Note on Latin Grammar

Jinyuan Wu

March 28, 2023

Chapter 1

Introduction

1.1 The language and the speaker

1.1.1 Latin as a classical language

Latin was the language of the Romans and the official language of both the Roman Republic and the Roman Empire, and hence the official language of the Catholic Church, which was *the* church for the Western Roman Empire. The international nature of the Roman Empire made Latin the international language around the Mediterranean Sea at that time – indeed, *Mare Nostrum* 'our sea' in Latin, and its importance in science, arts, law, religion, and literature lent it more than one thousand of years of life as a common literary language and a sacred language in western Europe after the collapse of the Western Roman Empire and the emergence of the Romance language family.

As recently as the nineteen century, Latin was still fluently used by scholars and in the Catholic Mass. A decline in the popularity of Latin was observed after that. The rapid development of English (at first, also French and German and sometimes Russian) as the language of science largely replaced the status of Latin as a scholar language. After Vatican allowed vernacular languages being used in liturgies, Latin also largely lose its position in the daily use in the Catholic church.

This note is about **Classical Latin** – the Latin of classical Latin writers – and Ecclesiastical Latin – the kind of Latin of the Catholic church. That's to say Old Latin, vulgar Latin (with prototypes of Romance articles), etc. are not discussed in detail in this note. Still, some historical knowledge is important for us to understand why Latin is the way it is.

1.1.2 Latin in ancient Mediterranean world

The historical and contemporary importance of Latin of course doesn't endorse it as a inherently superior language. Indeed, Latin used to be TODO: other languages were more important

1.2 Previous studies

1.3 Theoretical orientation

In a word, the theoretical framework of this note is Basic Linguistic Theory (BLT)(Dixon, 2009, 2010, 2012) with generative flavors. Here by *generative* I mean Minimalism plus Distributed Morphology plus Syntactic Cartography (but the Antisymmetry theory is not strictly followed here; I only use the idea of a semi-rigid template of functional projections). Although generativism is harshly criticized by Dixon, I believe this is largely due to notational reasons; some criticisms, like "generativism sticks to the universal notation of words" or "generativists blindly believe in a noun-verb distinction *exactly* the same as English", are invalid in the aforementioned framework. To connect the aforementioned school of generativism and BLT, I list some observations:

- First, note that "functional heads" are just an alternative way to say "grammatical categories" or "grammatical relations" in a constituency-based framework doing away with dependency relations. The constituency-dependency correspondence has long been discovered (Schneider, 1998; Osborne et al., 2011; Kahane and Mazziotta, 2015; Nefdt and Baggio, 2023). On the other hand,

"lexical heads" – nouns, verbs, etc. – lie at the *bottom* of an "extended noun phrase (NP)" (i.e. the DP projections) or an "extended verb phrase (VP)" (i.e. the CP projections). This settles the issue raised by many descriptive linguists: the term *head* is no longer used in the same way as it was in contemporary generativism. The *head* of descriptive linguists is essentially the *root* in Distributed Morphology.

- It's possible to "zip" the Minimalist constituency (or dependency) structure: removing invisible functional projections, replacing labels like SpecTP with "subject", using the term *head* to refer to the lexical head, etc. Thus, we are able to automatically obtain more traditional constituency analysis (as in Huddleston and Pullum (2002)) or dependency analysis from generative trees. The counterpart of c-command relations in the dependency analysis is how "tight" a dependency relation is: that the relation between the verb and the object is tighter than the relation between the verb and the subject is equivalent to that the subject has a higher position in the syntactic tree.

- One implicit message hidden in the idea that NPs and clauses are the only two types of constituents in Dixon (2009) is that when we finish building up an NP and insert it into an argument slot in a clause, the syntactic processing enters a new stage; on the other hand, the difference between a half-finished NP and a completely finished NP is not that huge. Now if we use constituency analysis all the way down, we are in the risk of losing this piece of information. This is settled by the concept of *phase* in modern generativism: when typologists argue for recognizing only noun phrases and clauses as constituents, they are essentially referring to phases.

- The phase theory also explains why some have the intuition that the *verb phrase* should exclude the object: because when the CP is being built, the arguments are already "frozen", and what are manipulated and realized together are verbal functional heads – that's exactly *their* verb phrase. The similar thing happens for a *word* (see the next point).

- Some people (many functionalists, but also some formal grammarians) really don't like the idea that differences in constituent order have their roots in the constituent structure and especially in movements. They say constituent order *directly* reflects grammatical relations and categories like topicalization, instead of the mainstream generative idea that constituent order reflects constituency relations, which then codes things like topicalization. The exact meaning of "directly reflect" however is rather hard to tell from an empirical perspective. What we already know is that quantitative researches suggest that at least a semi-configurational approach (i.e. a linear template with fixed constituent slots in it) is needed to fully capture Latin constituent order, because the diachronic change of the frequency of OVAux looks very different from the diachronic change of the overall OV frequency, which includes, say, OAuxV (Danckaert, 2015). But after we accept the semi-configurational approach, we can then do tests like, say, Principles A, B, and C, coordination and ellipsis tests, etc., on slots of these templates, and usually a hierarchy of relative "strength " of dependency relations can be established (Danckaert, 2017, § 1.6). Then, by the duality between constituency and dependency, usually we will find that a constituency-based analysis is *accurate* for a so-called non-configurational language, although it may not be *convenient* for its documentation.

- A word is just a mini-phrase (in the above phase-as-descriptivist-phrase sense, possibly a mini-tree, possibly a collective realization of a span of functional heads). The syntax all the way down analysis taken in this note therefore explains why we always have controversies concerning whether a unit is a word or a phrase (like *American history teacher* – note that its inner parts don't actively participate in other syntactic processes): this is no objective standard for drawing a line between the two. What are objective are the morphosyntactic units recognized: *American history teacher* is a nominal compounding structure (a certain kind of FP within the NumP projections), regardless of whether you say it's a word.

  According to Cartography syntax, cross-linguistically, we should find similar patterns of functional heads, so we should expect to see a morphosyntactic unit of a size similar to what we usually call words in English in another language, although the native speakers may not find this unit important in their society (that "word" unit may not be the unit for measuring the length of an article or for writing).

- Finally, the hierarchy of functional heads – or in other words, grammatical relations and categories – can be "routinized" and packaged, and how they are stored in the actual brain may have

more resemblance to Tree-adjoining grammar (TAG). Most inflection patterns, for example, seem to be packaged, which explains why sometimes they seem to be psychologically different from syntax, though Distributed Morphology has shown it's possible to treat the grammar as syntax all the way down. This is just what people call *construction*. However, it seems a construction is still not a packaged *linear* sequence: its inner structure still observes the usual rules for syntax and may (although of course sometimes may not) engage actively with productive syntactic elements. Thus a structuralist – as opposed to canonical constructivism – analysis is still valuable.

Elements of a language, in the perspective of Distributed Morphology, contain Lists A, B, and C, which are a list of roots and features, a list of how List A is phonetically realized (and also covers some syntactic selections that don't have semantic motivation: "if A and B meet, they never get appropriately spelt out and the derivation crashes – no why"), and the meaning of idiomatic phrases. With functional heads being kicked out in a practical language description project, the same amount of information needs to be shown in a different matter. What we need now are:

- Abstract syntax and "abstract morphology": facts like that Latin has six cases, that Latin has well-defined subjects and objects, etc. The grammatical concepts are knitted into subcategorization frames, each of which waits for a lexical head and some arguments being filled in. This covers the List A and the abstract part of List B (i.e. what structures can be spelt out, without considering what *is* the realization)
- How the above is realized: actual nominal and verbal paradigms, productive derivation rules, constituent order, which may involve "transformation rules". This covers the concrete part of List B.

  The term *transformation rule* is kind of misleading because it's possible – and even frequent – that a marked construction can't be obtained by transformation of the canonical construction in a natural way, because what really happens is that the two constructions undergo shared stages of syntactic structure building, and then diverge from each other, and the transformational rule linking the two is only a coarse phenomenological description of the relation between the two (Fig. 1.1).
- A dictionary, containing roots, how roots are placed into subcategorization frames, and established meanings of complex constructions (all things in List C). The discussion on (synchronic) roots may also involve historical morphology, the products of which are syntactically independent roots. The last includes some semi-fossilized derivations idioms in the everyday sense, formulaic speech, etc. Usually dictionaries don't really record roots: they record already inflected principal parts, from which the whole paradigm can be found.

The first two are about the grammar, while the third is about the dictionary or the so-called **lexicon**.

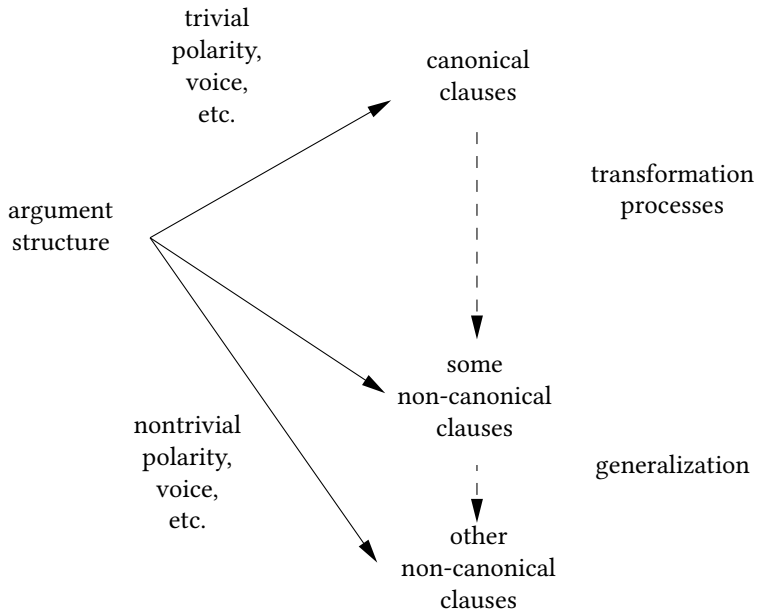

Figure 1.1: What is a transformation rule

1.4 About this note

This note roughly follows the example of

Of course, it's still possible to carry out grammatical description in different ways. For example, generally we shouldn't separate morphology and syntax categorically. For a heavily inflected language like Latin, however, it's an appealing idea to start a chapter named "verbal morphology" and cover all TAME marking in it, both the abstract concepts and the concrete paradigms.

1.5 Texts

TODO: classical writers

1.6 Remarkable features

Some typological parameters and peculiar points of Latin are listed here for the impatient.

1.6.1 Morphology

Latin is well known for its rich morphology, which enables a rather free – but still not completely arbitrary – constituent order. Latin has a clear inflection-derivation distinction. Despite its richness, Latin derivation is largely historical, with meanings of derived forms having shifted and no longer regularly inferrable. Latin inflection is always suffixal, while derivation is predominantly prefixal. Concatenative morphology (affixation and compounding) is prominent but isn't the only morphological device: the following non-concatenative mechanisms are all attested:

- *Reduplication*: formation of the perfect stem (TODO: ref)
- *Subtraction*: dropping of first-conjugation stem-final vowel (§ 7.4.1).
- *Infixation*: TODO: ref The imperfect *-ba-* is sometimes said to be an infix (as well as its counterparts like *-bi-*), though it fits in a concatenative picture of verbal morphology.

