangled** and **preserved** in the most disparate forms.’

Even though the sample analyzed is limited due to annotation constraints, and further investigation is needed to fully explore the phenomenon in Modern Italian, a clear pattern appears to emerge. The omission of the object in shared dependent constructions seem to be restricted to cases in which the object appears as a clitic or a relative pronoun. Instances involving nominal objects, such as the one reported in example (219), do not exhibit object omission; instead, the object is resumed by an object clitic. Constructions like the one in example (224) remain marginal and warrant further analysis.

#### 5.2.3.5 Lexical properties

Another interesting cluster that emerged from the analysis of the dataset consists of specific verbs, such as *potere* ‘can’, *sapere* ‘to know’, *credere* ‘to believe’, *dire* ‘to tell’, among others. Some instances of object omission with these verbs are attested also in Modern Italian, whereas others are ungrammatical.

In the following Examples (225-226), some instances are object omission with the modal verb *potere* ‘to can’ are reported.

(225) *Inf.* XXI, vv. 53-54

*Covertò convien che qui balli, sì che, se puoi, nascosamente accaffi*

‘It here behoves thee to dance covered, That, if (thou) **canst**, thou secretly mayest pilfer.’

Example (225) is taken from *Canto XXI* of the *Inferno* and the verses report a direct speech uttered by a devil who mockingly scorns a sinner. He gives the order to make the sinner dance in the boiling pitch, thus ironically alluding to his ability to sneak away, the same kind of deceitful behavior that

constitutes the sin punished in this *bolgia*, namely, barratry. In this case, the verb *potere* ‘can’ takes an omitted clausal complement whose antecedent is the subsequent clause, namely ‘if you are able to secretly pilfer’, thus exhibiting a postcedent.

By contrast, Example (226) illustrates a case of object omission in which the referent is an antecedent rather than a postcedent.

(226) *Inf.* XVI, vv. 124-125

*Sempre a quel ver c'ha faccia di menzogna de' l'uom chiuder le labbra fin ch'el **puote***

‘Faced with that truth which seems a lie, a man should always close his lips as long as he **can**’

Example (226) is taken from *Canto XVI* of the *Inferno*, and these verses prepare the reader for a mysterious revelation that will only be disclosed in the following *Canto*. In this passage, Dante asserts that one should remain silent about events that appear to be lies, even when they are true. However, in this case, he suggests that silence is no longer an option. The verb *potere* ‘can’ does not feature an overt object, and its referent is the preceding clause, namely, as long as he can close his lips, thus representing an instance of a clausal antecedent.

However, the omission of the object with the verb *potere* is also attested in Modern Italian, as shown in Example (227), and is therefore not a phenomenon restricted to Old Italian.

(227) ISDT – isst\_tanl-1907

*Lo studio quando **posso**.*

‘I study it whenever I can.’

In (227), the object of *posso* ‘can’ is not made explicit but remains implicit. Its referent is the main clause, thus constituting an instance of a clausal antecedent, as seen in (226) for Old Italian.

Turning to another verb, *sapere* ‘to know’, exhibits the same behavior as *potere*, occurring with both a postcedent referent (Example 228) and an antecedent one (Example 229).

(228) *Purg.* II, vv.59-60

*Se voi **sapete**, mostratene la via di gire al monte.*

‘Do show us, if you **know**, the way by which we can ascend this slope.’

Example (228) presents a case of reported speech, in which a group of newly arrived souls in *Purgatorio* ask Dante and Virgil for directions. In this example (228), the omitted object of *sapere* is

the main clause that follows it, namely, *sapete la via* ‘if you know the way’, thus constituting a postcedent. By contrast, Example (229) illustrates a case of object omission in which the referent is an antecedent.

(229) *Purg.* XXVI, vv. 89-90

*se forse a nome vuo' saper chi semo, tempo non è di dire, e non saprei.*

‘Wouldst thou perchance by name know who we are, There is not time to tell, nor **could I do** (it).’

In Example (229) the implicit object of *saprei* refers back to an antecedent that can be reconstructed in more than one way, either as *non saprei chi siete* ‘I don’t know who you are’ or *non saprei dire i nomi* ‘I don’t know how to tell (you) the names’. While a single, definite antecedent cannot be clearly identified, the context enables the reader to infer the referent through discursive cues. Despite the ambiguity in reconstructing an object or clausal complement for the verb *saprei*, its meaning remains fully recoverable from context.

When checking for occurrences of the verb *sapere* without an object or a complement clause in Modern Italian treebanks, no occurrences such as those attested in (228-229), where a textual referent is encoded, are found<sup>220</sup>. Instances of object omission with generic, or with a specific but non-textually anchored referent, are attested (as in Example 230). In contrast, when the referent is textually present and postponed (i.e., in cataphoric use), it is explicitly expressed (Example 231).

(230) ISDT-tut-2865

*La pantera è un animale schivo e notturno, ma non si sa mai*

The panther is a shy and nocturnal animal, but you never **know**.

(231) ISDT- isst\_tanl-3264

*Lo sanno oramai quasi tutti che un respiro affannoso è sintomo di una vita affannata.*

By now, almost everyone **knows** (it) that laboured breathing is a symptom of a laboured life

Therefore, differently from the verb *potere*, whose use without an object clitic is grammatical both in Modern and Old Italian, for the verb *sapere*, the omission of the object clitic seem to be more rigid in Modern Italian, where instances as (228-229) are ungrammatical. For the verbs with object omission listed in this section, no syntactic cues appear to account for the omission; rather, it seems to be driven by their lexical properties.

---

<sup>220</sup> The manual check was conducted on the ISDT treebank as a preliminary investigation.

### 5.2.3.5.1 Narrative/diegetic function

In some instances, specific verbs are used in a narrative or diegetic function to introduce a comment by Dante as the author, thereby reinforcing the narrative itself<sup>221</sup>. Such verbs typically occur parenthetically, without an expressed object or object clause. They are predominantly verbs like *dire* ‘to say’ or *credere* ‘to believe’, that is, verbs which convey the author’s assertion or stance regarding the truth of what is being narrated. These verbs are most often used in the present tense (see Examples 232 and 233):

(232) *Purg.* III, vv. 67-71

*Ancora era quel popol di lontano, i' dico dopo i nostri mille passi, quanto un buon gittator trarria con mano, quando si strinser tutti ai duri massi (...) e stetter fermi e stretti*

‘Still was that people as far off from us, After a thousand steps of ours I say, As a good thrower with his hand would reach When they all crowded unto the hard masses (...) and motionless stood and close’

Example (232) is taken from *Canto III* of the *Purgatorio* and describes Dante and Virgil approaching a group of souls that appear to the two poets from afar. To highlight the distance of the souls (*ancora era quel popol di lontano* ‘Still was that people as far off from us’ v. 67), Dante emphasizes their remoteness by using a first-person present tense verb *dico* ‘I say’, which interrupts the past perfect narration (*si strinser* ‘crowded’ v. 70; and *stetter* ‘stood’ v. 71) to insert a clarification that underscores the development and coherence of the narrative, enhancing the truthfulness of the narrative discourse.

It is interesting to note that Inglese’s edition (Inglese 2011: 64) chooses to isolate the parenthetical clause by using brackets, as shown in example (233) below:

(233) *Purg.* III, vv. 67-69 (Inglese)

*Ancora era quel popol di lontano (io dico dopo i nostri mille passi) quanto un buon gittator trarria con mano*

‘Still was that people as far off from us, After a thousand steps of ours I say, As a good thrower with his hand would reach’

---

<sup>221</sup> About the distinction of Dante as a poet and Dante as an author refers to: (Spitzer (1946); Contini (1970); Picone (1999)).

This diegetic function of *dico* is further emphasized through typographical punctuation. In this instance, the absence of an object for the verb *dico*, a feature also noted by Petrocchi (1994: 45), brings its meaning closer to that of ‘to mean’ or ‘to imply’, where the omitted object (that which is said or meant/implied) refers to the remoteness of the group of souls.

In Example (234), the same function performed by the verb *dico* in Example (232) is now fulfilled by the verb *credo* ‘believe’:

(234) *Par. XXVIII, vv. 37-39*

*e quello avea la fiamma più sincera cui men distava la favilla pura, **credo**, però che più di lei s'invera.*

‘And that one had its flame most crystalline From which less distant was the stainless spark, I **think** because more with its truth imbued.’

In Example (234), Dante author describes the movement of the angelic intelligences, which vary in speed and intensity of light depending on their distance from the center. The brightness of the innermost circle is explained by Dante as a consequence of its realization in pure radiance (*però che più di lei s'invera* ‘because more with its truth imbued’ v. 39). Once again, the use of a first-person verb *credo* ‘I think’ breaks the flow of the narration, reinforcing its truthfulness through the author’s own commentary, which serves to lend credibility to the account. Again, as in Example (232), the object of the first-person verb *credo* is left unexpressed, but it is easily inferred from the context and corresponds precisely to the sentence in which it is embedded as a parenthetical (*quello aveva la fiamma...* ‘one had its flame...’). This construction downplays the verb that expresses the narrator’s opinion, placed, in fact, in a parenthetical position, in order to foreground the progression of the events being presented to the reader.

However, such verbs can also occur within the dialogue of characters, thereby reinforcing the truthfulness of the narrative and introducing psychological nuances, as seen in Example (233), in a direct speech by Virgil.

(235) *Inf. XX, v. 64-66*

*Per mille fonti, **credo**, e più si bagna tra Garda e Val Camonica e Pennino de l'acqua che nel detto lago stagna.*

‘By a thousand springs, I **think**, and more, is bathed, ‘Twixt Garda and Val Camonica, Pennino, With water that grows stagnant in that lake.’

Example (235) is taken from *Canto V* of the *Inferno*, where Virgil describes his hometown, Mantua. These verses, whose exegesis remains debated (Chiavacci Leonardi 2012: 353), are part of a long direct speech delivered by Virgil, in which he describes a region situated between Lake Garda, Val Camonica and the Apennines, and crossed by many rivers whose sources are said to be countless (*mille fonti* ‘a thousand springs’), using the dubitative expression *credo*, which here serves to emphasize the uncontrollable abundance of those sources. As in Example (234), the verb presents no present no explicit object, which is instead recoverable from the sentence in which it appears.

#### 5.2.3.6 Non-sentential antecedent

This section examines instances of non-sentential referents, namely cases in which the specific referent of an omitted object is recovered from a preceding sentence. This is made possible not only by the reader’s background knowledge (as in Example 236), but also by the clarity and continuity of the discourse topic (Examples 237-238).

(236) *Inf.* XXXI, vv. 16-18

*Dopo la dolorosa rotta, quando Carlo Magno perdé la santa gesta, non **sonò** sì terribilmente Orlando.*

‘After the dolorous discomfiture when Charlemagne the holy emprise lost, so terribly Orlando **sounded** not.’

In Example (236), the verb *sonò* ‘sounded’ features an omitted and “potentially” specific object. By “potentially”, I mean that the referent of the omitted object can be inferred only by an attentive reader familiar with epic literature. In this passage, Dante relies on the reader’s knowledge of epic tradition to reconstruct the missing element. The verses establish a parallel between a horn Dante hears (explicitly mentioned and described in verse 12: *ma io senti’ sonare un alto corno*, ‘but I could hear the blare of a loud horn’), whose powerful sound captures his attention, and the horn blown by Orlando, Charlemagne’s paladin, during his defeat at the Battle of Roncesvalles. This intertextual echo evokes the major epic tradition of *La Chanson de Roland*. Although the object remains unexpressed in verse 18, the reference to a well-known epic context enables a skilled reader to recover the non-sentence antecedent, which occurs in verse 12.

(237) *Purg.* XXI, v. 64

*Prima **vuol** ben, ma non **lascia** il talento*

‘First it **wills** well; but the desire **permits** not,’

In Example (237), the verbs *vuol* ‘wills’ and *lascia* ‘permits’ do not feature an overt object, although a specific one is implied. The two verbs refer to the action of *salir sù* ‘to mount aloft’ (v. 60), which is the cause of the earthquake that Dante inquires about in this *Canto*. In *Purgatorio*, as explained by a spirit whom Dante and his guide Virgil encounter, later revealed to be Statius, earthquakes are manifestations of the purified soul’s will to ascend to God. This will is present in the souls from the beginning, but the *talento* ‘desire’ (v. 64), the instinctive desire shaped by divine punishment (Chiavacci Leonardi 2022: 379), initially prevents it from being realized. The implicit objects of the two verbs in verse 64 are specific and refer not to a sentence antecedent, but rather to material found in preceding verses, at a distance of one sentence.

A similar case, though with an even greater distance from non-sentence antecedent, is found in Example (238):

(238) *Par. XXV, vv. 76-77*

*Tu mi stillasti, con lo stillar suo, ne la pistola poi;*

And just as he instilled, you then instilled with your Epistle

In Example (238), the central theme of the *Canto* is the hope, the virtue represented by the saint encountered in this *Canto*, Saint James. In these verses, Dante escribes his understanding of hope to the saint, defining it as *uno attender certo de la gloria futura* ‘the certain expectation of future glory’ (vv. 67-68) and explains how this virtue was instilled in him, namely, by David and by the Apostle James himself, through his epistle. In verse 76, the object of the verb *stillasti* ‘instilled’, which is hope itself, the central topic of the entire discourse, is omitted and not explicitly present in the sentence. The term appears earlier in verse 67 (*Spene* ‘Hope’ v. 67), and is referred to again in verse 71 through the object clitic *la* ‘it’ (lit. ‘her’, as hope is feminine in Italian), which occurs with the same verb: *ma quei la distillò nel mio cor pria* ‘but he who first instilled it in my heart’ (v. 71). This is a clear example on how the object clitic can be freely omitted. Interestingly, when consulting Petrocchi’s edition, the editor notes that one manuscript attests to a different version in which the object clitic is present in verse 76: *Tul mi* (Urb) (Petrocchi 1994: 417). The order of the clitics, namely, the accusative (*l = la* ‘it’) preceding the dative (*mi* ‘to me’), is typical of Old Italian usage (Salvi et al., 2010: 443). This attestation reinforces the optionality of the object clitic.