These mechanisms, however, are largely historical, just like their concatenative counterparts.

Most clausal grammatical categories are marked on the verbal morphology. Sometimes a grammatical category is there but is not reflected in the morphology. For example, in English we have infinitive clauses, but strictly speaking, there is no such thing as "infinitive verb": the head verb of an infinitive clause has exactly the same form of a non-third person singular present tense verb. This is not the case in Latin. For example, the head verb of a infinitive clause in Latin indeed has a separate position in the paradigm. Thus, grammatical categories of the clause are listed in this section.

Box 1.1: Advices when reading

The morphological richness (and the scrambled constituent order) makes Latin hard to read especially for people whose first languages are, say, English or Mandarin. Whenever unsure about a sentence, do the follows:

1. Skim over the words and label the stems that can be easily recognized.
2. Skim over and circle uncontroversial grammatical items, like inflectional endings and prepositions. It's OK to be unable to interpret them immediately (and we need the steps below).
3. Choose a grammatical item and tentatively give a list of possible features it carries. For example, seeing *-v-* in a verb usually means it's based on the perfect stem (§ 7.3.3); *-um* may be second declension accusative, but there are other possibilities (Table 4.1).
4. Use constraints like "the preposition *in* licenses the accusative case or the ablative case" to narrow the possibilities identified above.
5. Draw unfinished dependency arrows: for a verb, draw arrows pointing to the subject and/or the object; for a nominative adjective, draw an arrow pointing to the modified

head noun. But note that it's possible that the subject is dropped, or there is no head noun (compare English *the poor*). Then try to pair the arrows.

Repeat the above procedure and finally the sentence can be understood. This procedure is demonstrated in § 11.2.1.

1.6.2 Lack of determiner

TODO: other IE languages

1.6.3 Alignment

Latin is a clear nominative-accusative language. Similar to what is documented in CGEL, Latin core arguments are coded as subject, object(s), and copular complements at the level of alignment. They can be distinguished by the semantic roles, case marking, possible contents, and transformational properties (§ 8.1). In

1.6.4 Peripheral arguments

There is no serial verb constructions in Latin (§ **??**), and thus semantic functions like location or instrument are always realized by typical peripheral arguments attached to the core argument structure. These peripheral argument positions sometimes can be filled by adverbs, which also reveals an origin of adverbs: fossilized case forms.

1.6.5 Constituent order

Although traditional Latin grammars focus on the dependency relations introduced by morphology, deeper examination of Latin grammar reveals that the dependency relations have a role in determining the constituent order, and thus Latin is better described as a discourse-configurational language, with multiple topicalization and focusing structures. There are several examples of the link between constituent order and dependency relations. The negator *non* usually appears before the verb (TODO: or aux?), and apparent violations seem to be constituent negation as opposed to sentential negation (Danckaert, 2017, p. 43); if we subtract sentences in which the OV/VO alternation can be alternatively analyzed as topicalization and only keep sentences in which the object stays in the VP (supposing there is a well-defined VP), then VO frequency no longer shows significant change as time went by (Danckaert, 2017, § 1.5, p. 29); constituency tests also robustly hint at an VP.

Box 1.2: Constituency deemphasized in Latin grammar

This is probably not surprising because even the most non-configurational languages show configurationality under scrutiny (Niedzielski, 2017; Morris, 2018; Legate, 2002, among others) and therefore a thorough disruption of the existing framework of generative syntax seems unnecessary. The question then becomes *how* "free-order" Latin is. Is it closer to a prototypical "non-configurational" language, say Warlpiri, or is it closer to Japanese where we have more localized scrambling? I will address this question in TODO: ref ,

For the sake of convenience, even though we know mainstream generative (constituency-based, though the introduction of movements and the structure of Cinque hierarchy gives it certain flavor of dependency grammars) approaches make perfect sense for Latin, a systematic and thorough description of Latin grammar would be better carried out in a dependency-relation based or BLT-based way. This is of course mostly notational change: for example, we only recognize the most "salient" types of constituents like NPs and clauses as constituents in our description, and the existence of fine-grained functional projections is covered by additional information like the "height" or "closeness" of dependency arcs corresponding to these functional projections (§ 1.3).

Chapter 2

Phonology and the writing system

2.1 The alphabet

The most accepted writing system of Latin developed into what we call Latin letters – or the Roman alphabet – today, which is the most widely used writing system in the world. **Old Italic scripts**, used by Early Old Latin inscriptions as well as neighbor languages, show a larger degree of variation, which clearly derived from Greek letters. The standard Latin alphabet derived from old Italic scripts,

Chapter 3

Parts of speech

3.1 Overview

Latin word classes can be defined easily via morphology, and these classes prove to have morphosyntactic significance. Traditionally speaking, word classes with none or poor morphology are called **particles**, and non-particle words can be divided into two large classes: those with similar morphology of prototypical nouns (i.e. **declension**) are **nominals**, while words with similar morphology of prototypical verbs (i.e. **conjugation**) form a uniform class rightfully called **verbs**. Nominals include **nouns** and **adjectives**, the distinction between the two can also be defined morphologically.

Latin particles include **prepositions**, **adverbs**, **interjections**, and **conjunctions**. The adverb class and the preposition class have a large overlap: often a preposition has an intransitive counterpart, which is similar to a prototypical adverb. Conjunctions may be seen as "prepositions for clauses". The functions and etymologies of particles are highly diverse.

Latin nouns, verbs, and adjectives are all open categories. They are able to head constituents, and so are correlatives (though correlatives can be listed in the grammar). The preposition class is closed and is a part of the grammar, just like conjunctions. However, conjunctions are purely functional, while certain prepositions may be argued to head attributive expressions: though prepositions are often said to be markers of a periphrastic case system, the semantics carried by certain Latin prepositions are too complicated for a case system. This is also the case of adverbs: some adverbs seem to be periphrastic markers of TAME categories and therefore may be considered as a part of the grammar, while others seem to carry "real" meanings. Fig. 3.1 is a visualization of the classification of Latin word classes.

> **Box 3.1: Lexical and function classes**
>
> By words with "real category labels", I mean words that have"real" meanings and serve as lexical heads of constituents (i.e. being surrounded by function words and dependents). Certain adverbs and prepositions have "real category labels", and they appear at the left side in Fig. 3.1. Prepositions can be enumerated and therefore are considered as a part of the grammar, so they are always at the lower side in Fig. 3.1. Other adverbs and prepositions are light in their semantic and are purely functional, so they appear in the southeast corner of Fig. 3.1.
>
> Note that this is not the standard terminology. Linguists use their own notion of *lexical class* and *function class* to cover what I say here.

Articles (English *a* or *the*), despite prevalent in other Indo-European languages, are missing in Latin. This, together with the fact that Classical Sanskrit and Old Persian didn't have articles and the Slavic languages still don't, is a strong indicator that proto-Indo-European (PIE) didn't have articles. Note that the fact that Latin lacks articles doesn't mean the determiner syntactic function doesn't exist: there are evidences suggesting certain aspects of the behavior of Latin NPs are just like English (Giusti and Iovino, 2014).

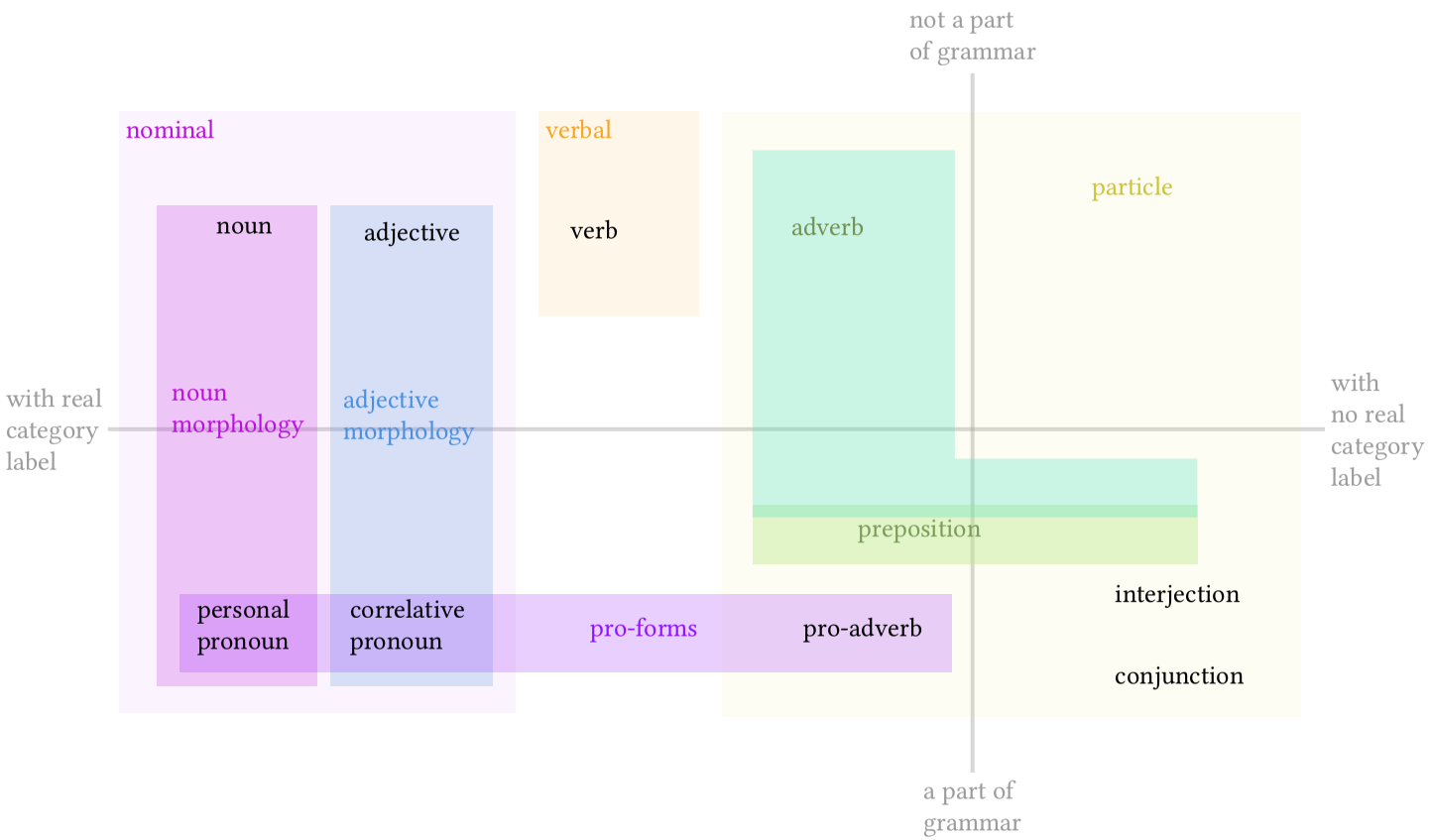

Figure 3.1: Latin word classes

Chapter 4

Noun and noun phrase

4.1 Introduction

4.2 Nominal derivation

4.3 Declension of regular nouns

4.3.1 The paradigms

There are five declension classes in Latin. The paradigms of each is well documented in Allen and Greenough (1903, p. 17). The full list of attested noun endings is Table 4.1. The list is still not the full picture of Latin nominal inflection: stem alternation is seen in the third declension.

Table 4.1: Declension endings; Roman numerals are declension classes

ending	declension
-a	I, SG.NOM, SG.VOC; IIN, PL.NOM, PL.ACC, PL.VOC; IIIN, PL.NOM, PL.ACC, PL.VOC
-ā	I, SG.ABL
-ae	I, SG.GEN, SG.DAT, PL.NOM, PL.VOC;
-am	I, SG.ACC
-ārum	I, PL.GEN
-ās	I, PL.ACC
-e	IIM, SG.VOC; IIIFMN, SG.ABL
-ē	V, SG.ABL
-ei/-ēi	V, SG.GEN, SG.DAT
-em	IIIFM, SG.ACC; V, SG.ACC
-ēbus	V, PL.DAT, PL.ABL
-ērum	V, PL.GEN
-ēs	IIIFM, PL.NOM, PL.ACC, PL.VOC; V, SG.NOM, SG.VOC, PL.NOM, PL.ACC, PL.VOC
-ī	IIM, SG.GEN, SG.VOC, PL.NOM, PL.VOC; IIN, SG.GEN; IIIFMN, SG.DAT
-ibus	IIIFMN, PL.DAT, PL.ABL; IVFMN, PL.DAT, PL.ABL
-is	IIIFMN, SG.GEN
-īs	I, PL.DAT, PL.ABL; IIMN, PL.DAT, PL.ABL
-ō	IIMN, SG.DAT, SG.ABL
-ōs	IIM, PL.ACC
-ōrum	IIMN, PL.GEN
-r	IIM, SG.NOM, SG.VOC
-ū	IVFM, SG.ABL; IVN, SG.NOM, SG.DAT, SG.ACC, SG.ABL, SG.VOC
-ua	IVN, PL.NOM, PL.ACC, PL.VOC
-uī	IVFM, SG.DAT
-um	IIMN, SG.ACC; IIN, SG.NOM, SG.VOC; IIIFMN, PL.GEN; IVFM, SG.ACC
-us	IIM, SG.NOM; IVFM, SG.NOM, SG.VOC
-ūs	IVFM, SG.GEN, PL.NOM, PL.ACC, PL.VOC; IVN, SG.GEN
-uum	IVFMN, PL.GEN

Box 4.1: Frequent confusions when analyzing noun endings

The ending sequence *-io* can'be be found in Table 4.1 and we may hurry to the conclusion that it's the third declension abstract noun ending *-io* in the nominative or accusative case. Not necessarily – it can also be *-ium* in the dative or ablative case (when the macron symbol for long vowels are not used).

4.4 The structure of the noun phrase

Although Latin lacks the article

4.4.1 Attributives

This section only discusses adjective or numeral attributives in detail. For in-depth discussion of relative clauses, see § 10.4.

4.4.2 Arguments of adjectives

TODO: case forms with adjectives

4.4.3 The possessive construction

4.4.4 Numerals in the noun phrase

4.4.5 Distribution of the cases

The roles of the five cases are not symmetric. Being nominative simply means being the subject in a finite clause or something agreeing to it and nothing else: the subject may be in a passive clause and is not agentive at all. The nominative case and the accusative case received by the direct object are *structural* cases: they are purely decided by the syntactic environment and don't have much semantic significance.

On the other hand, the rest cases are *inherent* cases: they are similar to prepositions, having semantic interpretations – "source" or "target" or … – themselves, and once an inherent case is assigned to an NP, the latter is "sealed" just like a prepositional phrase: the change of the outside syntactic environment doesn't change anything inside.

4.4.5.1 The nominative

4.4.5.2 The accusative

4.4.5.3 The dative

The dative case is assigned to the indirect object (§ 8.2.3)

4.4.6 Prepositions

Chapter 5

Adjectives

The adjective class and the adverb class are linked together by several factors: the adjective phrase and the adverb phrase are both prototypical modifiers, often with parallel structures; they both have the category of degree; adverbs can be formed regularly from adjectives.

5.1 Declension of regular adjectives

Peripheral arguments may also be regarded as adverbials. This chapter, however, is mainly about mean, TODO

tam quam

TODO: structure of some prepositions

Chapter 6

Minor nominal categories

Chapter 7

Verb inflection

7.1 The finite paradigm

7.1.1 The verb template

Some inflected Latin verbs and their parts are shown in Fig. 7.1. Traditionally, the verb is divided into the **stem** and the **ending**. Derivation in Latin is predominantly preverbal, and hence the conjugation is mostly about the final lexical morpheme in the verb stem, which is represented as the root in Fig. 7.1. There may be a perfect suffix after the root. Components of the verb ending include the **tense and mood marker**, and the person, number and voice marker, which is called the **personal ending** here, following the terminology in Allen and Greenough (1903, § 165).

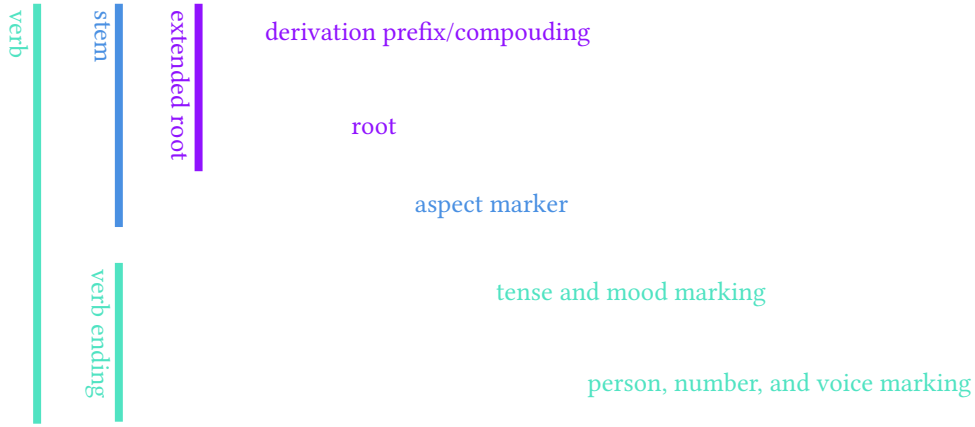

Figure 7.1: The template of Latin verbs. Indentation means linear order and not necessarily constituency structure.

The core stem assumes semi-regular alternations. The perfect marker is often but not always *-v-* (§ 7.3.3). Although the aspect marker is a part of the TAM marking (§ 7.1.3), it somehow is much more closely attached to the root in the morphophonological realization of the verbal system: how it is realized is not completely predictable, and therefore the two finite stem forms – the present stem and the perfect stem – are both required for complete characterization of the paradigm of a regular verb. To fully decide the paradigm we still need a further stem, the supine stem (§ 7.3.1).

Contextual allomorphs also exist for other morphemes in Fig. 7.1 (Embick and Halle, 2005, p. 11), as is shown in Table 7.1. The most salient change is that a vowel may be inserted between the root and the aspect suffix, which is also known as the **thematic vowel**. It is the residue of the PIE ablaut and labels the conjugation class. The tense and mood marker is influenced by the existence of the perfect suffix. The personal ending is also influenced by what precedes it, as in the the *-mus* and *-imus* alternation.

Below I discuss the subsystems in Fig. 7.1.

Table 7.1: Examples of Latin finite verbs

verb form	stem			tense and mood	personal ending	
	extended root		aspect marker		“thematic vowel”	personal ending
	extended root	thematic vowel				
amō	*am*					*ō*
laudāmus	*laud*	*ā*				*mus*
olēvimus	*ol*	*ē*	*v*		*i*	*mus*
amāveris	*am*	*ā*	*v*	*eri*		*s*

7.1.2 Voice

Latin doesn't have rich valency changing devices: there is only one clause-wide valency decreasing device – passivization – and there is no valency increasing device. Causative constructions are realized by complement clauses, not any change in the argument structure. Whether passivization happens is recorded by the category of **voice**. A verb (and hence the clause headed by it) is therefore either in **active voice**, or in **passive voice**.

> **Box 7.1: Valency changing**
>
> From a generative perspective, some languages realize valency changing by a series of *v*P structures, and then the case assignment of the arguments is trivial. Some languages use non-trivial structural case assignment mechanisms to achieve valency changing ("suppressing the agent argument, and leave the nominative probe to find the subject; the probe then has to choose the patient argument"). Of course, *v*P changes in the second type are still there, which may be a likely source of relevant verb morphology. Naturally, the second group of languages have more restricted valency changing devices; this is the case of Latin.

7.1.3 TAME categories

Latin has fused tense and aspect: the composition of three tense values and three aspect values gives nine options, but in Latin, there are only six morphologically distinguished options, as is shown in Table 7.2. When people talk about **tense** in Latin (and in many other Indo-European languages), they are often taking about things like the six options, instead of the past/present/future system. The imperfective/perfective distinction (for example the *be doing* construction in English) is not syntactically coded in Latin.

Table 7.2: Latin tense and aspect

	past	present	future
imperfect	IMPERFECT	PRESENT	FUTURE
simple	PERFECT		
perfect	PLUPERFECT	PERFECT	FUTURE PERFECT

> **Box 7.2: Mismathc between TAME constructions and fine-grained categories**
>
> Note that semantic TAME features are different from syntactic TAME features, and the two are in turn different from packaged TAME marking constructions that can be easily identified in surface-orientated analyses. This is illustrated in Table 7.2. Following the example in Grimm (2021), in this note, I use small capitals for the names of attested surface realizations of TAME and the default font for TAME values. (Some other grammars, like Jacques (2021); Friesen (2017), use initial capitals for the former.)

Similar fusion between categories is shown in the category of **mood**. It's the fusion of morphologically marked clause type (declarative and imperative) and morphologically marked modality. The verb morphology of interrogative clauses is exactly the same as declarative clauses: the interrogative clause type is marked by the existence of interrogative *pro*-forms. Thus, there are three moods in finite clauses in Latin: INDICATIVE, SUBJUNCTIVE, and IMPERATIVE. The INDICATIVE is the fusion of the declarative/interrogative clause type and the realis modality. The SUBJUNCTIVE mood is the fusion of the declarative/interrogative clause type and the irrealis modality. The IMPERATIVE is basically the imperative clause type: it doesn't allow modality marking. Sometimes people say the infinitive is the fourth mood, though it's a non-finite clause.