#### 5.2.4 Non canonical pattern

This section presents instances that, for various reasons, could not be included in the previously defined categories and have therefore been classified as Other in Table 20. More specifically, Section 4.2.4.1 discusses instances of object omission resulting from alternative encoding, namely, instances in which the canonical object is realized as an indirect rather than a direct object. Section 4.2.4.2 addresses instances of ambiguous valency, in which the clitic may be interpreted either as a direct object (accusative) or as an indirect object (dative). Section 4.2.4.3 focuses on issues related to the annotation scheme, while Section 4.2.4.4 presents instances involving divergences in textual reconstruction concerning the omitted object.

#### 5.2.4.1 Indirect object

Instances of construction with an indirect object, demonstrating a conscious return to classical models (Brambilla Ageno 1964: 49), is attested in *Purgatorio* XXXIII.

In this example (239) *offendere* ‘offend’ appears with a complement introduced by the preposition *a* ‘to’, although the verb *offendere* is also attested with a direct object (Example 240), where *Dio* ‘God’ is the object and it is not introduced by the preposition. These are two parallel and equally acceptable variants:

(239) *Purg.* XXXIII, v. 59  
*con bestemmia di fatto **offende a Dio***  
 ‘offends, with his blaspheming action, God’ [lit. ‘offends to God’]

(240) *Inf.* XI, vv. 83-84  
*come incontenenza men **Dio offende***  
 ‘and least offends God is incontinence’ [lit. ‘God offends’]

It is interesting to notice that Petrocchi (1994: 575) reports for the verses in Example (239), the *lectio* of other manuscripts that suggest a *lectio* without the preposition *a* (Laur, Po *offende dio*; *ofende idio* Rb, *offende idio* Urb). Whereas for Example (240) no attestation of the form with the preposition is reported (Inglese 2021: 144; Petrocchi 1994:185; Chiavacci Leonardi 2022: 201).

Interestingly, regarding the construction *offendere* + preposition, if we examine the early commentators, we can observe that up until the early 1500s, they tend to prefer the *lectio* without the preposition (see Examples (241-244)). This preference appears both in the textual choices adopted

for the *Comedy* and in their commentaries<sup>222</sup>, where they also employ the accusative form in Latin explanations:

(241) Benvenuto da Imola (1375-80):  
*nam qui offendit ecclesiam offendit Deum* (lat.)  
'for he who offends the Church offends God'

(242) Francesco da Buti (1385-95)  
*offende Iddio* (it.)  
'offends God'

(243) Johannis de Serravalle (1416-17)  
*de facto offendit Deum* (lat.)  
'in fact, he offends God'

(244) Cristoforo Landino (1481)  
*perchè offende Idio* (it.)  
*offende Idio con bestemia di facto* (it.)  
'because he offends God he offends God with blasphemy de facto.'

From the 1500s onward, and in more recent commentaries, the option with the preposition is introduced and accepted as a construction shared with Latin (see especially Scartazzini (1872-82 [2nd ed. 1900]), Sapegno (1955-57), and others.) Interestingly, Isidoro del Lungo still perceives this construction as contrastive, emphasizing its function as an indirect object.

(245) Isidoro del Lungo (1926)  
*Offende (neutro) a Dio: pecca contro a Dio* (it.)  
'Offende (used in the neuter sense) a Dio: he sins against God'. [lit. 'sins against **to** God']

Isidore, in his commentary, highlights the neuter use of the verb *offendere*, and he interprets God as an indirect object, thereby emphasizing it as the target against whom the sin is committed, paraphrasing it as *contro a Dio* ('against God').

---

<sup>222</sup> Comments have been taken from the Dartmouth project: <https://dante.dartmouth.edu>.

#### 5.2.4.2 Ambiguous valency

Another interesting case is offered by the verb *noiare* ‘to damage/disturb/hinder’, which is no longer attested in Modern Italian, having been replaced by the prefixed form *annoiare* with a slightly different meaning ‘to bore’. Interestingly, in Old Italian, both *noiare* and *annoiare* are attested, although *noiare* appears more frequently, with 105 occurrences compared to 44 for *annoiare*<sup>223</sup>. Specifically in the *Comedy*, only the form *noiare* is attested, with a total of 5 occurrences reported below in Example 246-250. This form is always used with a pronominal constructions (246-249), except for one instance (250) in which a noun phrase introduced by the preposition *a* ‘to’ clearly marks the oblique role of the nominal element involved.

(246) *Inf. XXIII, v. 15*  
*credo che lor nòì.*  
I am sure they feel deep indignation.

(247) *Purg IX, v. 87*  
*'l venir sù non vi nòì*  
your coming hither harm you not!

(248) *Par. IX, v. 35*  
*e non mi noia*  
I do not grieve

(249) *Par. XIV, v. 18*  
*al veder non vi nòì*  
to see such light and not be harmed.

(250) *Par. IX, v. 98*  
*noiando e a Sicheo e a Creusa*  
wronging both Sychaeus and Creusa [lit. ‘bother and **to** Sicheo and **to** Creusa’]

In Example (250), the dative construction is clear, as the preposition *a* ‘to’ marks *Sicheo* and *Creusa* as two oblique arguments. However, when it comes to the pronominal forms in Examples (246)-(249), some ambiguity remains.

More specifically, in Old Italian, the personal pronouns *mi* (1SG) and *vi* (2PL) in (247)-(249) can function as either accusative or dative (Salvi et al. 2010: 426), thus making the syntactic role of the pronoun unclear. The same ambiguity persists with the form *loro* (3PL) attested in (246), which can function either as a so-called “weak pronoun” in the role of a direct object (Salvi et al. 2010: 416), or as an oblique pronoun functioning as an indirect object appearing without a preposition and possibly

---

<sup>223</sup> The source has been conducted on the TLIO corpus, which is lemmatized:  
[http://tlioweb.ovi.cnr.it/\(S\(g2txtiypvdr1xm3tnewy2dkh\)\)/CatForm01.aspx](http://tlioweb.ovi.cnr.it/(S(g2txtiypvdr1xm3tnewy2dkh))/CatForm01.aspx)

classified as a weak dative pronoun (Salvi et al. 2010: 420). Once again, the function of the pronoun is not clearly identifiable.

It is true that a clear instance of a dative construction is attested, namely, the one with a prepositional phrase indicating the person who is addressed (Example 250). However, a single positive example is not sufficient to establish that the pronouns are being used as datives.

When it comes to the *Vocabolario Dantesco*<sup>224</sup>, the verb *noiare* is recorded as occurring in a dative construction, following the Provençal *enojar*, from which it is likely derived. However, there are indications that *noiare* could also be used in transitive constructions, where the affected person, the one who is annoyed, is encoded as a direct object rather than as an indirect object in the dative. The following examples from the TLIO corpus illustrate both dative and direct object constructions:

(251) Pucci, Libro, 1362 (fior.)

*E quando il Veglio voleva fare uccidere alcuno che **noiasse la sua signoria***

And whenever the Old Man wanted to have someone killed who troubled his authority.

[lit. bother the his authority]

(252) Filippo Villani, Cronica, p. 1363 (fior.)

*dell'altre furono abarrate le vie perché non **noiassono a' cittadini***

The roads from the other places were barred so that they would not trouble the citizens.

[lit. bother to citizens]

The texts are both Florentine and from the same period (1362-1363), thus demonstrating that both constructions were attested at the time. There is therefore no reason to exclude the possibility that the constructions in Examples (246)-(249) may be analyzed as involving direct objects. However, a more in-depth investigation, likely involving a comprehensive review of all construction attested in the TLIO, would be required. This line of inquiry is left for further research.

#### 5.2.4.3 Annotation issues

Some instances of object omission are retrieved due to annotation issues. More specifically, these instances involve double pronouns for which only a single dependency relation is encoded in the basic annotation layer (see Section 1.4). As previously discussed, such pronouns simultaneously fulfill two syntactic functions within the sentence (one in each of the clauses in which they appear),

---

<sup>224</sup> <https://www.treccani.it/vocabolario/noiare/>

thereby giving rise to doubling. Consequently, there may be instances in which the object relation associated with a double pronoun is not explicitly annotated, as only one of the syntactic functions it encodes is represented in the annotation. An example is provided below (253):

(253) *Purg.* XII, vv. 68-69  
non vide mei di me chi vide il vero, quant' io calcai  
Better than I saw not who saw the truth, All that I trod upon

In (253) the double pronoun *quant'* ‘all that’ performs two syntactic functions across the two clauses: it serves as the object of both *vide* ‘saw’ and *calcai* ‘trod upon’. These overlapping syntactic relations are not fully captured within the basic dependency annotation (see Section 1.4), and as a result, such cases must be manually checked and excluded.

#### 5.2.4.4 Contextual meaning and intransitive use

As explained in Section 4.2.2.2, some verbs marked as transitive by DanteSearch have been analyzed as conveying a different meaning, exhibiting an intransitive pattern. Such cases are mainly represented by the verbs *guardare* ‘to look’ and *riguardare* ‘to look again’, as illustrated in Example (74) (Section 4.2.2.2). More specifically, *guardare*, when used with an oblique complement or directional phrase, is interpreted as directing one’s gaze (Enciclopedia Dantesca 1970: 294), which supports an intransitive analysis. Similarly, *riguardare* (Enciclopedia Dantesca 1970: 927-928) can also bear an intransitive meaning, namely to re-direct one’s gaze. As mentioned before, all such instances have been treated as transitive by DanteSearch, but they warrant closer analysis to accurately capture the actual meaning of the verb. This section focuses on the verb *riguardare* to show how context is essential in determining the precise meaning of each occurrence.

Among the occurrences of finite verbs with the lemma *riguardare* that do not display an object or clausal complement, the distribution is as follows: 3 in *Inferno*; 2 in *Purgatorio*; 3 in *Paradiso*. In these cases, the verb’s interpretation alternates between a transitive (*osservare* ‘to observe’) and an intransitive (*rivolgere lo sguardo* ‘to direct one’s gaze’) meaning. When introduced by a locative adverb (Examples 254-255) or a directional oblique (Examples 256-257), the meaning is more likely to direct the gaze, therefore implying an intransitive use:

(254) *Par.* XXII, vv. 91-93  
*e se guardi 'l principio di ciascuno, poscia **riguardi là** dov' è trascorso, tu vederai del bianco fatto bruno.*

‘and if thou lookest at each one’s beginning, and then regardest whither he has run, Thou shalt behold the white changed into brown.’

(255) *Par.* XXXI, v. 67

*e se riguardi sù nel terzo giro*

‘And if thou lookest up to the third round’

(256) *Purg.* XXIX, vv. 67-69

*L' acqua imprende dal sinistro fianco (...) s' io riguardava in lei, come specchio anco.*

‘The water on my left flank was resplendent (...) E’en as a mirror, if I looked therein.’

(257) *Par.* XXXII, vv. 85-86

*Riguarda omai ne la faccia che a Cristo più si somiglia*

‘Look now into the face that unto Christ Hath most resemblance’

(258) *Purg.* VI, vv. 58-59

*Ma vedi là un' anima che, posta sola soletta, inverso noi riguarda*

‘But yonder there behold! a soul that stationed All, all alone is looking hitherward’

All the instances in Examples (255-258) can be associated with the meaning of *rivolgere lo sguardo* ‘to direct one’s gaze’ or *ricambiare lo sguardo* ‘to return a gaze’ and thus indicate a punctual and directional action. In some instances, a play on verbs of sight emerges, most notably in (255) and (258), creating a dynamic interplay. Specifically, in (255), the verses include three verbs of sight: *guardi* ‘lookest’, *riguardi* ‘regardest’ and *vederai* ‘behold’, creating a temporal progression that aligns with the topic, namely, the description of the Church’s gradual corruption. This progression of decay is reinforced by the use of three semantically related verbs of sight, which urge the poet (and therefore the reader) to closely observe this process of morale decline. The reader is thus compelled to *vederai del bianco fatto bruno* ‘thou shalt behold the white changed into brown’ (v. 93), a striking image of transformation and degradation.

Similarly, in (256), the author creates a layered interplay by combining the two verbs of sight, *vedi* ‘(you) yonder’ and *riguarda* ‘(it) is looking’. The scene depicts the encounter between the two poets, Dante and Virgil, and another poet, Sordello, who is portrayed as a solitary and secluded soul (*un’anima che, posta sola soletta* ‘a soul that stationed All, all alone’ vv. 58-59).

Interestingly, the scene revolves around an exchange of gazes between the poets. The adversative opening conjunction *ma* ‘but’ marks a break from the preceding discourse, drawing attention to a new narrative element: the encounter with a soul, who is soon revealed to be Sordello. The soul, introduced as the rhematic element in the main clause, conveying new information within the discourse, is immediately shifted to the thematic position (that which conveys given information) in the embedded relative clause, which now presents as rhematic content the soul’s returning gaze toward the two poets. This shift expresses a reciprocal exchange of gazes and redirects the focus of the scene onto the two poets, who are now the ones being observed by the soul.

When the verb *riguardare* is accompanied by an adverb, the latter can help clarify its meaning, as shown in Examples (259-260). In such cases, the adverbs *fiso* ‘fixedly’ and *ben* ‘well’ are particularly informative (Enciclopedia Dantesca 1970: 927). Specifically, in (259), the adverb *fiso* is semantically incompatible with a verb of movement. This constraint thus points toward an interpretation of *riguardare* as *osservare* (‘to observe’).

(259) *Inf. IV, vv. 5-6*

*e fiso riguardai per conoscer lo loco dov’ io fossi.*

‘and steadfastly I gazed, To recognise the place wherein I was.’