> **Box 7.3: The term *mood***
>
> BLT only calls the first category *mood*. Different linguists use the term *mood* and *modality* in radically different ways. In this note I just focus on the common practice in Latin grammar study.

7.1.4 Agreement

Latin is a typical nominative-accusative language, both morphologically and syntactically. In finite clauses, there is subject-verb agreement: the number and person of the subject is marked on the main verb. In the case of periphrastic conjugation, the features are marked on the copula.

7.1.5 Compatability of categories

There is no FUTURE tense and FUTURE PERFECT tense in subjunctive clauses, probably for the semantic reason that the future tense already contains certain sense of modality (an event predicted to happen), and thus is not compatible with the SUBJUNCTIVE mood. The IMPERATIVE mood is not compatible with other TAME markings except the PRESENT tense and the FUTURE tense. It's still compatible with the voice category, and allowed persons are second person singular/plural with the PRESENT tense, and second/third person singular/plural with the FUTURE tense. The absence of first person is also probably from semantic origin.

In conclusion, the categories involved in the finite verb paradigm of Latin are shown in Fig. 7.2. Here mood and tense are realized in one morpheme, and voice, person and number are realized in one morpheme. The paradigm is realized synthetically in all circumstances except in passive voice and perfect tense. In that case, the verb conjugation is realized like the English passive, i.e. via a copula and the perfect passive participle.

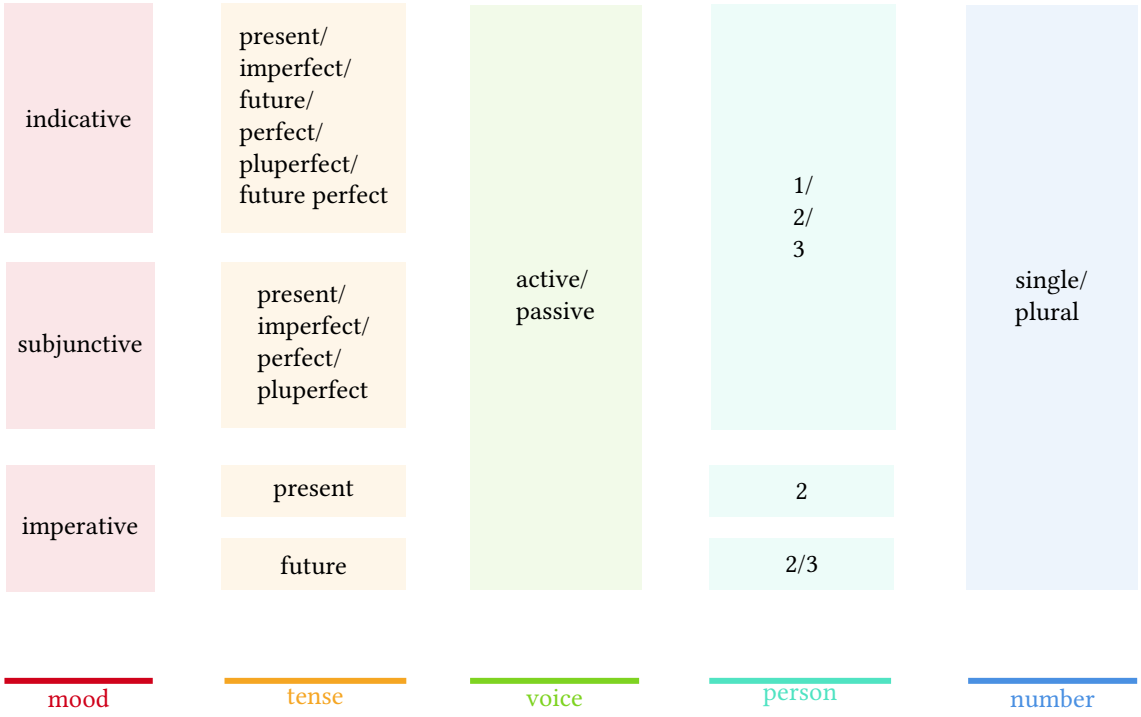

Figure 7.2: Categories in the finite paradigm

Box 7.4: Recording verb inflectional forms

Different people use the term *verb forms* – and count them – in different ways. The most generous – and the most syntactically relevant – way is to view the realization of every possible CP-TP-*v*P projection as a form of the main verb – the verb root at the core of the CP-TP-*v*P domains. This results in a paradigm in traditional grammar, essentially the traditional way to enumerate Latin verb forms ("the indicative active present second person form of a verb is …").

The problem with this approach is sometimes two cells in the paradigm are always identical. In this way, morphosyntactically there are indeed two different paradigm cells, but morphophonologically there is only one verb form. Take English as an example: a traditional grammar may say "the present subjunctive first person singular of English *take* is *take*". The problem

here is the present subjunctive first person singular *clause* always contains the same form of the *verb* with a present indicative first person singular clause, so it makes no sense to talk about "the present subjunctive first person singular *verb form*". A linguist stingy with the number of verb forms may then stipulate that conjugation forms are literally about *forms*, and thus there is no such thing as "the subjunctive form" of English verbs, because in subject *clauses*, the main verb always has the same form as the infinitive (Huddleston and Pullum, 2002, p. 76).

Another problem with this approach occurs when dealing with languages like Japanese. There are so many suffix slots, and the boundaries between suffixes are relatively clear, so the paradigm is too big to be displayed as a whole and too regular to be enumerated cell by cell. In this case, recording suffixes may be a better choice.

The analysis of conjugation forms of the verb, theoretically speaking, is more about vocabulary insertion and readjustment rules, instead of the syntax proper. This is an instance of the *separation principle*: morphophonological features can be separated from morphosyntactic features (Embick, 2000). Distinguishing between verb forms and clause categories isn't just a game about wording: in periphrastic conjugation, we have auxiliary verb(s) plus a non-finite verb form, but here the non-finite verb form is just the spellout of several features together with the verb root and is definitely not thea head of a non-finite *clause*: what we have here is one clause, not clause embedding. Thus it makes no sense to say "we use a non-finite verb form in a periphrastic construction", because finiteness is a category of a clause, and here is no clause combining. This is also relevant for surface-oriented descriptive linguistics: Huddleston and Pullum (2002, p. 74,83) rejects the notion of the *infinitive form* of the verb, and replace the term by *default form*, because the so-called infinitive form also appears in the subjunctive mood or the imperative mood. Despite this, to respect the tradition, I will still use the term *non-finite verbs* or wordings like "the perfect passives are formed by attaching forms of copula to the perfect participle".

The generous paradigm-cell-as-verb-form approach fortunately works in Latin because Latin is morphologically rich and thanks to historical changes, the boundaries between suffixes marking each grammatical category are already vague enough, so the Japanese School Grammar approach is also not applicable. So it does make sense to talk about "the indicative active perfect second person singular form" of a verb. Similarly, we also talk about non-finite verb forms (§ 7.2), though strictly speaking, finiteness is a category of the clause.

7.1.6 Overview of conjugation classes

The way realization of the paradigm for a verb may be divided into four conjugation classes (§ 7.4), and there are also deponent verbs (§ 7.6) and irregular verbs (§ 7.7).

7.2 The non-finite paradigm

According to the morphology, Latin non-finite verb forms can be classified into the infinitives (§ 7.5.1) and the nominal forms (§ 7.5.2), the latter having noun-like or adjective-like morphology. Non-finite verb forms don't agree with the subjects they take, so there is no number or person category marked on them in the same way as Fig. 7.2, though for nominal verb forms there are number and person categories marked in the same way as the nominal morphology.

The infinitives include present active, present passive, perfect active, perfect passive, future active, and future passive infinitives. The latter three are realized periphrastically (Fig. 7.3).

The nominal verb forms include the **simple active**, the **perfect passive** (often just called the perfect participle), and the **future active** participles, the **gerund**, the **gerundive** which is also known as the **future passive** participle, and two supine forms. The **first supine** is identical in the form to the singular neutral accusative perfect participle, without any reference to the number category of any argument it takes. The **second supine** is identical to the singular neutral ablative or dative past participle, also with no inflection with respect to the number category of any argument it takes.

Box 7.5: Whether to keep supine as a verb form

The idea of the stingy linguist may lead one to reject the notion of supine in Latin grammar (Box 7.4). However, for the same reason the infinitive (or the "plain form", since the infinitive is actually a label of clauses – see the discussion and the separation principle in Box 7.4) is recognized as a form independent from the present form in English in Huddleston and Pullum (2002, p. 74), the status of supine as a separate form is recognized in this note. The reasons include TODO

In Classical Latin, the gerund and participle forms are significantly more noun-like than their counterparts in English, and this also justifies the term *nominal form*, because they are not far from prototypically nominalization: although they are still modified by adverbs, they are unable to take arguments. In Ecclesiastical Latin, the so-called nominal forms are more verb-like (TODO: ref), being able to take arguments, and are therefore no longer "nominal".

7.3 Formation of stems

7.3.1 The three verb stems

Box 7.6: About the concept of stem

What's a stem Prototypically, the verb conjugation in a language is described by a series of morphological devices that take *the* verb stem as the input, and give conjugated verb forms as the final product. This is indeed the case for Latin nouns (§ 4.3) and for English regular verbs: the infinitive form is taken in, and third-person singular *-s*, past tense *-ed*, past participle *-ed*, and the gerund-participle *-ing* are attached according to the syntactic environment. Sometimes the process is a little more irregular but not that irregular: *several* stems can be identified, each of which is fed into different morphosyntactic machines. In other words, we have irregular stem alternation. Again, for English irregular verbs, there are three stems: the infinitive stem (e.g. *go*), the preterite stem (e.g. *went*) and the past participle stem (e.g. *gone*). The step to feed stems into morphosyntactic machine is irregular, but everything else is regular: irregular, in this case, does appear, but it appear *regularly*: it only appears in certain parts. The irregularity of stem alternation is so prevalent that if the conjugation paradigm of a verb can be described with a few stems, the verb is deemed as regular, despite the fact that such verbs are obviously irregular by the standard of English.

Stem alternation as mini conjugation classes This phenomenon – that a verb has more than one stem, i.e. irregular stem alternation – is frequent cross-linguistically (Jacques 2021 § 12.2, Forker 2020 § 11.2, among others). Usually, certain correlation between the stem varieties can still be recognized, and verbs can be grouped accordingly, which, if the linguist truly will, can be (though tediously) summarized as more fine-grained conjugation classes. This is also the case for Latin (§ 7.3).

Stem excluded from primary concepts in morphosyntax The notion of **stems** isn't really essential in the description of morphosyntactic: it can well be modeled by environment-dependent vocabulary insertion rules and/or post-syntactic operations. When certain correlations can still be built between so-called suppletive forms, what happens may be analyzed as in Embick and Halle (2005), where certain stems receive morphophonological readjustment (according to the aforementioned hyper fine-grained conjugation subclasses). Thus, it's not true suppletion: it's just a corner case of non-concatenative morphology. When these readjustment rules are fossilized, suppletion – like the English *good*/*better* – may just be the result of conditional insertion, as is outlined in Siddiqi (2009).

Strong irregularity in stem alternation usually restricted to light verbs A general tendency about suppletion is truly suppletive verbs are usually light verbs (in the surface-oriented sense), with meanings like 'do', 'come', etc. This may come from the fact that conditional realization of the root – as opposed to grammatical items – is somehow "heavy" and not favorable. When we forbid conditional realization of roots, (Embick and Halle, 2005), unrestricted suppletion can only be the result of vocabulary insertion rules of functional heads; certain degree of suppletion can still be realized by readjustment rules, which however are restricted in their

> computational capacity. Thus, real lexical verbs are highly unlikely to have truly irregular suppletion; if a verb is truly suppletive, then it's highly likely to be the spellout of *v*P functional heads.

Latin shows not completely predictable stem alternation. All forms can be obtained by three stems (Allen and Greenough, 1903, § 164), if the verb is regular:

- The **present stem**, which, after attached with proper endings, forms
  - The PRESENT, IMPERFECT, and future forms, regardless of whether they are indicative or subjunctive, active or passive. (There is no future or future perfect subjunctive).
  - All the imperatives.
  - The present infinitives, active and passive.
  - The present participle, the gerundive, and the gerund.
- The **perfect stem**, which, after attached with proper endings, forms
  - The perfect, pluperfect, and future perfect active, indicative or subjunctive. Again, there is no future or future perfect subjunctive. Note that the passives are *not* formed by the perfect stem.
  - The perfect active infinitive. (Or the perfective infinitive active, since infinitive is considered as a mood by some people.)

  Note that the perfect passive participle is *not* obtained from the perfect stem.
- The **supine stem**, which, after attached with proper endings or used together with proper forms of *sum*, forms
  - The perfect passive participle, which, by being used with proper forms of *sum*, forms
    - The perfect, pluperfect, and future perfect passive forms, indicative or subjunctive. Again, there is no future or future perfect subjunctive. This is periphrastic conjugation: it is done by using proper forms of *sum* with the perfect passive participle.
    - The perfect infinitive passive.
  - The future active participle, which, used together with *esse*, makes the future active infinitive.
  - The future passive infinitive, by being used together with *īrī*.

This process is summarized in Fig. 7.3. In a dictionary, typically the stems are not directly given – which are given are representative verb forms, from which the stems and the conjugation class can be inferred (§ 7.8).

Note that in Medieval Latin, often, instead of *iri* plus the first supine, *fore* plus the perfect participle is used to form the future passive infinitive. TODO: find a reference https://www.nationalarchives.gov.uk/latin/stage-2-latin/lessons/lesson-24-infinitives-accusative-and-infinitive-clause/

7.3.2 Formation of the present stem

7.3.3 Formation of the perfect stem

7.4 The finite paradigm

7.4.1 Tense and mood

The contextual alternation of the tense and mood marker is listed below. Note that the tense and mood marker is also subject to phonological rules (diachronic or synchronic). The following rule is the most important one that apples in verb conjugation: *a long vowel is shortened before* -m, -r, -t, -nt, -ntur. Considering *bāris*, this doesn't look like a synchronic rule?? TODO

- The indicative:

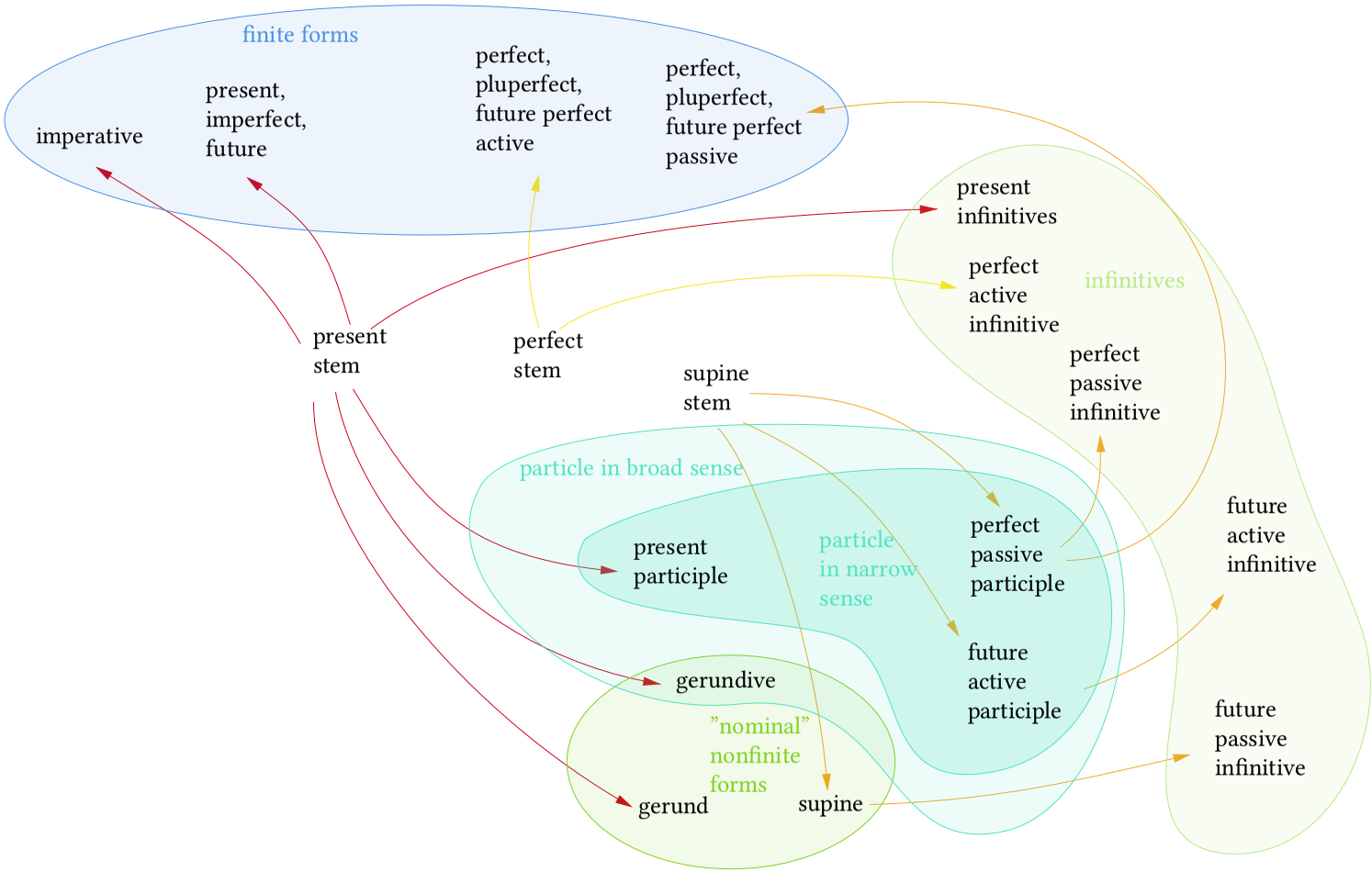

Figure 7.3: How to get all conjugation forms from the three stems

  - Present: zero suffixation, but there is change on the stem-final vowel:
    * For first conjugation verbs, the final *ā* is dropped.
    * For second conjugation verbs, the final *ē* → *e*.
    *
  - Imperfect: *-bā-*, possibly shortened.
  - Future:
    * For first and second conjugation verbs, the tense-mood morpheme is *-bi-*, except for first-person singular (which is *-b-*) and third-person plural (which is *-bu-*).
    * For third and fourth conjugation verbs, change stem-final stem.
  - Perfect:
    * *-ī-*: first-person singular, third-person singular, first-person plural. Shortened for the latter two.
    * *-is-*: second-person singular, second-person plural.
    * *-ēru-*: third-person plural.
  - Pluperfect: *-erā-*, possibly shortened.
  - Future perfect:
    * *-eri-* for all cases except first person singular.
    * *-er-* for first-person singular.

- The subjunctive:
  - Present: no suffixation, but there is regular change on the stem-final vowel:
    * For first conjugation verbs, *ā-* → *ē-*.
    * For second conjugation verbs, → *eā-*.
    * For third conjugation verbs, → *ā-*.
    * For fourth conjugation verbs, → *iā-*.

    These alternations apply for both active and passive verbs, so they have nothing to do with polarity, and this is why I put them in this section.
  - Imperfect: *-rē-*, possibly shortened.
  - Perfect: *-eri-* for all non-periphrastic cases, i.e. active.
  - Pluperfect: *-issē-* for all non-periphrastic cases, i.e. active. Possibly shortened.

> **Box 7.7: Shortening or prolonging?**
>
> At the first glance, it may be attempting as well to consider *-ba-* as the indicative imperfect suffix, since *-bā-* does not outnumber it. However, note that TODO: justification of analyzing *-bā-* as the fundamental form. The same line of argumentation can be applied to justify the status of *-rē-* as the somehow canonical subjunctive imperfect suffix.

7.4.2 The personal ending

Here I list possible personal endings for verbs that are indicative or subjunctive.

- The active:
  - First-person singular:
    * *-ō*: present indicative, future indicative (first and second conjugations only), future perfect indicative.
    * *-m*: imperfect indicative, future indicative (third and fourth conjugations only), pluperfect indicative, subjunctive regardless of tense.
    * *-ī*: perfect indicative.
  - Second-person singular:

        - *-s*: compatible with all tenses and moods, except the perfect indicative.
        - *-tī*: perfect indicative.
    - Third-person singular: *-t*, with all tenses and moods.
    - First-person plural: *-mus*, with all tenses and moods.
    - Second-person plural: *-tis*, with all tenses and moods.
    - Third-person plural: *-nt*, with all tenses and moods.
- The passive:
    - First-person singular:
        - *-r*: compatible with all tenses and moods, except the present indicative.
        - *-or*: present indicative. Also, note that the future indicative (first and second conjugations only) ending is *-bor*, which may be analyzed as *-b-or*.
    - Second-person singular:
        - *-ris*: compatible with all non-periphrastic tenses and moods.
        - *-re*: alternative form of second-person singular compatible with all non-periphrastic tenses and moods. If this personal ending is used, then the tense and mood marking is none. Note that the resulting verb form is the same as the infinitive participle part.
    - Third-person singular: *-tur*, with all non-periphrastic tenses and moods.
    - First-person plural: *-mur*, with all non-periphrastic tenses and moods.
    - Second-person plural: *-minī*, with all non-periphrastic tense and moods.
    - Third-person plural: *-ntur*, with all non-periphrastic tenses and moods.

7.4.3 Periphrastic conjugations

7.5 Non-finite forms

7.5.1 The infinitives

7.5.2 The gerund and participles

7.5.2.1 The gerund

Note that the nominative case is missing – when a non-finite clause is required in the subject position, it's always an infinitive.

7.6 Deponent verbs

> **Box 7.8: Distributed Morphology of deponent verbs**
>
> Embick (2000) analyzes deponent verbs as roots carrying a passive feature themselves. When functional heads higher than the roots are realized, the passive feature – which may come from the root or from the passive light verb – guides the realization of the person and number categories.