Similarly, in Example (260), the adverb *ben* ‘well’, which can be paraphrased as ‘carefully’, functions as an intensifier and is compatible with the interpretation *guardare intensamente* ‘to look intensely’ (Enciclopedia Dantesca, 1970: 927).

(260) *Inf. XIII, vv. 20-21*

*Però riguarda ben, sì vederai cose che torrien fede al mio sermone.*

‘Therefore look well around, and thou shalt see Things that will credence give unto my speech.’

Finally, when neither oblique complements nor adverbs are present, the interpretation becomes more difficult to determine, as shown in Example (261):

(261) *Inf. III, vv. 51-52*

*Non ragioniamo di lor, ma guarda e passa». E io, che riguardai, vidi una 'nsegna*

‘Let us not speak of them, but look, and pass.’ And I, who **looked again**, beheld a banner’

In Example (261), the meaning of *riguardai* ('looked again') is subject to debate. In the *Enciclopedia Dantesca* (Enciclopedia Dantesca 1970: 927), the verb is interpreted as *rivolgere lo sguardo* 'to direct one's gaze', an alternative reading is proposed by Chimenz, who prefers to interpret the verb as *guardare attentamente* 'to look attentively'. In this specific case, there are no definitive cues that clearly support one interpretation over the other, thereby underscoring the importance of the reader's interpretive sensibility in shaping meaning.

To conclude, we noted that in some cases, such as in (261), it is not easy to reconstruct the exact meaning of each verb without resorting to interpretation with the risk of conjecture, potentially leading to different readings of the verses, which is also part of the work of the lector especially given the narrative nature of the text. However, certain clusters can help guide the interpretation, such as the presence of adverbial modifiers and oblique complements (as in Examples 254-260). Further analysis of the usage of this verb, through the collection of additional occurrences in other texts from the same period, may be useful in assessing whether other syntactic cues can aid in determining its interpretation.<sup>225</sup> Nonetheless, it is important to remember that, especially in this type of narrative and poetic text, part of the interpretation is deliberately left to the reader's imagination.

For the purposes of the present study, instance of *riguardare* interpreted as *rivolgere lo sguardo* 'to direct one's gaze' have been considered intransitive occurrences and thus classified as non-canonical patterns, in order to distinguish them from uses of *riguardare* meaning *osservare* ('to observe'), which have been classified as transitive and included in both the data collection and manual analysis. With regard to the ambiguous case, Example (261) has been analyzed as an instance of *rivolgere lo sguardo* and was therefore excluded from the final analysis.

#### 5.2.4.5 Divergences in Textual Reconstruction

It is interesting to note that in some cases the object pronoun is introduced by certain editors in the *lectio* of the text. One example can be found in *Inferno* XV, v. 90:

(262) *Inf.* XV, v. 90:

*E serbolo a chiosar con altro testo / a donna che **saprà**, s'a lei arrivo.* (ed. Petrocchi)

*E serbolo a chiosar con altro testo / a donna che **'l saprà**, s'a lei arrivo.* (ed. Inglese)

'I keep it with another text, for comment by one who'll understand, if I may reach her.'

---

<sup>225</sup> A similar effort is currently underway in the TLIO dictionary project; however, as the compilation has not yet reached the letter R, the entry for the verb *riguardare* is still unavailable.

Example (262) is taken from a response by Dante to Brunetto Latini, his teacher, who is condemned to Hell, and who had prophesied the poet's future glory (*La tua fortuna tanto onor ti serba* 'Your fortune holds in store such honor for you', v. 70). In his reply, Dante says he will retain in his memory Brunetto's words, *ciò che narrate di mio Corso* 'what you have told me of my course' (v. 89), and will preserve it so that it may be glossed, continuing the metaphor of the book of life, by a woman capable of doing so. The woman in question is Beatrice.

The verse in (262) is reported twice. The first version corresponds to the Petrocchi edition, which is also the one used in the Italian-Old treebank (see Chapter 2). The second comes from the edition by Inglese (2007). The difference between the two *lectiones* lies precisely in the presence of an object pronoun with the verb *saprà* 'will understand': in Inglese's edition, an object pronoun is introduced, whereas in Petrocchi's it is omitted. Both editors note that some manuscripts include the object pronoun (Triv, Ham, Eg, Laur, Rb; see Inglese 2007: 183), while others do not transmit it (beta, and the text of c and b; see Petrocchi 1994: 253). Inglese, who supports the presence of the object pronoun, justifies its omission in the manuscripts that lack it as the result of assimilation ("non esclude un'assimilazione di *l.*", Inglese 2007: 183), and therefore favors the *lectio* with the explicit pronoun. Petrocchi, on the other hand, argues that the use of *(i)l* trivializes (*banalizza* 'Petrocchi 1994: 253') the expression, and thus prefers the version without it.

Similarly to this instance, also in the following Example (263), Inglese (2007: 234) differs from Petrocchi (1994: 289):

(263) *Par. XVII, v. 116*

*ho io appreso quel che s'io **ridico***, (ed. Petrocchi)

*ho io appreso quel che, s'io **'l ridico***, (ed. Inglese)

I learned that which, if I retell it

As for Example (262), also in (263) Inglese preserves the version with the explicit object pronoun, whereas Petrocchi prefers the version in which the clitic is omitted.

#### 5.2.4.6 Use of the object pronoun as expletive:

Although the present study focuses on object omission, attention should also be drawn to a reverse instance, namely, a case in which the pronoun is present in the text, even though it functions pleonastically (Petrocchi 1994, 253–54; Chiavacci Leonardi 2012, 272; Inglese 2007, 183), referring

to something general rather than to a specific element in the text. An example is provided by *Inferno* XV, v. 99, reported in Example (264):

- (264) *Inf.* XV v. 99:  
*poi disse: «Bene ascolta chi la nota».*  
He listeneth well who noteth it.

In this instance, the object pronoun *la* ‘her’ (translated with *it*) does not have a specific referent in the text, but rather functions similarly to idiomatic expressions such as *chi la fa, la aspetti* ‘what goes around comes around’ (lit. ‘who does it, waits for it’), where *la* stands for a general, unspecified thing (*cosa*).

#### 5.2.5 Omission of the reflexive clitic: neuter and *uso assoluto*

In some cases, the omitted element is not a canonical object but rather the reflexive particle *si* (Brambilla Ageno 1990: 228-229; Brambilla Ageno 1964: 132-136; 175-247). These instances of *si* omission differ from those discussed in Section 5.2.3.4.2, where the omission is triggered by the presence of an antecedent. In the cases under consideration here, no textual antecedent can be identified. This omission of the reflexive clitic has been analyzed for anticausative in Old Florentine by Cennamo (2012) who noticed that it is frequently attested especially in compounded tense (Cennamo 2012: 417), reflecting “the gradual reconstruction of the tense-aspectual and voice systems following the loss of the grammatical dimension of voice in the passage to Romance” (Cennamo 2012: 405-406). A detailed analysis of such constructions will not be undertaken here, as they have already been addressed in previous studies (Cennamo 1993; Cennamo & Jezek 2011; Pescarini 2015; among others). However, examples from the Comedy of reflexive omission will be provided that are consistent with the retrieval criteria adopted in this study.

An instance of reflexive omission is illustrated in Example (265) with the verb *rinovellare*. The verb *rinovellare* (Brambilla Ageno 1964: 106) is classified among denominal verbs derived from adjectives, which exhibit alternation between a neutrum use and transitive use (Brambilla Ageno 1964: 97). Specifically, for *rinovellare*, attested uses in the *Comedy* include neutrum instances such as in example (265), reflexive constructions as in example (265), as well as transitive cases with an expressed object, as in example (266).

- (265) *Par.* XXVI, vv. 127-129

*ché nullo effetto mai ragionabile, per lo piacere uman che **rinovella** seguendo il cielo, sempre fu durabile.*

‘For nevermore result of reasoning (Because of human pleasure that doth **change**, Obedient to the heavens) was durable.’

Example (265) is taken from *Canto XXVI* of *Paradiso*. Adam, the first man on Earth, responds to Dante’s questions, one of which concerns the language he spoke, a language described as mutable. In the verses cited in (265), the mutability of language is explained as the result of its being a human product, shaped by people who tend to introduce changes over time. The verb *rinovella* ‘change’ is used here in its neuter form, without an object or reflexive clitic. As mentioned before, however, the neuter use is not the only attested for the verb *rinovellare* in the Comedy, for which also a reflexive (266) and transitive uses (267) are attested.

(266) *Purg.* XXXII, vv. 52-56

*Come le nostre piante (...) turgide fansi, e poi **si rinovella** di suo color ciascuna*

‘In the same manner as our trees (...) Begin to swell, and then **renew** themselves, Each one with its own colour,’

Example (266) contains a periphrasis introducing the arrival of spring, marked by the position of the sun in the constellation of Aries. The periphrasis describes the blossoming of plants, typical of the spring season, and their changing colours. To express such changing the verb *rinovella* ‘renew’ is employed in its reflexive form. By contrast, in Example (267), the same verb is used in its transitive form:

(267) *Inf.* XXXIII, vv. 4-5

*Tu vuo' ch'io **rinovelli** disperato **dolor***

‘You want me to **renew** despairing **pain**’

The verses mark the beginning of Ugolino’s direct speech, which he iconically opens by emphasizing the pain that the memory evokes. Here, the verb *rinovellare* (*rinovelli* ‘renew’) is used with a direct object, namely *dolor* ‘pain’. Among the verbs that displays such optional omission (neuter use), also others are attested such as the one reported below in Example (268) and (269):

(268) *Par.* XII, vv. 86-87

*tal che si mise a circuir la vigna che tosto imbianca*

‘That he began to go about the vineyard, Which **fadeeth** soon’

Example (268) provides another instance of a denominal verb derived from an adjective (Brambilla Ageno 1964: 97), for which a neuter use is attested. The verb in question is *imbiancare* ‘to turn white’. In this verse, it describes the withering of the vineyard, used as a metaphor of the Church. In this case, the verb is used without the reflexive clitic *si*. However, the same verb *imbiancare* is also attested with the clitic, conveying the same meaning as in (268), ‘to turn white’ in *Purgatorio IX*, Example (269):

(269) *Purg. IX, vv. 1-2*

*La concubina di Titone antico già s'imbiancava al balco d' oriente*

‘The concubine of old Tithonus now **Gleamed white** upon the eastern balcony’

In Example (269) the author describes the dawn using a periphrasis: Aurora, the companion of Tithonus, turns white as she leans out from the balcony of the East.<sup>226</sup> Finally, a noteworthy instance of reflexive clitic omission is found in Example (270):

(270) *Inf. II, v. 67*

*Or movi,*

Go now,

In this example, the clitic, expected here in the second person singular form, is omitted with the verb *muovere* ‘to move’ (Brambilla Ageno 1964: 73-74). Such usage would be ungrammatical in Modern Italian, where the presence of the reflexive clitic is obligatory. An example of the same verb used with the required clitic in Modern Italian is provided in (271):

(271) PoSTWITA-3560

*Coglione muoviti :’3 ho fatto le corse per te*

‘Come on, idiot :’3 I rushed around for you!’

---

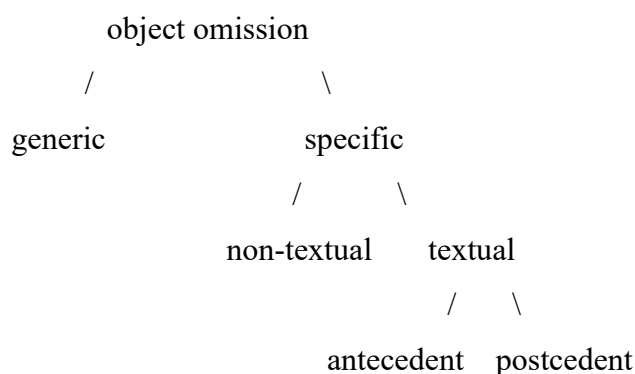
<sup>226</sup> It should be noted that the meaning of ‘turning white’ is also supported by some manuscript attestations (Ham), which report the verb *imbiaccare* in the sense of “applying white lead to the face” (Petrocchi, 1994: 137; Inglese, 2007: 124).

However, the form without the clitic is not the only one attested, as shown by other manuscripts, for instance, manuscript Pa reports *or moviti* ‘move yourself’ (Petrocchi 1994: 29). The clitic form is also attested in other verses of the *Comedy*, such as in Example (272):

(272) *Inf.* XVII, v. 97  
*Gerion, **moviti** omai*  
 ‘Now, Geryon, **move on**’

### 5.2.6 Summary

Section 5.2 presents a data-driven analysis of object omission in the *Divine Comedy*. Following the automatic extraction of transitive verbs lacking an object or complement clause (as described in Section 4.2), a manual review and analysis of the retrieved instances was conducted. The analysis draws on previous literature on omitted objects in English (Fillmore 1986), and more specifically in Italian (Lo Duca 2000; Cennamo 2015, 2017), which distinguishes between generic and specific objects. The distinction adopted in this analysis follows the analytical flow outlined in Image 170:



**Image. 170.** Steps for object omission analysis

Once the object omission has been identified, a distinction is made between generic and specific referent. If the referent of the omitted object is specific, a further distinction is drawn between textual and non-textual referents. Finally, for textual referents, it is indicated whether the referent is an antecedent or a postcedent. Information about the verb’s lemma and its syntactic relation in the tree is automatically retrieved from the treebank. Through manual analysis of such omissions, several observations emerged.