7.7 Irregular verbs

7.7.1 The verb *sum*

7.7.1.1 Overview

The verb *sum* has lots of uses in Latin grammar (§ 8.5.3.1), and its inflection is (unfortunately but expectedly) highly irregular. It's also defective: it has no passive forms, either finite or nonfinite. The

principal parts (§ 7.8.1) are *sum*, *esse*, *fuī*, with the supine form being absent – usually replaced by the future active participle *futūrus*.

From the principal parts, we find the perfect stem is *fu-*, and the supine stem – if we insist on defining it – is the same, although the perfect passive participle is absent and so is the supine, and therefore the supine stem only appears in the future active participle.

The present stem is not well-defined: the second principal form *esse* doesn't have the regular infinitive ending *-re*, though we can roughly recognize something like *es-* or *e-*; the first principal form *sum* gives *su-* or *s-*. The two stems appear in the finite paradigm in an unpredictable manner, also with irregular though still recognizable endings. Besides *s-* and *es-*, there is also *fo-* seen in one variant of the future active infinitive (§ 7.7.1.2), which also appears in variants in the subjunctive active imperfect part of the finite paradigm.

7.7.1.2 The nonfinite paradigm

The only nominal form is the future active participle *futūrus*. The three active infinitives forms are all attested. The present active infinitive is *esse*. The perfect active infinitive is *fuisse*, regularly formed by the perfect stem *fu-*.

The future active infinitive can be regularly formed by adding *esse* to the future active participle, and therefore is *futūrum esse*. There is also a free variant *fore*.

7.7.1.3 The perfect system

The perfect forms – finite forms and the perfect active infinitive – of *sum* can be formed regularly (§ 7.4) according to the perfect stem *fu-*.

7.7.1.4 The imperative system

The present imperative system, which is known for reflecting the present stem, is formed regularly using *es-*: the singular second person present imperative is *es* and the plural second person present imperative is *este*.

7.7.1.5 The present system

The imperfect forms of *sum* are highly irregular, though patterns can still be found. In the indicative part (Table 7.3):

- The PRESENT forms show no pattern except the personal endings. Note that here *-m* instead of *-ō* is used for the first person singular form.
- The IMPERFECT forms are formed by adding the standard personal endings (*-m*, *-s*, *-t*, *-mus*, *-tis*, *-nt*) to *erā*, the vowel *ā* of which undergoes shortening according to rules in § 7.4.1.
- The FUTURE forms are formed by the same personal endings seen in the first and the second conjugations, although the tense marker isn't the same: the stem-tense marker complex is *er-* instead of the stem plus *-b-*.

Table 7.3: The indicative paradigm of *sum*

PRESENT	IMPERFECT	FUTURE
sum	*eram*	*erō*
es	*erās*	*eris*
est	*erat*	*erit*
sumus	*erāmus*	*erimus*
estis	*erātis*	*eritis*
sunt	*erant*	*erunt*

In the subjunctive paradigm (Table 7.4), we find that in the PRESENT system, the stem-tense marker complex is fused into *sī-*, and in the IMPERFECT system, the stem-tense marker complex is fused into

essē- or *forē-*, both of which are then attached to the standard *-m*, *-s*, etc. personal endings, and the vowel shortening rule in § 7.4.1 works.

Table 7.4: The subjunctive paradigm of *sum*

PRESENT	IMPERFECT
sim	*essem*, *forem*
sīs	*essēs*, *forēs*
sit	*esset*, *foret*
sīmus	*essēmus*, *forēmus*
sītis	*essētis*, *forētis*
sint	*essent*, *forent*

7.7.2 The verb *faciō*

The verb *faciō* looks pretty regular regarding the endings, except for one thing: its stem alternates according to the voice.

7.8 Guide for parsing and using Latin verbs

7.8.1 Principal forms and stems

In practice, the three stems aren't what stored in the dictionary, for two reasons. First, for fluent users, recording actually attested word forms is easier compared with the morpheme-based and "anatomized" approach. Second, Latin has four conjugation types, and hence the three stems themselves aren't sufficient to decide how to conjugate the verb: more information is needed, and by storing already conjugated verb forms, the conjugation class can be decided by observing the endings. What are stored are the following **principal forms**, from which the three stems and the conjugation class can be solved out (Allen and Greenough, 1903, § 172):

1. *The first-person present active indicative*: formed from the present stem.
2. *The present infinitive*: formed from the present stem. By observing its ending, the conjugation class can be decided, and by comparing with the first principal form, the present stem is obtained.
3. *The first-person perfect active indicative*: showing the perfect stem.
4. *The neutral accusative past participle*, i.e. the form of supine: showing the supine stem.

The ways to obtain the stems from the principal forms are:

- *The present stem* can be found by dropping *-re* in the PRESENT INFINITIVE (Allen and Greenough, 1903, § 175).
- *The perfect stem* can be found from the third principal part: just remove *-ī*.
- *The supine stem* can be found by dropping *-um* in the supine i.e. the fourth principal form (Allen and Greenough, 1903, § 178).

7.8.2 Constructing non-finite forms

7.8.2.1 Nominal forms

- *The present active participle (i.e. the present participle)*: replace the *-re* ending of the present active infinitive by *-nt* (or in other words, add *-nt* to the present stem) and the result is the nominative.
- *The perfect passive participle (i.e. the perfect participle or the past participle)*: this can be found by declining the neutral accusative past participle, i.e. the fourth principal part.
- *The future active participle (i.e. the future participle)*: add *-turus* to the supine stem.

Chapter 8

Verb phrase

8.1 Introduction

This chapter gives an overview of clausal dependents, especially about the mapping from purely semantic argument roles to clause dependent slots. This chapter is mainly about verbs that don't take complement clauses as arguments. The phenomena discussed in this chapter mostly apply to complement clause constructions as well, but complement clause constructions have their own peculiarities (§ 10.3.1).

A list of semantic classification of verbs – and hence valency classes – can be found in Dixon (2005, Part B), Dixon (2010, § 18.5), and Dixon (2009, § 3.3). Dixon classifies the verb class into three subgroups: 1. Primary-A, which contains verbs that don't take arguments with meanings similar to those of complement clauses, 2. Primary-B, which are semantically **complement-taking** (Box 8.2) and **lexical**, which have arguments that are semantically equivalent to complement clauses (but not necessarily syntactically coded as complement clauses) and have meanings more complicated then what's expected for grammatical items, and 3. Secondary, members of which have the same *meaning* of certain grammatical constructions in the verbal system, but not the same syntactic properties (for example, they may just take complement clauses instead of being auxiliary verbs).

Allowed combinations of clausal dependents are determined by the valency class of the verb and how it engages with valency changing devices, which is strongly related to but is not determined by the semantics of the verb. In English, for example, *I like this* where *this* can represent an event is semantically complement-taking but involves no complement clause construction syntactically; and it's possible to leave out a semantic argument in the syntactic frame of a verb. Thus, research mainly focused on *syntactic* properties of arguments is also needed.

> **Box 8.1: A-positions**
>
> In generative syntax, we say the positions in the syntactic frame of a verb are **A-positions**. To achieve a more disciplined analysis, we may adopt a multiple-step analysis of A-positions: at least two steps – the *v*P step and the TP step – are to be distinguished, and in both steps there are sub-steps.
>
> **A-positions and semantics** I should focus again that arguments with same semantic roles may appear in different A-positions, and one A-positions may host various semantic roles. A Secondary verb is different from, say, an auxiliary verb in the verbal system in the eye of syntax and at the syntax-semantic interface: the lexical verb *start*, for example, introduces a new event besides the event that the agent is start to do, while an inchoative aspect, if in the grammatical aspect region and not the lexical aspect region, reflects the speaker's attitude (possibly by shifting the time referred to to the initial part of the whole situation), but they are of course equivalent to each other, although their interpretations immediately at the syntax-semantic interface are different. On the other hand, the subject A-position may host the participant in an event that initiates the event or it may just be an experiencer.
>
> One thing to note, however, is that if we do coarse-graining to the various verb-specific semantic roles and compare them with positions within the *v*P, we *do* find a strong correlation, known as the Uniformity of Theta-Assignment Hypothesis (UTAH). This means the coarse-grained semantic roles (things like "agent", "theme", etc., as opposed to "Manipulator" or "Tar-

get") are syntactic objects as well, as they label the positions in *v*P; this creates a confusion of the meaning of the term *semantic role*; in this note both types of semantic roles are discussed, but usually the meaning of the term *semantic role* can be inferred from the context.

A-positions in an accusative language The first step is roughly a "translation" of the semantic argument roles into their syntactic counterparts, according to the so-called Uniformity of Theta Assignment Hypothesis. (There are subtleties in topics like whether the experiencer roles in psych-verbs as in *he fears the police* and *the police frightens him* are actually different in the *v*P step, or they are the same in the *v*P step but something else decides which argument becomes the subject.) Assignment of the accusative case is said to be done within the *v*Pregion. Then comes the TP step, in which usually the highest argument position in the *v*P step – also the most agentive one – becomes SpecTP, which is better known as the subject.

Non-trivial correspondence between derivational steps in clausal syntax [a] The **argument structure** (argument labels in which are "agent", "patient", etc., i.e. the type of semantic roles bearing direct syntactic significance discussed above) is the structure of the "canonical" *v*P containing a verb root. For roles in the argument structure, see, for example, Huddleston and Pullum (2002, § 4.2), which lies around the syntax-semantic interface and are actually syntactic categories. On the other hand, the **clausal complement types** like subject, object, etc. are largely decided by what happens after TP is finished. Lots of things can happen between the two. The most well known case is passivization, where although we have evidences that the agent argument still occupies a higher position, it receives an inherent case (quite similar to how DPs are licensed by prepositions) and is thus unable to move to the subject position. It's therefore worthwhile to talk about "deep" roles and "surface" roles – although they are defined as S, A, P, etc. and are known as generalized semantic roles, these are *syntactic* labels, not semantic ones. For unaccusative verbs, the agentive argument may be absent, and the deep P argument becomes the surface S argument.[b] Thus, in a truly well organized grammar, we need to first study deep syntactic argument slots and then surface syntactic argument slots.

Ergativity and split of grammatical relation labels It should be noted that labels for A-positions, like *subject*, implicitly means there is an argument which is spontaneously in several syntactic functions (in the case of *subject*, it's the external argument which governs all internal complements, the receiver of the nominative case, and, say, SpecDoP, where Do or something else is the highest light verb in the *v*P field). When these syntactic functions are disassembled with each other, the corresponding collective label no longer makes sense. Thus in a morphological ergative language, the "nominative case receiver" syntactic function is absent, and the A argument of a transitive verb receives an inherent case, so the A argument loses the morphological resemblance with the S argument for intransitive verbs, but it keeps the syntactic functions of the highest argument; for syntactic ergativity, the external argument property is given to the P argument (Aldridge, 2008). In both cases, grammatical relations condensed into the label *subject* need to be taken out one by one and redistributed to new collective terms.