The first observation is that generic and specific but non-textual objects display points of overlap that could justify grouping them together, especially for NLP tasks, as will be explained shortly. More specifically, while it is sometimes straightforward to reconstruct an omitted object by recovering its referent through world knowledge (see Example 182), the precise semantic field cannot always be uniquely determined, as it was shown in Example (185), thus resulting in a continuum of precision in referent attribution, which must be assessed on a case-by-case basis. Moreover, there may be instances in which the ambiguity in retrieving the omitted object, or even in selecting the correct meaning of the verb, as in Example (188), is deliberate and intentional, and is closely tied to the creativity of language, which is particularly heightened in poetic texts such as the one under analysis. On the other hand, even the generalization to “something” (the test proposed by Fillmore (1986) to classify generic omission) may be undermined by associative processes that lead the reader to reconstruct a context. This is because reconstructing something that is not present in the text, not even via a textual antecedent, is inherently risky and leaves even more room for subjective interpretation. For instance, in example (180), a more fine-grained interpretation, rather than the generic “something”, could have been “perceptions of life”, thus shifting this example into the category of specific referents without textual anchor. All these considerations highlight the difficulty of reconstructing a possible omitted object, a process in which, when no textual referent is provided, the only true criterion is the reader’s interpretive judgment and background knowledge, thereby opening the door to multiple possible readings. Given the complexity and subjectivity of this work for a human annotator, it is easy to see how it could present significant challenges for NLP applications as well.

When it comes to omitted objects with a specific textual referent, the degree of arbitrariness is mitigated by the text itself. Within this category, certain clusters of phenomena emerged. Among the verbs lacking a direct object or complement clause but allowing for the retrieval of an antecedent or postcedent, the following types of instances were analyzed: verbs accompanied by the deictic *così* ‘so’; comparative clauses in which the object is implicit and recoverable from the main term of comparison; instances of reduplication in which only one of the repeated verbs bears the object; instances of shared dependents occurring in coordinated structures; instances attributed to so-called lexical properties, where no clear syntactic cues account for the omission, which instead appears to be determined by the lexical characteristics of the verb itself; and finally, instances of non-sentential referents, in which the omitted object is not easily retrievable from the immediate textual context.

Some instances retrieved by the query do not fit into the categories outlined above and have therefore been grouped under non-canonical patterns. Among these, specific cases of indirect object encoding were identified, in which Old Italian, due to Latin influence (Brambilla Ageno 1964), employs alternative strategies for object encoding. Additionally, ambiguous instances were found

where the valency pattern requires further investigation. Other instances of apparent object omission were in fact due to annotation issues and thus should not be counted as genuine omissions. Moreover, variation in *lectio* were noted, illustrating that the alternation in clitic usage was still in progress at the time (Luraghi 1998). Finally, the final cluster concerns the omission of the reflexive clitic.

## 6) Summary and conclusion

This dissertation has presented the creation of a syntactically dependency-annotated resource of Dante Alighieri’s *Divine Comedy*, named the *Italian-Old treebank*<sup>227</sup>. The treebank is integrated into the Universal Dependencies (UD) framework (de Marneffe et al. 2021), which currently represents the state of the art for dependency-annotated corpora and covers more than 150 languages<sup>228</sup>, including several historical ones. UD is widely regarded as the state of the art in terms of typological diversity and data availability for NLP tools. Nowadays, the growing use of NLP applications, and the possibility of conducting quantitative analyses that are targeted, universally comparable, and openly accessible, exerts pressure on linguistic data as well. This was the motivation for aligning Old Italian with this direction, with the hope that more resources will become accessible and more data made available for sharing and reproducibility, thereby contributing to growth and advancement in research. The decision to focus on a historical language such as Old Italian stems not only from personal interest, but also from the conviction that such resources can enable studies that enrich qualitative research, supporting them with quantitative investigations, and open new perspectives that shed light on the texts themselves.

With regard to the choice of text, the *Divine Comedy* is a poem written in Old Italian (more specifically in 13th-century Florentine) and constitutes a foundational text for both the Italian language and culture (Manni 2013). Its importance was therefore one of the main reasons for selecting this work as the first resource to document Old Italian within the UD treebank landscape. Another important reason for this choice was the availability of a pre-existing corpus of Dante’s works, including the *Divine Comedy* itself, DanteSearch (Tavoni 2012), already annotated, though according to a formalism different from that of UD. The decision to reuse this resource aligns with the FAIR principles<sup>229</sup> of making scientific data standardized and reusable. The reuse of part of the original annotation from DanteSearch was not exempt from modifications and adaptations, as described in Chapter 2. The original tagset, particularly the morphological one, was converted to conform to UD guidelines, followed by a manual dependency-based syntactic annotation. The syntactic annotation was carried out entirely from scratch for the first cantica, the *Inferno* (for a total of 41,369 tokens and 1,228 sentences). Once the annotation of the *Inferno* was completed, thereby providing a pool of gold data for training in-domain models, the second Cantica, *Purgatorio*, was pre-parsed using a model

---

<sup>227</sup> The treebank is available at [https://github.com/UniversalDependencies/UD\\_Italian-Old/blob/master/README.md](https://github.com/UniversalDependencies/UD_Italian-Old/blob/master/README.md)

<sup>228</sup> <https://universaldependencies.org>

<sup>229</sup> <https://www.go-fair.org/fair-principles/>

trained with UDPipe1 (Straka et al., 2016). The *Purgatorio* was then manually revised (encompassing 41,270 token and 1,174 sentences) and once completed it served in turn as training data for a Stanza model (Qi et al., 2020) including both *Inferno* and *Purgatorio* gold annotations, which was used to pre-parse the *Paradiso* (consisting of 39,385 token and 1,017 sentences). The choice of employing a Stanza model, as well as the decision of not merging Old Italian data with Modern Italian UD treebanks, followed from a series of experiments described in Chapter 2. Chapter 2 also discussed specific annotation choices related to linguistic phenomena and outlined possible directions for future research.

Once the treebank was completed, *Italian-Old* was used to record the “sound of silence” in the *Divine Comedy*, to analyze ellipsis, a phenomenon defined as the absence of a form in the presence of its meaning (Merchant 2018). Ellipsis has been widely discussed and debated in the literature, particularly from a syntactic perspective. The main divergences in its treatment arise from the theoretical framework adopted. When ellipsis is conceived as an empty node in the syntactic structure, constituency-based frameworks have produced a vast body of research on the topic, largely due to the possibility and widespread acceptance of empty nodes in syntactic representations. By contrast, dependency-based accounts have been less extensive in their discussion of ellipsis, often focusing on a specific type, gapping, given the tendency in this framework to avoid the insertion of empty nodes in syntactic structures. Chapter 3 is devoted to a general overview of ellipsis. The aim is not to provide an exhaustive survey of the literature, which would be nearly impossible given the volume of work on the topic, especially within constituency frameworks, but rather to outline a general scheme highlighting the key concepts involved in the study of ellipsis (Section 3.1), the main frameworks and theories applied to its analysis (Section 3.1), and finally a classification of the types of ellipsis reported in the literature (Section 3.2). Following this general and theoretical introduction, special attention is devoted to ellipsis in diachrony (Section 3.3), with a focus on studies available for Old Italian. Finally, Section 3.4 also presents the scheme adopted within UD to handle ellipsis, raising a number of critical issues that are explored in greater detail over the course of the dissertation.

Chapter 4 is dedicated to ellipsis and NLP and addresses the problematic question of how to tackle ellipsis, namely, “data in absence”. The aim of this section is to outline a dual workflow: one theoretical and one more practical, directly applicable to the data available in UD, which represent a valuable resource for large-scale and cross-linguistic analyses of the phenomenon. The framework adopted here is specifically structured around one type of ellipsis, namely predicate ellipsis. In the theoretical section (Section 4.1), the process of investigating ellipsis is described. Ellipsis is, by definition, a phenomenon that is not explicitly retrievable in the text or data. The process is therefore divided into three stages: detection, identification, and reconstruction. For each step, practical

solutions to explicitly annotate predicate ellipsis in data are proposed, together with a discussion of the critical issues involved. These theoretical and practical workflows were studied and tested on Old Italian data, in particular on the *Italian-Old* treebank, where specific solutions to challenging issues were proposed (Section 4.2). To date, the same methodology has also been applied in a pilot study on Latin (Albertelli et al. 2025), revealing its overall compatibility while also raising other language-specific requirements. The long-term goal is to extend this approach, necessarily modelled on Italian-Old data in this study, to a wider range of languages, building also on the queries provided in Chapter 4 for retrieving predicate ellipses designed within the UD formalism, which is thus compatible, though adjustable, with the other languages attested in UD as well. The aim is to refine the approach and foster discussion on possible improvements that could make the proposal genuinely cross-linguistic and universally applicable. Although the general focus is on predicate ellipsis, this study has also taken into account object omission, an interesting phenomenon in Old Italian, already investigated by Luraghi (1998) and Marengo (2013), for which additional data were collected in order to test the claims against the available evidence. In the case of Old Italian, object omission could be retrieved thanks to transitivity marking obtained through the conversion of the original DanteSearch corpus. However, such information is not typically encoded in UD, which makes the procedure difficult to reproduce for other languages and therefore specific to Old Italian.

Finally, Chapter 5 presents the analysis of predicate ellipsis and object omission in the *Divine Comedy*. The analysis is distinctive in being data-driven. This means that, despite relying on a dependency-based treebank, the extraction was not constrained by a framework-specific classification of elliptical constructions. Namely, not only gapping structures, explicitly annotated as elliptical in UD, were extracted. Rather, the procedure was designed to be as agnostic as possible, within the limits imposed by the annotation criteria. More specifically, predicate ellipsis was identified in sentences where the predicate, understood, following Serianni (1991), either as a verbal predicate (consisting of a verb) or as a nominal predicate (consisting of a nominal accompanied by a copular element, which in Italian is the verb *essere* ‘to be’), is absent. Instances of object omission were also retrieved on the basis of transitivity annotation and the absence of a syntactic object. The data were then manually clustered into groups according to their properties.

For predicate ellipsis (Section 5.1), a quantitative description was provided, based on the syntactic position occupied by predicate ellipsis. Each case was then analyzed individually and classified, revealing a wide variety of constructions in which the predicate is omitted and a continuum of structures. At one end are cases where the verb is not only omitted but genuinely absent, being unnecessary for predication itself (as in presentative and locative constructions), which in turn raises the question of how the predicate itself should be defined. At the other end are more explicit instances

of verb omission, such as gapping, where the absence is concretely realized as a true gap in the syntactic structure (this omission has direct repercussions on the structure itself and is therefore addressed in UD through the `orphan` relation). Examples from each *Cantica* are provided for every type. The quantitative analysis based on syntactic position was then integrated with the manual qualitative investigation, focusing on the most frequent syntactic contexts (main clause, coordination, reported speech, and parataxis). This provided a more fine-grained account of the distribution of predicate ellipsis across syntactic clauses and showed that many types of predicate omission are in fact peculiar to the textual genre rather than being determined by syntactic position alone, thus aligning with observations already made in the literature, which emphasize the importance of not limiting the analysis of ellipsis to syntax alone (Kuno 1976; Ramat 1982; Kehler 2000, 2002). A specific example is given by the high number of elliptical constructions in main clauses, which, after the quantitative–qualitative filtering, turned out to be determined by a specific type of structure characteristic of the narrative system, here referred to as elliptical narrative framing of reported speech. Therefore, the analysis conducted in this work once again raises the issue of reducing ellipsis to a purely syntactic, framework-dependent concept tied to the notion of empty nodes in syntactic structures. Such a perspective risks polarizing the study of ellipsis, creating discrepancies and disagreements even in the very definition of the phenomenon. As mentioned above, while constituency frameworks tend to propose a broad catalogue of ellipsis types, dependency frameworks are more cautious, often limiting their attention to gapping, where the syntactic structure is directly affected, while underrepresenting other types that nevertheless emerge from the data. This divergence makes dialogue and comparison between approaches particularly challenging and, to some extent, not comparable at all, an undesirable situation for scientific progress aimed at analyzing a unique linguistic phenomenon. The proposal advanced here, when dealing with ellipsis, is therefore to shift the focus of analysis away from ellipsis as mere absence and from the debate on the presence of empty nodes in syntactic structures (which, however, remains relevant for NLP tasks, as addressed in Section 4), and instead to treat absence as a starting point for investigating what remains and how the utterance is structured by virtue of that absence. This, in turn, opens the way to a broader discussion of the very notion of predicate and extends the analysis and classification of ellipsis to include the role of textual genre and stylistic factors, thereby situating the study of ellipsis within the dynamics of textual typology and its function for communicative purposes.

Concerning the analysis of object omission (Section 5.2), a systematic classification was provided based on the presence of antecedents, distinguishing between generic and specific ones, and, in the latter case, further differentiating between textual and non-textual antecedents. This classification builds on the studies of Fillmore (1986) and was subsequently applied to Italian by Lo

Duca (2000) and Cennamo (2017). The analysis based on this classification revealed a complex distribution of cases, particularly for specific objects with a textual referent, for which the more fine-grained analysis identified several clusters, ranging from stylistic omissions (such as reduplications, the use of the deictic *così*, and narrative functions) to cognitively driven peculiarities, such as comparatives, that deserve more in depth and cross-linguistic studies, such as omission in comparatives, or omissions triggered by the lexical properties of specific verbs. Another interesting insight is offered by the analysis of shared dependents (Section 5.2.3.4), which revealed their involvement in the optionality of pronominal clitic objects in Old Italian, a phenomenon studied by Luraghi (1999) and here further confirmed, also by textual editions (Inglese vs Petrocchi).

## References

- Aelbrecht, Lobke, and William Harwood. 2018. Predicate Ellipsis. in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. pp. 504–525.
- Agnoloni Tommaso, Roberto Bartolini, Francesca Frontini, Simonetta Montemagni, Carlo Marchetti, Valeria Quochi, Manuela Ruisi, and Giulia Venturi. 2022. Making Italian Parliamentary Records Machine-Actionable: The Construction of the ParlaMint-IT corpus. In *Proceedings of the Workshop ParlaCLARIN III within the 13th Language Resources and Evaluation Conference*, pages 117–124, Marseille, France. European Language Resources Association.
- Albertelli, Lisa Sophie, Lorenzo, Augello, Giulia, Calvi, Annachiara, Clementelli, Federica, Iurescia, Claudia, Corbetta. 2025. Ellipsis in Enhanced Dependencies: A Case Study on Latin. *CLiC-it 2025*. Forthcoming.
- Alighieri, Dante. 1972. *La Divina Commedia, riveduta nel testo e commentata* G. A. Scartazzini. Leipzig, F.A. Brockhaus, 1874-82, Vol. III; ristampa nast. Bologna. Forni.
- Alighieri, Dante. 1955-1957-1968. *La Divina Commedia*. A cura di Nicolò Sapegno, Firenze, La nuova Italia.
- Alighieri, Dante. 1994. [1966-1968]. *La Commedia secondo l'antica vulgata*. a cura di Petrocchi Giorgio. Milano. Mondadori. Vol. I, Introduzione. Vol. II, Inferno. Vol. III, Purgatorio. Vol. IV, Paradiso. Edizione nazionale delle opere di Dante, a cura della Soc. Dantesca Italiana, Seconda ristampa riveduta, Firenze, Le Lettere.
- Alighieri, Dante. 1996. [1995]. *La Commedia*. Nuovo testo critico secondo i più antichi manoscritti fiorentini. a cura di Lanza Antonio, Anzio, De Rubeis.
- Alighieri, Dante. 2001. *Comedia*. Edizione critica per cura di Sanguineti Federico. Firenze. Edizioni del Galluzzo. ID. Id. Appendice bibliografica 1988-2000, ivi, Id. 2005.
- Alighieri, Dante. 2016. *Epistole; Egloge; Questio de aqua et terra*. Edizione critica per cura di Baglio Marco, Azzetta Luca, Petoletti Marco, Rinaldi Michele con introduzione di Mazzucchi Andrea. Roma: Salerno.
- Alighieri, Dante. 2021 [2007]. *Commedia*. Revisione del testo e commento di Inglese Giorgio. Vol. I Inferno. Vol. II Purgatorio. Vol. III Paradiso. Roma. Carocci.
- Alighieri, Dante. 2022 [1999] *Commedia*. Con il commento di Chiavacci Leonardi Anna Maria. Vol. I Inferno. Vol. II Purgatorio. Vol. III Paradiso. Mondadori.
- Alighieri, Dante. 2022. *La Divina Commedia. Inferno*. A cura di Malato, Enrico. Salerno Editrice.
- Alighieri, Dante. 2022. *Commedia*. Edizione critica per cura di Elisabetta Tonello e Paolo Trovato, Padova, [libreriauniversitaria.it](http://libreriauniversitaria.it) Edizioni, 2022.
- Allerton, David J. 1975. Deletion and proform reduction. *Journal of linguistics* 11.2. 213-237.
- AnderBois, Scott. 2010. Sluicing as anaphora to issues. In: *Semantics and linguistic theory*. p. 451-470.

- Anthony, Laurence. 2013. A critical look at software tools in corpus linguistics. *Linguistic research* 30(2): 141-161.
- Ascoli, Graziadio Isaia. 1882–1885. *L'Italia dialettale*. *Archivio glottologico italiano*, VIII:98–128. Available at <https://archive.org/details/archivioglottolo08fireuoft/page/n5/mode/2up>.
- Auerbach, Erich. 2007. *Studi su Dante*. Universale economica Feltrinelli. Editore Milano: Feltrinelli.
- Baldelli, Ignazio. 1965. Sulla teoria linguistica di Dante. *Cultura e scuola*, n. 13–14. 705–713.
- Baldelli, Ignazio. 1996. *Dante e la lingua italiana*. Firenze: Accademia della Crusca.
- Battaglia, Salvatore. 1966. *Grande dizionario della lingua italiana*. Dizionari Unione tipografico-editrice torinese.
- Barros, Matthew, and Luis Vicente. 2016. A remnant condition for ellipsis. *Proceedings of WCCFL*. Vol. 33. Somerville, MA: Cascadilla Proceedings Project. Pages 57-66
- Belletti, Adriana. 2001 [1991]. Funzioni delle frasi subordinate: Frasi avverbiali: comparative. In Renzi, Lorenzo, Salvi, Giampaolo, Cardinaletti, Anna (eds.), *Grande Grammatica Italiana di Consultazione*. Vol. II. I sintagmi verbale, aggettivale, avverbiale. La subordinazione. Bologna: il Mulino. 832-853.
- Belletti, Adriana. 2010. La comparazione. In Salvi, Giampaolo, and Lorenzo Renzi, (eds.) *Grammatica dell'italiano antico*. Vol. II. Bologna: il Mulino. 1135-1146.
- Bellomo, Saverio. 2020. *Filologia e critica dantesca*. (Nuova ed. riveduta e ampliata). Scholè.
- Beltrami, Pietro G. 2003. Il Tesoro della Lingua Italiana delle Origini (tlio). *Italia linguistica anno Mille, Italia linguistica anno Duemila: atti del XXXIV Congresso internazionale di studi della Società di linguistica italiana (SLI)*, Firenze 19-21 ottobre 2000. (Pubblicazioni della Società linguistica italiana; 45. Bulzoni.
- Benincà, Paola, Cecilia Poletto 2010. L'ordine delle parole e la struttura della frase parte. In Salvi, Giampaolo, and Lorenzo Renzi, (eds.) *Grammatica dell'italiano antico*. Vol. 1. Bologna: il Mulino. 27-78.
- Benveniste Émile. 1971. *Problemi di linguistica generale*. Milano : Il Saggiatore.
- Berruto, Gaetano. 1993. Le varietà del repertorio. Sobrero Alberto (édit.), *L'italiano contemporaneo: la variazione e gli usi*, Bari, Laterza, II: 3-36.
- Bhatt, Rajesh, Takahashi, Shoichi. 2011. Reduced and unreduced phrasal comparatives. *Natural Language & Linguistic Theory* 29, 581–620. <https://doi.org/10.1007/s11049-011-9137-1>
- Bloomfield, Leonard. 1965 [1933]. *Language history: from Language* (1933 edition). edited by Harry Hayer. - New York. Holt, Rineart and Winston.
- Bonial, Claire, Julia Bonn, Kathryn Conger, Jena D. Hwang, and Martha Palmer. 2014. PropBank: Semantics of New Predicate Types. In *Proceedings of the Ninth International Conference on*

Language Resources and Evaluation (LREC'14), pages 3013–3019, Reykjavik, Iceland. European Language Resources Association (ELRA).

Bonsignori, Veronica. 2018. Introduction to ellipsis in English texts and discourse: A descriptive study. Pisa University Press.

Bosco, Cristina, Manuela Sanguinetti, and Leonardo Lesmo. 2012. The parallel-TUT: a multilingual and multiformat treebank. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), pages 1932–1938, Istanbul, Turkey. European Language Resources Association (ELRA).

Bosco, Cristina, Simonetta Montemagni, Maria Simi. 2013. Converting Italian Treebanks: Towards an Italian Stanford Dependency Treebank. In Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse (LAW VII & ID at ACL-2013), Sofia, Bulgaria, August 8-9, pp. 61-69.

Bosco, Cristina, Felice Dell'Orletta, Simonetta Montemagni, Manuela Sanguinetti, Maria Simi. 2014. The EVALITA 2014 dependency parsing task. Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014: 9-11 December 2014, Pisa. Pisa University Press.

Bouma, Gosse, Gertjan Van Noord, and Robert Malouf. 2001. Alpino: Wide-coverage computational analysis of Dutch. Computational linguistics in the Netherlands 2000. Brill. 45-59.

Brambilla Ageno, Franca. 1964. Il verbo nell'italiano antico: ricerche di sintassi. Milano; Napoli. R. Ricciardi.

Brambilla Ageno, Franca. 1990. Studi danteschi. Padova: Antenore.

Bresnan, Joan W. 1973. Syntax of the comparative clause construction in English. Linguistic inquiry 4.3. 275-343.

Bresnan, Joan. 2001. Lexical-Functional Syntax. Oxford: Blackwell.

Buchholz, Sabine and Erwin Marsi. 2006. CoNLL-X Shared Task on Multilingual Dependency Parsing. In Lluís Màrquez and Dan Klein, editors, Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X), pages 149–164, New York City, NJ, USA, June. Association for Computational Linguistics (ACL).

Čavar, Damir, Ludovic Mompelat, Muhammad Abdo. 2024a. 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. St. Julian's, Malta. Association for Computational Linguistics. 46–54.

Čavar, Damir, Zoran Tiganj, Ludovic Veta Mompelat, and Billy Dickson. 2024b. Computing ellipsis constructions: Comparing classical NLP and LLM approaches. In Proceedings of the Society for Computation in Linguistics 2024b. 217–226.

Castellani, Arrigo. 2000. Grammatica storica della lingua italiana . Vol. I. Bologna : Il Mulino.

Cecchini, Flavio Massimiliano, Sprugnoli, Rachele, Moretti, Giovanni, and Passarotti, Marco. 2020. UDante: First Steps towards the Universal Dependencies treebank of Dante's Latin Works. In Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it 2020. 66-72

Cennamo, Michela. 1993. The reanalysis of reflexives: A diachronic perspective. Liguori.

Cennamo, Michela, Elisabetta Jezek. 2011. The anticausative alternation in Italian. Luoghi della traduzione: le interfacce: atti del XLIII Congresso internazionale di studi della Società di linguistica Italiana (SLI), Verona, 24-26 settembre 2009. (Pubblicazioni della Società di linguistica italiana; 54). Bulzoni.

Cennamo, Michela. 2012. Aspectual constraints on the (anti) causative alternation in Old Italian. Transactions of the Philological Society 110.3: 394-421.

Cennamo, Michela. 2015. Valency patterns in Italian. Introducing the Framework, and Case Studies from Africa and Eurasia 1. 417-481.

Cennamo, Michela. 2017. Object omission and the semantics of predicates in Italian in a comparative perspective. Contrastive studies in verbal valency. John Benjamins Publishing Company. 251-273.

Chang, Pi-Chuan, Huihsin Tseng, Dan Jurafsky, and Christopher D. Manning. 2009. Discriminative Reordering with Chinese Grammatical Relations Features. In Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation (SSST-3) at NAACL HLT 2009, pages 51–59, Boulder, Colorado. Association for Computational Linguistics.

Chung, Sandra, William A. Ladusaw, and James McCloskey. 1995. Sluicing and logical form. Natural language semantics 3.3. pages 239-282.

Chung, Sandra. 2005. Sluicing and the lexicon: The point of no return. In: Annual Meeting of the Berkeley Linguistics Society. pages 73-92.

Cignarella Alessandra Teresa, Cristina Bosco, and Paolo Rosso. 2019. Presenting TWITTIRÒ-UD: An Italian Twitter Treebank in Universal Dependencies. In Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019), pages 190–197, Paris, France. Association for Computational Linguistics.

Contini, Gianfranco. 1970. Un'idea di Dante. Torino: Piccola biblioteca Einaudi.

Chomsky, Noam. 1957. Syntactic structures. The Hague: Mouton.

Chomsky, Noam. Lectures on government and binding: The Pisa lectures. No. 9. Walter de Gruyter, 1993.

Consales, Ilde, Adriana Pelo. 2003. "Fare" vicario", "fare+ N", "fare+ V". Per un'analisi del verbo "fare" nell'italiano antico. Verbo italiano: studi diacronici, sincronici, contrastivi, didattici: atti del 35. Congresso internazionale di studi, Parigi, 20-22 settembre 2001.-(Pubblicazioni della Società di linguistica italiana). Bulzoni.

Contini Giancarlo. 1970. Dante come personaggio-poeta della «Commedia», in Varianti e altra linguistica. Una raccolta di saggi (1938-1968), Torino, Einaudi. 335-61.

Corbetta, Claudia, Marco Passarotti, Flavio Massimiliano Cecchini, and Giovanni Moretti. 2023. Highway to Hell. Towards a Universal Dependencies Treebank for Dante Alighieri's Comedy. In Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023), pages 154–161, Venice, Italy. CEUR Workshop Proceedings.

Corbetta, Claudia, Marco Passarotti, and Giovanni Moretti. 2024 a. The Rise and Fall of Dependency Parsing in Dante Alighieri's Divine Comedy. In Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA) @ LREC-COLING-2024, pages 50–56, Torino, Italia. ELRA and ICCL.

Corbetta, Claudia, Giovanni Moretti, and Marco Passarotti. 2024 b. Join Together? Combining Data to Parse Italian Texts. In Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), pages 251–257, Pisa, Italy. CEUR Workshop Proceedings.

Corbetta, Claudia, Giovanni Moretti, and Marco Passarotti. 2025 a. How to ... Ellipsis: a Proposal from Old Italian for the Annotation of Elliptical Structures in UD. UniDive 3rd General Meeting Hungarian Research Centre for Linguistics, Budapest, Hungary 29-30 January 2025.

Corbetta, Claudia, Marco Passarotti, Flavio Massimiliano Cecchini, & Giovanni Moretti. 2025 b. Hell Awaits: Building a Universal Dependencies Treebank for Dante Alighieri's Comedy. IJCoL. Italian Journal of Computational Linguistics, 11(11-1).

Corbetta, Claudia, Anna Erminia Colombi, Giovanni Moretti, and Marco Passarotti. 2025 c. What Is Better for Syntactic Parsing? A Comparison Between Supervised and Unsupervised Models on Dante and Cavalcanti. In Proceedings of the 12th Italian Conference on Computational Linguistics (CLiC-it 2025). CEUR Workshop Proceedings.

Corbetta, Claudia, Federica Iurescia, and Marco Carlo Passarotti. 2025 d. «Are you Afraid of Ghosts?» A Proposal for Busting Predicate Ellipsis in Universal Dependencies. In Proceedings of the 23rd International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2025), Ljubljana, Slovenia. Association for Computational Linguistics, 52–63.

van Craenenbroeck, Jeroen, and Tanja Temmerman (eds), 2018. The Oxford Handbook of Ellipsis, Oxford Handbooks (online edn, Oxford Academic, 8 Jan. 2019).

Croft, William. 2001. Radical construction grammar: Syntactic theory in typological perspective. Oxford University Press.