[a] Here *derivational* means what it means in modern generative grammar, that grammatical structures are built by successive applications of (possibly internal) Merge. It *doesn't* mean *transformational operations* in early versions of generative grammar.

[b] Note that unaccusativity has nothing to do with alignment: we can have unaccusative verbs in a typical accusative language, like English.

8.2 Core, oblique and peripheral arguments

This section examines complement and adjunct positions in Latin clauses. Parameters used in the classification include their correspondence with semantic roles (agent, patient, source, etc.), their internal structures (case, preposition, NP or clause, etc.), agreement, and their behaviors in valency alternation.

Box 8.2: The term *complement*

In Huddleston and Pullum (2002) the term *complement* means A-positions mentioned in Box 8.1. Thus the subject, several kinds of objects, the copular complement (Huddleston and Pullum (2002) calls it *predicative complement*) are all complements, and they are labels implying several grammatical relations: the label *subject* roughly corresponding to "SpecTP" or "what is in *v*P

and receives the object case from a high light verb".

In traditional Latin grammar, however, *complement* means the copular complement. This note follows the terminology used in most descriptive grammars, so use the term *copular complement* to refer to the CGEL *predicative complement*. Also, as is seen in § 8.1, the term *complement-taking verb*, despite being confusing, is used to refer to a verb that take a complement clause or something semantically equivalent to a complement clause as one of its arguments.

Beside the subject, various types of objects, and copular complements, there are more clause dependents corresponding to less frequently seen semantic roles like purpose, direction, location, etc. They may be licensed or even required by the verb (**oblique argument**), or they may be modifying the whole clause and therefore are usually optional (**peripheral argument**, or "adjuncts"). Besides clauses and NPs (with or without prepositions), their categories also include adverb phrase (AdvP)s, a majority of the latter originating from case forms.

A clear complement-adjunct distinction – telling peripheral arguments from core arguments or oblique arguments – is hard to establish in Latin. Latin peripheral arguments do not necessarily have prepositions. Latin is highly free-ordered and therefore all clause dependents can leave their base positions. Latin is also highly *pro*-drop, and even uncontroversial core arguments can be omitted. Oblique arguments are frequent in Latin, as is the case in English (consider *run away from* or *get into*). Thus criteria of category, position, and argument in Huddleston and Pullum (2002, § 4.1.2) all fail to work. Latin doesn't have systematic way to replace the core predicate (i.e. without adjuncts) by an anaphora, and that criterion does not work, either. TODO: really? The remaining criteria are about selection, licensing, and obligatoriness; these criteria are however hard to use for a classical language. Thus, despite I'm fully aware that clausal dependents concerning place, instrument, mean, etc. may be licensed by both the argument structure of the verb and by clausal adjunct positions and may have clear structural differences in other languages (as in English), currently no distinction between the two cases is made.

Box 8.3: How to document complement types

The traditional practice of Latin grammar research is to classify clausal complement and adjunct types according to their case marking. This strategy is also found in modern grammars. Some introduce clausal complement types just in chapters about case marking (Jacques, 2021, chap. 8), while other grammars, despite giving a brief description of the context of case marking, spare some time to discuss complement types in the chapters about valency and clause structure (Forker, 2020, § 3.4, chap. 19, chap. 22). From a TAG perspective (§ 1.3), the two extremes are different in how they treat function labels: in the former, the function label of a construction appears together with its category label on the root node, while in the latter, the function is described separately from the form of what fills that position. The former is more bottom-up, while the latter is more top-down. The choice between the two, however, is usually language-dependent: grammars for analytic languages, of course, have to lean even further to the "complement type as clause slot" extreme and away from the "complement type as case-form context" extreme. Allen and Greenough (1903) uses a hybrid method: the discussion about case marking (§ 39) is separated from the discussion about complement and adjunct types (§ 338), so the top-down approach seems to be adopted, but the latter is still arranged in terms of case. This arises both from the distinct features of Latin and the intended readers: the relation between complement types and cases is regular enough in Latin, and what is most important for Latinists is to understand, at least sketchily, ancient writings, so a parsing-oriented grammar is much handier.

8.2.1 The subject

Latin is an accusative language. A **subject** can be identified for all clauses, though it is frequently omitted. Grammatical behaviors of the subject are summarized in the following list:

- *Coding of semantic role*: In an active clause, the subject is always the most agentive argument, i.e. the S argument in a prototypical intransitive argument structure and the A argument in a prototypical transitive argument structure. In a passive clause, the subject corresponds to the "promoted argument" (§ 8.3).

- *Case marking*: Subjects are always nominative for finite clauses, whenever the case system is in action, i.e. whenever the subject is an NP or a gerund. Nonfinite clauses may be argued to be subjectless in the surface form (a reasonable claim, since they have deficient TP layers, and hence it is possible that no canonical subject position exists), but in accusative-infinitive constructions, the accusative may be seen as the non-canonical subject of the infinitive.
- *Agreement*: the number and person features on the subject leave marking on the verb complex. Latin does not have verbal agreement with arguments other than the agreement with the subject.
- *Category*: a subject is an NP or a complement clause (§ 10.3.1), usually an infinitive but never a gerund (§ 7.5.2.1). This constraint isn't seen in any other clausal complement types.

8.2.2 The direct object

Here is a list of grammatical properties of the direct object:

- *Coding of semantic role*: In a prototypical transitive argument structure, the direct object is the P argument, i.e. the most patientive argument.
- *Case marking*: Direct objects are always accusative when it makes sense to talk about case – but not all accusative arguments are direct objects (§ 4.4.5.2).
- *Passivization*: If an argument is coded as the direct object, then it can regularly be promoted to the subject position in a passive clause (§ 8.3). Secondary objects are less frequently promoted in passivization (§ 4.4.5.2).

8.2.3 The indirect object and the secondary object

Latin also has two complement positions named as object: the indirect object and the secondary object. The indirect object is distinguished by the following grammatical properties:

- *Coding of semantic role*: in a AGT-type argument structure, the indirect object is usually the G argument. Intransitive clauses sometimes also have indirect objects, and an indirect object, in this case, is also a G argument.
- *Case marking*: indirect objects are always dative.
- *Passivization*: indirect objects are always retained in passive clauses. They are never promoted to subjects in passivization.

The secondary object is distinguished by the following grammatical properties:

- *Coding of semantic role*: in an AGT-type argument structure that is always about information flowing, the T argument (i.e. the thing asked about or taught about) is the secondary object. The G argument (i.e. the person who is asked or taught) is the direct object. Sometimes the G argument is ablative, and in this case, there is only one accusative argument: the secondary object. Another place where secondary objects appear is clauses headed by a verb with a compounded accusative preposition.
- *Case marking*: secondary objects are always accusative.
- *Passivization*: secondary objects can be passivized, but much more rarely than direct objects.
-

The distributions of the secondary object and the indirect object are mutually exclusive. This means for ditransitive verbs of type GIVING, Latin shows a clear and strong tendency to identify the T argument with the monotransitive O, while for ditransitive verbs about teaching, the inverse is true.

> **Box 8.4: Comparison with English**
>
> It can be found that the Latin indirect object has more similarity with the English *to*-PP, which is also called the indirect object in some grammars, but not CGEL. The Latin indirect object differs from the English (accusative) indirect object in passivization. Since in Latin, verbs with

AGT-type argument structure do not have alternation of complementation pattern – in English we have *give sth. to sb.* and *give sb. sth.*, while in Latin there is only the former one, but *to sb.* is replaced by a dative, (always with no preposition) – the G argument is identified with the E argument, and the T argument is identified with the P argument. In other words, in Latin, there is only the *John gave [goods]$_T$ to [charity]$_G$* pattern: the double-object *John gave charity goods* pattern is absent.

Therefore, for typical ditransitive verbs, i.e. verbs like *give*, Latin shows a clear and strong tendency to identify the T argument with the monotransitive O, which is more typical than English[a], but for verbs with meaning of TEACH or ASK, there is also a clear and strong tendency to identify the G argument with the monotransitive O. The term *secondary object* is coined to cover this grammatical position.

[a] In English, in the *give sb. sth.* construction, it is the person i.e. the G argument that is passivized, while the T argument i.e. *sth.* cannot, though the latter is identified with monotransitive O according to other criteria.

8.2.4 Copular complements

Latin also has copular complements. A copular complement, just like its counterpart in English, basically can be viewed as a displaced attributive or appositive (and hence is prototypically filled by an NP or an AdjP) but is a little more peripheral (manner, state, factitive, etc.) in its meaning than an attributive or appositive.

Latin has nominative predicate and accusative predicate: as hinted by their names, the nominative predicate gives a property of the subject and agrees with it, and the accusative predicate gives a property of the direct object and agrees with it. In passivization of the direct object, the accusative predicate becomes the nominative predicate.

Other types of copular complements without agreeing with the subject exist. TODO: ablative of quality, price, etc. The syntactic status of copular complements here are closer to PPs: we may say they receive *inherent cases*, while the nominative and accusative copular complements receive *structural cases* (§ 4.4.5).

8.2.5 (Change of) location

TODO: considering moving this section to the case section

Various semantic roles can be summarized as SOURCE, and the source clausal dependents – adjunct or complement – have the following properties. Note that we are dealing with a *group* of clausal dependents.

- *Coding of semantic role*: A source argument can be the position from which an object moves (**ablative of source**) or the source in a separation event (**ablative of separation**: 'remove', 'deprive'), or the place where something comes into being (**ablative of material**, 'birth', 'origin'), or the cause of something ("the source of the event", **ablative of cause**); the agent in the passive voice possibly comes from one of the figurative use of the ablative as well.
- *Case marking*: a source argument is in the ablative case. It may come together with the prepositions *ex* or *ab*.
- *Passivization*: not available.

8.2.6 Others

There are other clausal dependents with semantic roles and case/preposition markings different from any other type mentioned above; they are skipped here for brevity. A full list of these clausal dependent slots can be found by checking the usage of each preposition.

8.3 Passivization

8.4 Preverbs and other verbal derivations

8.5 Verb frames

I will generally follow the classification in Pinkster (2015, chap. 4); some verb classes enumerated by Pinkster are discussed in more details in § 10.3, since a full account of their behaviors is closely related to the structure of complement clauses.

8.5.1 Prototypical intransitive verbs

8.5.1.1 The MOTION type

8.5.2 Prototypical transitive verbs

With complement-taking verbs temporarily excluded, a prototypical transitive verb is more or less close to the AFFECT type, with an A argument which is the causer of the event

8.5.3 Copular verbs

8.5.3.1 The verb *sum*

It's also possible to use *sum* with an indirect object, and the meaning because 'something be to [someone]$_{\text{indirect object}}$'. In this case we get the possessive dative construction (Allen and Greenough, 1903, § 373).