Culicover, Peter W. and Ray Jackendoff. 2005. Simpler Syntax. Oxford University Press: Oxford, pp. 233-301.

D'Achille, Paolo and Grossmann Maria. 2017. Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio, Firenze: Franco Cesati.

D'Amico, Marta. 2014. La sintassi della frase complessa nella Commedia di Dante. PhD Thesis. University of Pisa.

Dagnac, Anne. 2010. Modal Ellipsis in French, Spanish and Italian. Karlos Arregi, Zsuzsanna Fagyal, Silvina A. Montrul and Annie Tremblay. Romance Linguistics 2008 : Interactions in Romance, 313, John Benjamins, pp.157-170, 2010, Current Issues in Linguistic Theory.

- Dalrymple, Mary, Stuart M. Shieber, and Fernando Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and philosophy* 14.4: 399-452.
- Dardano, Maurizio. 2012. *Sintassi dell'italiano antico: la prosa del Duecento e del Trecento*. Roma: Carocci.
- Dardano, Maurizio, and Gianluca Frenguelli. 2004. *SintAnt: la sintassi dell'italiano antico*. Atti del convegno internazionale di studi, Università Roma Tre, 18-21 settembre 2002, Roma: Aracne.
- De Cesare, Anna Maria. 2019. *Le parti invariabili del discorso*. Le Bussole: Carocci.
- Delcorno, Carlo. 1970. Fare. In *Enciclopedia Dantesca*, vol. II, Rome: Istituto dell'Enciclopedia Italiana. 794–803
- Delmonte, Rodolfo, Antonella Bristot, and Sara Tonelli. 2007. VIT - Venice Italian Treebank: Syntactic and Quantitative Features. In *Processing Sixth International Workshop on Treebanks and Linguistic Theories*.
- Delmonte, Rodolfo. 2009. Treebanking in VIT: from Phrase Structure to Dependency Representation. In Sergei Nirenburg (ed.) *Language Engineering for Lesser-Studied Languages*, pages 51–81. IOS Press, Amsterdam, The Netherlands.
- de Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating Typed Dependency Parses from Phrase Structure Parses. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).
- de Marneffe, Marie-Catherine and Christopher D. Manning. 2008. The Stanford Typed Dependencies Representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8, Manchester, UK. Coling 2008 Organizing Committee.
- de Marneffe, Marie-Catherine, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, and Christopher D. Manning. 2014. Universal Stanford dependencies: A cross-linguistic typology. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4585–4592, Reykjavik, Iceland. European Language Resources Association (ELRA).
- de Marneffe, Marie-Catherine, and Joakim Nivre. 2019. Dependency grammar. *Annual Review of Linguistics* 5.1: 197-218.
- de Marneffe, Marie-Catherine, Christopher D. Manning, Joakim Nivre, Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2) : 255–308.
- Di Nuovo, Elisa, Manuela Sanguinetti, Alessandro Mazzei, Elisa Corino, and Cristina Bosco. 2022. VALICO-UD: Treebanking an Italian Learner Corpus in Universal Dependencies. *IJCoL. Italian Journal of Computational Linguistics*, 8(8-1).
- Depiante, Marcela. 2018. Null Complement Anaphora. in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. 657–680.
- De Roberto, Elisa. 2012. Le proposizioni relative. Dardano Maurizio (eds). In *Sintassi dell'italiano antico: La prosa del Duecento e del Trecento*. Roma: Carocci.: 196-269.

de Saint-Exupéry, Antoine. 2015. [1943]. *The Little Prince*. Trans. Gregory Norminton. Richmond: Alma Books.

Droganova, Kira, Filip Ginter, Jenna Kanerva, Daniel Zeman. 2018. Mind the gap: Data enrichment in dependency parsing of elliptical constructions. In *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*. 47–54.

Dryer, Matthew S. & Haspelmath, Martin (eds.) 2013. *WALS Online*. v2020.4

Du Bois, John W. 2014. Towards a dialogic syntax. *Cognitive linguistics* 25.3. 359-410.

Egerland, Verner, Anna Cardinaletti. 2010. I pronomi personali e riflessivi. In Salvi, Giampaolo, and Lorenzo Renzi, (eds.) *Grammatica dell'italiano antico*. Vol. 1. Bologna: il Mulino. 401-468.

Enciclopedia Dantesca. 1970. Dir. Umberto Bosco. Roma, Istituto della Enciclopedia Italiana.

Faibisoff, Leah. 2019. *ISCAD Annotated Bibliography: Contrapasso, International Seminar on Critical Approaches to Dante*. Website, May 2019. Online:  
<https://dante.medieval.utoronto.ca/contrapasso/>

Favaro, Manuel. Elisa Guadagnini, Eva Sassolini, Marco Biffi, Simonetta Montemagni. 2022. Towards the creation of a diachronic corpus for Italian: A case study on the GDLI quotations. In *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA 2022)–LREC 2022 Workshop Language Resources and Evaluation Conference*: 94-100.

Fillmore, Charles J. 1986. Pragmatically controlled zero anaphora. Annual meeting of the Berkeley linguistics society.

Fillmore, Charles J. and Paul Kay. 1993. *Construction grammar coursebook, chapters 1 thru 11 (reading materials for Ling. X20)*. Berkeley: University of California.

Frosini, Giovanna. 2017. *Inventare una lingua. Note sulla lingua della «Commedia»*. disponibile online:  
[https://graficheincomune.comune.milano.it/fgm/datigic/allegati/bachecaroot/danteincasatrivulzio/approfondimenti\\_ita/Frosini\\_Inventare%20una%20lingua.pdf](https://graficheincomune.comune.milano.it/fgm/datigic/allegati/bachecaroot/danteincasatrivulzio/approfondimenti_ita/Frosini_Inventare%20una%20lingua.pdf)

Frosini, Giovanni. 2020. *Storia dell'italiano. La lingua, i testi*. Roma: Salerno Editrice. 113–132

Gaeta, Livio, Silvia Luraghi. 2001. Gapping in classical Greek prose. *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”* 25.1. 89-113.

Garavelli, Bice Mortara. 2003. [1997]. *Manuale di retorica*. Milan: Bompiani.

Gengel, Kirsten. 2013. *Pseudogapping and ellipsis*. Oxford: Oxford University Press.

Gergel, Remus, Sonia Cyrino (forth.) On the diachrony of ellipsis. Ms. Saarland University/U. Campinas, Under contract for *The Wiley Blackwell Companion to Diachronic Linguistics*, ed. by E. Aldridge, A. Breitbarth, K. E. Kiss, A. Ledgeway, J. Salmons, & A. Simonenko. [Prepublication draft:  
[https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/gergel/\\_Ellipsis-Revision\\_2023.12.01.pdf](https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/gergel/_Ellipsis-Revision_2023.12.01.pdf)]

- Giacalone Ramat, Anna, 1990. Discourse function and syntactic change. In Henning Andersen & Konrad Koerner, eds., *Historical Linguistics 1987*, Amsterdam, Benjamins, pp. 175-190.
- Gigli, Sara. 2004. *La codifica sintattica della Commedia di Dante*. PhD Thesis. Università di Pisa.
- Ginzburg, Jonathan, and Ivan A. Sag. 2000. *Interrogative investigations*. Stanford, Calif.: CSLI Publications.
- Goldberg, Adele E. 2015. Compositionality. *The Routledge handbook of semantics*. Routledge. 419-433.
- Goodall, Grant. 2017. Coordination in syntax. *Oxford Research Encyclopedia of Linguistics*.
- van der Goot, Rob, Ahmet Üstün, Alan Ramponi, Ibrahim Sharaf, and Barbara Plank. 2021. Massive Choice, Ample Tasks (MaChAmp): A Toolkit for Multi-task Learning in NLP. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 176–197, Online. Association for Computational Linguistics.
- Guibon Gaël, Marine Courtin, Kim Gerdes, and Bruno Guillaume. 2020. When Collaborative Treebank Curation Meets Graph Grammars. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, France. European Language Resources Association. 5291–5300.
- Haegeman, Liliane. 2007. Subject omission in present-day written English. *Rivista di Grammatica Generativa*, vol. 32. 91-124.
- Hall, Alison, 2018. Fragments, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. 605–623.
- Halliday, Michael A. K., Ruqaiya, Hasan. 2013. [1976]. *Cohesion in English*. Routledge.
- Hankamer, Jorge. 1973. Why there are two than's in English. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*. Vol. 9. No. 1. Chicago Linguistic Society. 179-191.
- Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic inquiry* 7.3. 391-428.
- Hajič, Jan. 1998. Building a syntactically annotated corpus: The Prague dependency treebank. *Issues of Valency and Meaning*. Studies in Honour of Jarmila Panevová: 106-132.
- Hajič, Jan, Barbora Vidová-Hladká, and Petr Pajas. 2001. The Prague dependency treebank: Annotation structure and support. *Proceedings of the IRCS workshop on linguistic databases*. Philadelphia: University of Pennsylvania.
- Hajič, Jan, Eva Hajičová, Marie Mikulová, Jiří Mírovský. 2017. Prague dependency treebank. In *Handbook of Linguistic Annotation*. Dordrecht: Springer Netherlands. 555-594
- Hardt, Daniel F. 1992. An Algorithm for VP Ellipsis. In *30th Annual Meeting of the Association for Computational Linguistics*, pages 9–14, Newark, Delaware, USA. Association for Computational Linguistics.

- Hardt, Daniel F. 1993. *Verb phrase ellipsis: Form, meaning, and processing*. University of Pennsylvania.
- Hardt, Daniel F. 1997. An empirical approach to VP ellipsis. *Computational Linguistics* 23.4. 525-541.
- Hartmann, Iren, Haspelmath, Martin, Taylor, Bradley. (eds.). 2013. *The Valency Patterns Leipzig online database*. Max Planck Institute for Evolutionary Anthropology, Leipzig. (Available online at <https://valpal.info>)
- Havelka, Jiri. 2007. *Mathematical Properties of Dependency Trees and their Application to Natural Language Syntax*. Ph.D. thesis, Univerzita Karlova – Matematicko-fyzikální fakulta, Prague, Czech Republic, June.
- Haverinen, Katri, Jenna Nyblom, Timo Viljanen, Veronika Laippala, Samuel Kohonen, Anna Missilä, Stina Ojala, Tapio Salakoski & Filip Ginter. 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. *Lang Resources & Evaluation* 48, 493–531.
- Haug, Dag, Silvia Luraghi. 2012. Introduction. *Acta Linguistica Hafniensia*, 44(2), 109–115.
- Heinecke, Johannes. 2019. ConlluEditor: a fully graphical editor for Universal dependencies treebank files. In: *Universal Dependencies Workshop 2019*. Paris.
- Hromei, Claudiu Daniel, and Roberto Basili. 2024. U-DepPLLAMA: Universal Dependency Parsing via Auto-regressive Large Language. *IJCoL-Italian Journal of Computational Linguistics* vol. 10, n. 1 june 2024: 21.
- Hudson, Richard. 2007. *Language networks: The new word grammar*. Oxford University Press.
- Hudson, Richard. 2017. *Dependency Grammar*. In: Hippisley A, Stump G, eds. *The Cambridge Handbook of Morphology*. Cambridge Handbooks in Language and Linguistics. Cambridge University Press :657-682.
- Inglese, Giorgio. 2012. *Dante: guida alla Divina Commedia*. Nuova edizione. Carocci.
- Isidoro, di Siviglia. 2004. *Etimologie, o Origini*. Vol 1. Libri 1-11. Unione tipografico-editrice torinese.
- Kehler, Andrew. 2018. Ellipsis and Discourse. in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks (online edn, Oxford Academic, 8 Jan. 2019), 314–341.
- Kehler, Andrew. 2000. Coherence and the Resolution of Ellipsis. *Linguistics and Philosophy* 23, 533–575.
- Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford, CA.
- Keydana, Götz, Silvia Luraghi. 2012. Definite referential null objects in Vedic Sanskrit and Ancient Greek. *Acta Linguistica Hafniensia* 44.2. 116-128.

- Kertz, Laura. 2010. *Ellipsis reconsidered*. University of California, San Diego.
- Khan, Mohammad, Markus Dickinson, and Sandra Kübler. 2013a. Towards domain adaptation for parsing web data. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 357–364, Hissar, Bulgaria. INCOMA Ltd. Shoumen, Bulgaria.
- Khan, Mohammad, Markus Dickinson, and Sandra Kuebler. 2013b. Does size matter? text and grammar revision for parsing social media data. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 1–10, Atlanta, Georgia. Association for Computational Linguistics.
- Kondratyuk, Dan, and Milan Straka. 2019. 75 Languages, 1 Model: Parsing Universal Dependencies Universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China. Association for Computational Linguistics.
- Krisch, Thomas. 2009. On the “syntax of silence” in Proto-Indo-European. *Information Structure and Language Change New Approaches to Word Order Variation in Germanic*. 191-222.
- Kromann Matthias Trautner. 2003. The Danish Dependency Treebank and the DTAG treebank tool. In *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories (TLT)*, pages 217–220.
- Lipenkova Janna and Milan Souček. 2014. Converting Russian Dependency Treebank to Stanford Typed Dependencies Representation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, volume 2: Short Papers, pages 143–147, Gothenburg, Sweden. Association for Computational Linguistics.
- Kuhn, Thomas S. 1970. [1962]. *The Structure of Scientific Revolutions*. 2d ed. Chicago: University of Chicago Press.
- Kuno, Susumu. 1976. Gapping: A functional analysis. *Linguistic Inquiry* 7.2. 300-318.
- Johnson, Kyle. 2019. Gapping and Stripping. in Jeroen van Craenenbroeck, and Tanja Temmerman (eds). *The Oxford Handbook of Ellipsis*, Oxford Handbooks. pp. 562–604.
- Lechner, Winfried. 2004. Ellipsis in comparatives. *Studies in Generative Grammar*. vol. 72.
- Lechner, Winfried, 2018. Comparative Deletion, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. 624–656
- Lee, Geonhee, Jong-Bok Kim. 2021. Negative Stripping in English: A Corpus-based Approach. *Language and Information* 25.3.
- Lerner, Jan, Pinkal, Manfred. 1995. Comparative ellipsis and variable binding. In: *Semantics and Linguistic Theory*. 222-236.
- Lo Duca, Maria G. 2000. Proprietà valenziali e criteri di descrizione lessicografica: un caso di alternanza argomentale. *Studi italiani di linguistica teorica ed applicata* 29.2. 219-242.
- Longo, Nicola. 2013. L'arte della ripetizione nella Divina Commedia. *Atti e memorie dell'Arcadia*: 2, 2013. 23-38.

- Luraghi, Silvia. 1997. Omission of the direct object in Classical Latin. *Indogermanische Forschungen* 102, 239–257.
- Luraghi, Silvia. 1998. Omissione dell'oggetto diretto in frasi coordinate: dal latino all'italiano. In *Sintassi storica. Atti del XXX Congresso SLI*, ed. P. Ramat. Roma: Bulzoni. Pp 183-196.
- Luraghi, Silvia. 2003. Definite referential null objects in Ancient Greek. *Indogermanische Forschungen* 108. 167-194.
- Luraghi, Silvia. 2004. Null objects in Latin and Greek and the relevance of linguistic typology for language reconstruction. *Proceedings of the 15th Annual UCLA Indo-European Conference, JIES Monograph. Vol. 49.*
- Luraghi, Silvia, Elisa Roma. 2021. Valency and transitivity over time: An introduction. *Valency over Time: Diachronic Perspectives on Valency Patterns and Valency Orientation* 368: 1.
- Luraghi, Silvia, Alessio P. Apro시오, Chiara Zanchi, Martina Giuliani. 2024. Introducing PaVeDa–Pavia Verbs Database: Valency Patterns and Pattern Comparison in Ancient Indo-European Languages. *Proceedings of the Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)@ LREC-COLING-2024.* 79-88.
- Lobeck, Anne C. 1995. *Ellipsis: Functional heads, licensing, and identification.* Oxford University Press.
- Mambrini Francesco and Marco Carlo Passarotti. 2012. Will a parser overtake Achilles? First experiments on parsing the ancient Greek dependency treebank. In *Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories (TLT11)*. 30 November–1 December 2012, Lisbon, Portugal, pages 133–144. Edições Colibri.
- Manni, Paola. 2013. *La lingua di Dante.* Bologna: Il Mulino.
- Marcus, Mitchell, Santorini, Beatrice, Marcinkiewicz, Mary Ann, 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, Vol 19.
- Marcus, Mitchell, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, Britta Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey*, March 8-11.
- Marello, Carlo. 2010. Ellissi. In Salvi, Giampaolo, and Lorenzo Renzi, (eds.) *Grammatica dell'italiano antico*. Vol. 2. Bologna: il Mulino. 1369-1386.
- Marti, Mario. 1972. *Storia dello Stil Nuovo*, Collezione di Studi e Testi, Milella, Lecce.
- Martin, Andrea E., Brian McElree. 2009. Memory operations that support language comprehension: Evidence from verb-phrase ellipsis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1231–1239. <https://doi.org/10.1037/a0016271>
- Martin, Andrea E., Brian McElree. 2011. Direct-access retrieval during sentence comprehension: Evidence from sluicing. *Journal of Memory and Language*, Volume 64, Issue 4. Pages 327-343.

- McEnergy, Tony, Hardie, Andrew. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: CUP.
- McDonald, Ryan, Joakim Nivre, Yvonne Quirnbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. Universal Dependency Annotation for Multilingual Parsing. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 92–97, Sofia, Bulgaria. Association for Computational Linguistics.
- McShane, Marjorie Joan. 2005. *A theory of ellipsis*. Oxford: Oxford University Press.
- Mel'čuk, Igor. 1988. *Dependency syntax: Theory and practice*. Albany, NY: State University of New York Press.
- Mel'čuk, Igor. 2009. *Dependency in natural language*. *Dependency in linguistic description*. John Benjamins Publishing Company, 2009. 1-110.
- Merchant, Jason. 2001. *The syntax of silence*. Oxford: Oxford University Press.
- Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661–738.
- Merchant, Jason. 2006. Sluicing. in Martin Everaert and Henk van Riemsdijk (eds), *The Blackwell companion to syntax*. Maldon, MA and Oxford: Blackwell, 269–89.
- Merchant, Jason. 2008. An asymmetry in voice mismatches between VP-ellipsis and pseudogapping. *Linguistic Inquiry* 39:169-179.
- Merchant Jason. 2012. Ellipsis. to appear in *Syntax: An international handbook of contemporary syntactic research*, by Tibor Kiss and Artemis Alexiadou, Berlin: Walter de Gruyter. Available online: <https://home.uchicago.edu/~merchant/pubs/merchant.ellipsis.pdf>
- Merchant, Jason. 2018. Ellipsis: A survey of analytical approaches, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks (online edn, Oxford Academic, 8 Jan. 2019)
- Michaelis, Susanne Maria, Maurer Philippe, Haspelmath Martin, Huber Magnus (eds.). 2013. *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://apics-online.info>
- Micheli, M. Silvia. 2022. CODIT. A new resource for the study of Italian from a diachronic perspective: Design and applications in the morphological field, *Corpus* (23).
- Migliorini, Bruno. 1960 [2001]. *Storia della lingua italiana*. Firenze: Sansoni.
- Mikulová, Marie. 2014. Semantic representation of ellipsis in the prague dependency treebanks. *Proceedings of the 26th Conference on Computational Linguistics and Speech Processing (ROCLING 2014)*. 125-138.
- Miller, Philip, 1992. Postlexical cliticization vs. affixation: Coordination criteria, in *Chicago Linguistic Society* 28, Chicago, CLS .382-396.

- Miller, Philip Harold, Geoffrey K. Pullum. 2014. Exophoric VP Ellipsis. Hofmeister, P.; Norcliffe, E. *The Core and the Periphery: Data-driven Perspectives on Syntax Inspired by Ivan A. Sag*, CSLI Publications. 5-32.
- Molinelli, Piera. 2010. Le strutture coordinate. In Salvi, Giampaolo, and Lorenzo Renzi, (eds.) *Grammatica dell'italiano antico*. Vol. 1. Bologna: il Mulino. 241-271.
- Moshtaghi, Nima, Masoud Dehghan, and Shahla Raghieb Doust. 2020. Cohesion in Discourse of Elderly with Alzheimer's. *Iranian Journal of Health Psychology* 3.1. 53-60.
- Murphy, Gregory L. 1985. Processes of understanding anaphora. *Journal of Memory and Language*, 24(3), 290–303.
- Napoli, Donna Jo. 1983. Comparative ellipsis: A phrase structure analysis. *Linguistic Inquiry* 14.4. 675-694.
- Napoli, Donna Jo, Marina Nespør. 1986. Comparative structures in Italian. *Language* 62.3. 622-653.
- Nardi, Bruno. 1983 [1942]. *Dante e la cultura medievale*. Nuova edizione a cura di Mazzantini, Introduzione di Gregory. Roma-Bari. Laterza.
- Nencioni, Giovanni. 1989a. *Saggi di lingua antica e moderna*. Rosenberg & Sellier.
- Nencioni, Giovanni. 1989b. Il contributo dell'esilio alla lingua di Dante, in *Dante e le città dell'esilio*. Atti del Convegno Internazionale di studi (Ravenna, 11-13 settembre 1987), Ravenna, Longo Editore. 177-198.
- Nguyen, Minh Van, Viet Dac Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021. Trankit: A Light-Weight Transformer-based Toolkit for Multilingual Natural Language Processing. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 80–90, Online. Association for Computational Linguistics.
- Nivre, Joakim, and Beata Megyesi. 2007. Bootstrapping a Swedish Treebank using cross-corpus harmonization and annotation projection. *Proceedings of the 6th international workshop on treebanks and linguistic theories*. Pennsylvania, PA: Association for Computational Linguistics.
- Nivre, Joakim, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666, Portorož, Slovenia. European Language Resources Association (ELRA).
- Nivre, Joakim, 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 Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.
- Nyblom, Jenna, Samuel Kohonen, Katri Haverinen, Tapio Salakoski, and Filip Ginter. 2013. Predicting Conjoint Propagation and Other Extended Stanford Dependencies. In *Proceedings of the*

Second International Conference on Dependency Linguistics (DepLing 2013), Prague, Czech Republic. Charles University in Prague, Matfyzpress, Prague, Czech Republic, pages 252–261.

Oflazer, Kemal, Bilge Say, Dilek Zeynep Hakkani-Tür, and Gökhan Tür. 2003. Building a Turkish treebank. In *Treebanks*. Springer, pages 261–277.

Ortega-Santos, Iván, Masaya Yoshida, Chizuru Nakao. 2014. On ellipsis structures involving a wh-remnant and a non-wh-remnant simultaneously, *Lingua*. Volume 138. Pages 55-85.

Osborne, Timothy, Michael Putnam, and Thomas Groß. 2012. Catenae: Introducing a novel unit of syntactic analysis. *Syntax* 15.4. 354-396.

Osborne, Timothy. 2018. Ellipsis in Dependency Grammar, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. 142-161.

Osborne, Timothy. 2019. Ellipsis. In *A Dependency Grammar of English*. John Benjamins Publishing Company. 349–378.

Osborne Timothy, Chenchen Song. 2025. Head-initial and head-Final coordinate structures in two annotation schemes of dependency grammar. In *Proceedings of the Eighth International Conference on Dependency Linguistics (DepLing, SyntaxFest 2025)*, Ljubljana, Slovenia. Association for Computational Linguistics, 74–83.

Osenova, Petya, Kiril Simov. 2004. BTB-TR05: BulTreeBank Stylebook. BulTreeBank Version 1.0. BulTreeBank Technical Report BTB-TR05.

Paccosi, Teresa and Alessio Palmero Aprosio. 2022. It Is MarkIT That Is New: An Italian Treebank of Marked Constructions. In *Proceedings of CLiC-it 2021 Italian Conference on Computational Linguistics*.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics* 31.1: 71-106.

Parodi, Ernesto Giacomo. 1896. La rima e i vocaboli in rima nella Divina Commedia, «*Bollettino della Società Dantesca*». 3, 6-9:81-156. (<https://bulletino.dantesca.it/dnt-fo-viewer/pages/document-details.jsf?documentId=506>)

Passarotti, Marco Carlo. 2018. Well, It Depends. Reflections on the Dependency Turn in Computational Linguistics. *Formal Representation and the Digital Humanities*. Cambridge Scholars Publishing. 141-158.

Passarotti, Marco, Cecchini, Flavio Massimiliano, Sprugnoli, Rachele, and Moretti, Giovanni. 2021. UDante. L'annotazione sintattica dei testi latini di Dante. *Studi Danteschi* 86: 309-338.

Pellegrini, Giovanni Battista. 1964. Lo stato attuale dei dialetti italiani e il problema dei confini dialettali. Estr. *Convegno preparazione Carta dei dialetti italiani* 16-17 maggio 1964.

Pelo, Adriana. 2012. Le proposizioni comparative. Dardano Maurizio (eds). In *Sintassi dell'italiano antico: La prosa del Duecento e del Trecento*. Roma: Carocci. 441-465.

Pescarini, Diego. 2015. *Le costruzioni con si*. Italiano, dialetti, lingue romanze. Roma: Carocci.

- Petrov, Slav, Dipanjan Das, and Ryan McDonald. 2012. A Universal Part-of-Speech Tagset. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey. European Language Resources Association (ELRA): 2089–2096.
- Phillips, Colin, Dan Parker, 2014. The psycholinguistics of ellipsis. *Lingua*, Volume 151, Part A. 78-95.
- Picone, Michelangelo. 1999. Dante come autore/narratore della Commedia. *Nuova rivista di letteratura italiana* 2.1.
- Pinkster, Harm. 2021. *The Oxford Latin syntax. Volume II: The complex sentence and discourse.* Oxford University Press.
- Plewnia, Albrecht. 2013. Möglichkeiten und Wahrscheinlichkeiten. Was man in der Dependenzgrammatik mit syntaktischen Leerstellen tun kann. *Die Ellipse. Neue Perspektiven auf ein altes Phänomen.* de Gruyter. 227-251.
- Poesio, Massimo, Roland Stuckardt, and Yannick Versley. 2016. *Anaphora resolution.* Berlin: Springer.
- Poletto, Cecilia. 2014. *Word order in Old Italian. Vol. 7.* OUP Oxford.
- Pollard, Carl, and Ivan A. Sag. 1994. *Head-driven phrase structure grammar.* University of Chicago Press.
- Popel, Martin, Zdeněk Žabokrtský, and Martin Vojtek. 2017. Udapi: Universal API for Universal Dependencies. In Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017), Gothenburg, Sweden. Association for Computational Linguistics. 96–101.
- Poppels, Till. 2022. Explaining ellipsis without identity. *The Linguistic Review*, vol. 39, no. 3. 341-400.
- Progovac, Ljiljana, et al. "The syntax of nonsententials." (2006): 1-384.
- Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pages 101–108, Online. Association for Computational Linguistics.
- Ramat, Paolo. 1982. Per una tipologia del “gapping”. *Linguistica contrastiva. Atti del XIII Congresso Internazionale di studi della Società di Linguistica Italiana (Asti, 1979)*, Roma, Bulzoni.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic inquiry* 17.3. 501-557.
- Rohlf, Gerhard. 1966-1969. *Grammatica storica della lingua italiana e dei suoi dialetti. Vol. I-III.* Temistocle Franceschi e Maria Caciagli Fancelli (trad.). Torino: Einaudi.
- Rohlf, Gerhard. 1972. *Studi e ricerche su lingua e dialetti d'Italia.* Firenze: Sansoni.

- Rosa, Rudolf, Jan Mašek, David Mareček, Martin Popel, Daniel Zeman, and Zdeněk Žabokrtský. 2014. HamleDT 2.0: Thirty Dependency Treebanks Stanfordized. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pages 2334–2341, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Ross, John Robert. 1969. Guess who?. in Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan (eds), Proceedings of the 5th annual meeting of the Chicago Linguistic Society. Chicago: Chicago Linguistic Society. 252–86.
- Ross, John Robert. 1970. Gapping and the order of constituents. In: Bierwisch, M., Heidolph, K.E. (Eds.), Progress in Linguistics. Mouton, The Hague. 249–259.
- Saab, Andrés, 2018. Nominal Ellipsis, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), The Oxford Handbook of Ellipsis, Oxford Handbooks. 526–561.
- Sabatini, Francesco. 1990. La comunicazione e gli usi della lingua. Torino. Loescher.
- Sabatini, Francesco. 1999. «Rigidità-esplicitzza» vs «elasticità-implicitzza»: possibili parametri massimi per una tipologia dei testi, in Gunver Skytte / Francesco Sabatini (edd.), Linguistica testuale comparativa. In memoriam Maria Elisabeth Conte. Atti del Convegno interannuale della Società di Linguistica Italiana (Copenaghen, 5-7 febbraio 1998), Copenaghen, Museum Tusulanum Press. 141-172.
- Sag, Ivan A. 1976. Deletion and logical form. Diss. Massachusetts Institute of Technology.
- Salvi, Giampaolo, Lorenzo Renzi, Anna Cardinaletti. 2001. (eds.) Grande grammatica italiana di consultazione. Nuova ed. Bologna: Il Mulino.
- Salvi, Giampaolo, and Lorenzo Renzi, 2010. (eds.) Grammatica dell'italiano antico. Vol. 1-2. Bologna: il Mulino.
- Salvi, Giampaolo. 2010. La realizzazione sintattica della struttura argomentale. In Grammatica dell'italiano antico. Vol. 1 Salvi, Giampaolo, Lorenzo Renzi (eds.). Bologna: il Mulino. 123-189.
- Sanguinetti, Manuela, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in Universal Dependencies. Proceedings of LREC 2018
- Sanguinetti Manuela, Cristina Bosco. 2014 a. PartTUT: The Turin University Parallel Treebank. In Basili, Bosco, Delmonte, Moschitti, Simi (editors) Harmonization and development of resources and tools for Italian Natural Language Processing within the PARLI project, LNCS, Springer Verlag.
- Sanguinetti Manuela, Cristina Bosco. 2014 b. Converting the parallel treebank ParTUT in Universal Stanford Dependencies. In Proceedings of the 1st Conference for Italian Computational Linguistics (CLiC-it 2014), Pisa (Italy)
- Schuster Sebastian, and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Portorož, Slovenia. European Language Resources Association (ELRA): 2371–2378.

- Seraji, Mojgan, Carina Jahani, Beáta Megyesi, and Joakim Nivre. 2014. A Persian treebank with Stanford typed dependencies. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*. Reykjavik, Iceland, May. European Languages Resources Association (ELRA): 796–801.
- Serianni, Luca, Alberto Castelvechi. 1991. *Grammatica Italiana*. Universitaria. UTET Università, Turin, Italy, second edition.
- Serianni, Luca. 2007. Sul colorito linguistico della *Commedia*. in «*Letteratura italiana antica*» 8: 141-150.
- Serianni, Luca. 2021. *Parola di Dante*. Il Mulino. Bologna.
- Shopen, Tim. 1973. Ellipsis as grammatical indeterminacy. *Foundations of language* 10.1. 65-77.
- Simi, Maria, and Simonetta Montemagni. 2018. Bootstrapping enhanced universal dependencies for Italian. *Ceur Workshop Proceedings*. Vol. 2253. CEUR-WS.
- Simov Kiril, Petya Osenova, Milena Slavcheva, Sia Kolkovska, Elisaveta Balabanova, Dimitar Doikoff, Krassimira Ivanova, Alexander Simov, and Milen Kouylekov. 2002. Building a Linguistically Interpreted Corpus of Bulgarian: the BulTreeBank. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).
- Spitzer, Leo. 1946. Note on the Poetic and the Empirical 'I'. in *Medieval Authors*. *Traditio* 4: 414-422.
- Squillaciotti, Paolo. 2021. I progetti digitali dell'OVI. *Griseldaonline*, 20.2: 197-203.
- Stabler, Edward P. 1999. Remnant movement and complexity. *Constraints and resources in natural language syntax and semantics*. 2: 299-326.
- Stainton, Robert. 2006. *Words and thoughts: Subsentences, ellipsis, and the philosophy of language*. Clarendon Press.
- Straka, Milan, Jan Hajič, and Jana Straková. 2016. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asunción Moreno, Jan Odijk Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (Irec'16)*, pages 4290–4297, Portorož, Slovenia, May. European Language Resources Association (ELRA).
- Straka, Milan and Jana Straková. 2017. Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In Jan Hajič and Dan Zeman, editors, *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, BC, Canada, August. Association for Computational Linguistics (ACL).
- Taylor, Ann, Marcus Mitchell, and Beatrice Santorini. 2003. The Penn treebank: an overview. *Treebanks: Building and using parsed corpora*. Springer: 5-22.

- Tavoni, Mirko. 2011. DanteSearch: il corpus delle opere volgari e latine di Dante lemmatizzate con marcatura grammaticale e sintattica. *Lectura Dantis 2002-2009. Omaggio a Vincenzo Placella per i suoi settanta anni. Vol. 2. Università degli Studi di Napoli "L'Orientale", Il Torcoliere. 583-608.*
- Tavoni, Mirko. 2015. DanteSearch: Istruzioni per l'uso. Interrogazione morfologica e sintattica delle opere volgari e latine di Dante. *Sintassi dell'italiano antico e sintassi di Dante*, edited by Marta D'Amico, Pisa: Felici, 59–79.
- Tavoni, Mirko. 2022. Allestimento, fruizione e prospettive di DanteSearch. *Corpora e Studi Linguistici. Atti del LIV Congresso della Società di Linguistica Italiana (Online, 8-10 settembre 2021). No. 6.*
- Tesnière, Lucien. 1959 [2015]. *Elements of structural syntax*. John Benjamins Publishing Company.
- Tesi, Riccardo. 2004. Parametri sintattici per la definizione di "Italiano antico". In Maurizio Dardano and Gianluca Frenguelli, (eds.), *SintAnt. La sintassi dell'italiano antico*. Roma: Aracne. 425–444.
- Thoms, Gary. 2013. Lexical mismatches in ellipsis and the identity condition. *NELS 42*. GLSA Publications.
- Tomasin, Lorenzo. 2013. Dante e l'idea di lingua italiana. In *Dante e la lingua italiana*, a cura di M. Tavoni. Ravenna. Longo. 29-46.
- Tomasin, Lorenzo. 2019. *Il caos e l'ordine*. Piccola Biblioteca Einaudi. Giulio Einaudi, Turin, Italy.
- Tomazzoli, Gaia. 2024. Digital resources for Dante studies: a critical survey. *Ecdotica 21.1: 77-112*.
- Tonelli, Sara, Rodolfo Delmonte, and Antonella Bristot. 2008. Enriching the Venice Italian Treebank with Dependency and Grammatical Relations. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Tsarfaty, Reut. 2013. A Unified Morpho-Syntactic Scheme of Stanford Dependencies. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 578–584, Sofia, Bulgaria. Association for Computational Linguistics.
- Vicente, Luis. 2018. Sluicing and Its Subtypes, in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. 479–503.
- Weir, Andrew. 2014. *Fragments and clausal ellipsis*. Diss. University of Massachusetts Amherst.
- Wilder, Christopher, 2018. Conjunction Reduction and Right-Node Raising. in Jeroen van Craenenbroeck, and Tanja Temmerman (eds), *The Oxford Handbook of Ellipsis*, Oxford Handbooks. pp. 681–718.
- Zanchi, Chiara, Silvia Luraghi, and Claudia Roberta Combei. 2022. PaVeDa-Pavia Verbs Database: Challenges and Perspectives. *Proceedings of the 4th workshop on research in computational linguistic typology and multilingual NLP*. pp. 99-102.

Zeman, Daniel. 2008. Reusable Tagset Conversion Using Tagset Drivers. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Marrakech, Morocco. European Language Resources Association (ELRA).

## Appendix

Predicate ellipsis	Nominal ellipsis	Clausal ellipsis
Pseudo-gapping Gapping VP-ellipsis Modal Complement Ellipsis	Subject ellipsis Object ellipsis	Stripping Fragments Sluicing Ellipsis in comparatives and comparative deletion

**Table 3.** Table with ellipsis classification (Section 3.2)

Types of predicate ellipsis	Example	What is omitted	What is <u>not</u> omitted
Pseudo-gapping	“Some brought roses, and others did lilies”	Lexical verb	Inflectional domain (= auxiliary); Internal argument (= object)
Gapping	“Jones likes seafood and Smith, bread”	Predicate (lexical verb + auxiliary)	Internal argument (= object)
VP-ellipsis	“Mr Honeypie was happily eating pumpkins, but his wife certainly wasn’t.”	Verbal element and internal argument	Auxiliary
Modal Complement Ellipsis	“I want to help you, but I can’t”	Verbal element and internal argument	Auxiliary → root modal

**Table 4.** Table of predicate ellipsis (Section 3.2.1)

### Queries:

Udapi query to retrieve orphan cases:

```
udapy util.Eval node='
if node.deprel == "orphan" : print(node.parent.deprel)' <
file.conllu | sort | uniq -c | sort -nr
```

Udapi query to retrieve promotion cases (excluded coordination):

```

udapy util.Eval node='
targets = {"acl", "acl:relcl", "advcl", "advcl:cmp",
"advcl:relcl", "ccomp", "ccomp:reported", "csubj",
"csubj:pass", "parataxis", "root"}
if node.deprel in targets and node.upos != "VERB" \
and not any(child.deprel == "cop" for child in node.children)
\
and not any(child.deprel == "orphan" for child in
node.children): \
print(node.deprel)' < file.conllu | sort | uniq -c | sort -nr

```

Udapi query to retrieve coordination in promotion cases:

```

udapy util.Eval node='
parent_targets =
{"acl","acl:relcl","advcl","advcl:cmp","advcl:relcl",
"ccomp","ccomp:reported","csubj","csubj:pass","parataxis","roo
t"}
if node.deprel == "conj" and node.upos != "VERB" \
and node.parent and node.parent.deprel in parent_targets \
and not any(child.deprel == "cop" for child in
node.children) \
and not any(child.deprel == "orphan" for child in
node.children):
print(node.deprel)' < file.conllu | sort | uniq -c | sort
-nr

```

Udapi query to retrieve coordination in promotion cases detailed per parent deprel:

```

udapy util.Eval node='
parent_targets =
{"acl","acl:relcl","advcl","advcl:cmp","advcl:relcl",
"ccomp","ccomp:reported","csubj","csubj:pass","parataxis","roo
t"}
if node.deprel == "conj" and node.upos != "VERB" \
and node.parent and node.parent.deprel in parent_targets \
and not any(child.deprel == "cop" for child in
node.children) \
and not any(child.deprel == "orphan" for child in
node.children):
print(node.parent.deprel + "\t in " + node.deprel) ' <
file.conllu | sort | uniq -c | sort -nr

```

Udapi query to retrieve occurrences of clausal deprel (excluded coordination):

```

udapy util.Eval node='
targets = {"acl", "acl:relcl", "advcl", "advcl:cmp",
"advcl:relcl", "ccomp", "ccomp:reported", "csubj",
"csubj:pass", "parataxis", "root"}
if node.deprel in targets : \

```

```
print(node.deprel)' < / file.conllu | sort | uniq -c | sort -nr
```

Udapi query to retrieve occurrences of clausal coordination:

```
udapy util.Eval node='
parent_targets =
{"acl","acl:relcl","advcl","advcl:cmp","advcl:relcl",
"ccomp","ccomp:reported","csubj","csubj:pass","parataxis","root"}
if node.deprel == "conj" \
and node.parent and node.parent.deprel in parent_targets : \
print(node.deprel)' < file.conllu | sort | uniq -c | sort -nr
```

The data used for the present analysis are available in the following repository:

<https://github.com/ClaudiaCorbe/Divine-Comedy-of-Dante-Alighieri>

The UD data (*Italian-Old* treebank) are available in the following repository:

[https://github.com/UniversalDependencies/UD\\_Italian-Old](https://github.com/UniversalDependencies/UD_Italian-Old)

The two datasets are identical in content; the main difference lies in their update status and organization. The dataset collected in the repository “Divine Comedy of Dante Alighieri” is more up to date, as it incorporates revisions and possible error corrections, whereas the UD data follow official release cycles and may therefore not always reflect the most recent updates, particularly with respect to the main branch and the online version of the treebank. In addition, the data in the repository are divided by *Cantica*, while the UD split follows the community standards (namely, train, dev, and test) which do not correspond to the structural division of the work.

## Acknowledgements

This work received support from the CA21167 COST action UniDive, funded by COST (European Cooperation in Science and Technology). I would like to thank the UniDive community for providing a stimulating research environment that greatly enriched my knowledge and enabled fruitful exchanges of scientific ideas.

I would like to express my gratitude to my supervisors, Marco Passarotti and Silvia Luraghi, for their guidance and continuous support throughout my doctoral research. I am also grateful to the faculty members and host institutions at the University of Pavia and the University of Bergamo, and in particular to the CIRCSE research center and all its members, for their valuable support and for providing a welcoming and stimulating academic environment.

I would also like to express my special thanks to Daniel Zeman, Petya Osenova, and Kiril Simov, who supported and guided me throughout my research period abroad.

*E quindi uscimmo a riveder le stelle.*  
'Thence we came forth to rebehold the stars.'  
*Inf. XXXIV, v. 139*