Chapter 9

Clause structure

9.1 Small clauses

accusative in *Deo gratias*

9.2 Constituent order and the information structure

Constituency tests reveal there is a VP unit in Latin (Danckaert 2017, § 1.6; TODO: ref to my own analysis in conjunction), and the *non*-before-auxiliary constraint (Danckaert 2017, § 1.5; TODO: my own ref) also means there are This means Latin

9.3 Clause combining

Box 9.1: Types of coordination and subordination

It's hard to draw a line between coordination and asymmetric (i.e. subordinating) clause linking (like concessive clauses). Theoretically, this is because any clause combining construction follows the X-bar scheme: one clause is the Specifier, and another clause is the Complement, and certain asymmetry has to be introduced. In English, the FANBOYS – *for*, *and*, *nor*, *but*, *or*, *yet*, *so* – are usually regarded as coordinating conjunctions. But what's the essential difference between *although* and *but*?

On the other hand, adverbial clause constructions are uncontroversially asymmetric and can in theory be distinguished from clause linking: in clause linking, the less important clause is base-generated in one Specifier position in the CP layer of the main clause, so the two combined units are of roughly the same structure, while adverbial clauses appear in the TP layer, so the two combined units are of different structures: the adverbial clause is a CP, while the main clause, when the adverbial clause enters derivation, is a TP. Complement clauses, on the other hand, are first introduced in the *v*P layer: they are TPs or CPs, while the main clause, when complement clauses enter the derivation, are *v*Ps. But there are still certain subtleties regarding the boundaries of *v*P, TP, and CP.

Relative clauses are introduced in DPs, so the probability to confuse a relative clause construction with a complement clause construction is small – but still not zero. It can be expected that *I like the man dancing* and *I like the dancing man* are realized in quite similar ways. Besides, some languages lack prototypical complement clause constructions but have complementation strategies. That is, when they talk about *I like the dancing man*, a speaker of such a language may be implying that he or she actually likes the man's dancing, though not the man's personality. Now comes the question: when there are vague evidences indicating the grammaticalization of this construction, should we now claim the language has already developed a complement clause construction?

It's still possible to do the same thing – largely symmetric coordination and certainly asymmetric subordination – completely with *v*Ps. The former results in clause chaining (Nonato, 2014), while the latter results in serial verb constructions. These construction types, however,

> are absent in Latin, and I will not go deep into them in this note.

In Latin there is no serial verb constructions. Subordination strategies can be neatly summarized into complement clauses, relative clauses and adverbial clauses.

Chapter 10

Clause combining

10.1 Clause linking and conjunctions

10.2 Adverbial clauses

10.3 Complement clause constructions

10.3.1 Overview

10.4 Relative constructions

10.4.1 Agreement properties

The case of a relative pronoun is determined by its syntactic position in the relative clause, and *not* the case of the antecedent, though the number and gender categories are determined by agreement with the antecedent.

Chapter 11

Examples of texts

Below are some examples of Latin texts, in an order from the easiest to the hardest, with remarks on their vocabulary and grammar.

11.1 Liturgy texts

11.1.1 Short formulae in the Roman Mass

Examples in this section are short formulae found in the Roman Mass in the order of their appearance. In (1, 2), *nomine* and *patris* are third third declension nouns, while *spiritus* is a fourth declension noun.

(1) In Nomine Patris, et Filii, et Spiritus Sancti.
in name-SG.ABL Father-SG.GEN and Son-SG.GEN and SPIRIT(M)-SG.GEN holy-M.SG.GEN
'In the name of the Father, and of the Son, and of the Holy Spirit.'

(2) – Dominus vobiscum. – Et cum spiritu tuo.
Lord(M)-SG.NOM 2PL.ABL and with spirit(M)-SG.ABL your-M.SG.ABL
'– The Lord be with you. – And with your spirit.'

(3)

11.1.2 Nicene Creed

(4) Credo in unum Deum, Patrem omnipotentem,
believe-IND.PRES.1SG in one-M.SG.ACC God(M)-SG.ACC father(M)-SG.ACC omnipotent-M.SG.ACC
'I believe in one God, (the) omnipotent Father,'

(5) factorem caeli et terrae, visibilium omnium et
maker-
'maker of'

11.2 Vulgate bible

11.2.1 Excerpts in John 1

(6) in principio erat Verbum et Verbum erat apud Deum et Deus erat Verbum
in be.IMPF
'In the beginning' (John 1:1)

(7) omnia per ipsum facta sunt et sine ipso
all-N.PL.NOM through DEM-ACC make.PPRT-N.PL.NOM be.IND.PRES.3PL and without DEM.ABL
factum est nihil quod factum
make.PPRT-N.SG.NOM be.IND.PRES.3SG nothing.NOM REL.N.3SG make.PPRT-N.SG.NOM
est
be.IND.PRES.3SG

'All have been made through exactly this (i.e. the Word of Lord), and without exactly this, nothing that has been made has been made.'

As an example, below I show how (7) can be parsed. First we can see a *et* dividing the sentence into two branches.

1. For the first branch, we know *omni-* is a quantifier meaning *all*, and morphologically it's a twin-termination third declension adjective; then from Table 4.1 and the fact that we are dealing with a third declension word, the ending *-a* means neutral and PL.NOM/ACC/VOC. The vocative case is of course impossible here.

2. The word *per* is a preposition taking an accusative object. *Ipsum* is a basic identity demonstrative, with the meaning of "exactly this". Since it follows *per*, the ending *-um* here seems to be the accusative case marker, instead of a neutral nominative case marker.

3. The sequence *facta sunt* contains the indicative perfect 3pl copula *sunt*, and in *facta*, we see the supine stem *fact-* of the verb *faciō* 'make'. The second fact means *facta* should be the perfect passive participle in a certain inflection form. Then *facta sunt*, collectively, is the indicative passive perfect 3pl form of *faciō*. (Here we are fortunate: it's possible that *facta* and *sunt* get scattered to different places.) The *-a* ending can again be looked up for in Table 4.1: the possibilities are PL.NOM/ACC/VOC – note that the first declension singular possibilities are excluded by the fact that *sunt* is in plural form. We expect *facta* to be nominative because it has to agree with the subject, which is always nominative and it turns to be possible.

4. Now we should link things together. The open ends are: the case of *omnia*, and the (3pl) subject of *facta sunt*. Then quite obviously, we find *omnia* should be in the subject position, and therefore everything works well.

5. We can also check gender agreements to make sure our reading is correct.

The second half is done in similar manners. The structure of the text looks like this:

[[omnia]$_{\text{subject}}$ [per ipsum]$_{\text{instrument:PP}}$ [facta sunt]$_{\text{verbal complex}}$]$_{\text{coord}}$ et [[sine ipso]$_{\text{adverbial:PP}}$ [factum est]$_{\text{verbal complex}}$ [nihil [quod factum est]$_{\text{rel}}$]$_{\text{subject}}$]$_{\text{coord}}$

11.3 Aeneid

11.3.1 Introduction

(8) Arma virumque cano, Troiae qui
weapon(N)-PL.ACC man(M)-SG.ACC=and sing-IND.PRES.1SG Troy-SC.GEN REL.M.SG.NOM
primus ab oris Italiam, fato profugus,
first-M.SG.NOM from shore(F)-PL.ABL Italy(F)-SG.ACC fate(N)-SG.ABL exiled-M.SG.NOM
Laviniaque venit litora, multum ille et terris iactatus et alto vi
Lavinia-TODO=and go.to-IND.PRES.3SG shore(N)-PL.ACC
superum saevae memorem Iunonis ob iram; multa quoque et bello passus, dum conderet

urbem, inferretque deos Latio, genus unde Latinum, Albanique patres, atque altae moenia

Romae.

'I sing weapons and a man, who was the first from the shores of Troy to Italy, was by fate exiled, and '

In (8), it should be noted that *arma* is in plural only. The *qui* clause is an example of the rule that the relative pronoun doesn't agree in case with the antecedent (§ 10.4.1). The copula is omitted in the *qui* clause.

References

Edith Aldridge. Generative approaches to ergativity. *Language and Linguistics Compass*, 2(5):966–995, 2008.

Joseph Henry Allen and James Bradstreet Greenough. *Allen and Greenough's New Latin grammar for schools and colleges: Founded on comparative grammar*. Ginn, 1903.

Lieven Danckaert. Studying word order changes in latin. *Perspectives on historical syntax*, 169:233, 2015.

Lieven Danckaert. *The development of Latin clause structure: A study of the extended verb phrase*, volume 24. Oxford University Press, 2017.

RM Dixon. Basic linguistic theory. volume 3: Further grammatical topics, 2012.

Robert MW Dixon. *A semantic approach to English grammar*. OUP Oxford, 2005.

Robert MW Dixon. *Basic linguistic theory volume 1: Methodology*, volume 1. OUP Oxford, 2009.

Robert MW Dixon. *Basic linguistic theory volume 2: Grammatical topics*, volume 2. Oxford University Press on Demand, 2010.

David Embick. Features, syntax, and categories in the latin perfect. *Linguistic inquiry*, 31(2):185–230, 2000.

David Embick and Morris Halle. On the status of stems in morphological theory. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4*, 270:37, 2005.

Diana Forker. *A grammar of Sanzhi Dargwa*. Language Science Press, 2020.

Dianne Friesen. *A grammar of Moloko*. Language Science Press, 2017.

Giuliana Giusti and Rossella Iovino. A split-dp hypothesis for latin and italo-romance. *Complex Visibles Out There. Olomouc Modern Language Monographs*, 4:127–143, 2014.

Nadine Grimm. *A grammar of Gyeli*. Language Science Press, 2021.

Rodney Huddleston and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002. doi: 10.1017/9781316423530.

Guillaume Jacques. *A grammar of Japhug*, volume 1. Language Science Press, 2021.

Sylvain Kahane and Nicolas Mazziotta. Syntactic polygraphs. a formalism extending both constituency and dependency. In *Mathematics of Language*, 2015.

Julie Anne Legate. *Warlpiri: theoretical implications*. PhD thesis, Massachusetts Institute of Technology, 2002.

Paul A Morris. Evidence of a configurational structure in meskwaki. *Proceedings of the Linguistic Society of America*, 3(1):64–1, 2018.

Ryan M Nefdt and Giosué Baggio. Notational variants and cognition: The case of dependency grammar. *Erkenntnis*, pages 1–31, 2023.

Patrick M Niedzielski. *Clausal Syntax in the Sumerian Language*. PhD thesis, BA Thesis. Ithaca, NY: Cornell University, 2017.

Rafael Nonato. *Clause chaining, switch reference and coordination.* PhD thesis, Massachusetts Institute of Technology, 2014.

Timothy Osborne, Michael Putnam, and Thomas M Gross. Bare phrase structure, label-less trees, and specifier-less syntax. is minimalism becoming a dependency grammar? *The Linguistic Review*, 2011.

Harm Pinkster. *The Oxford Latin Syntax: Volume 1: The Simple Clause.* Oxford University Press, 08 2015. ISBN 9780199283613. doi: 10.1093/acprof:oso/9780199283613.003.0004.

Gerold Schneider. *A linguistic comparison of constituency, dependency and link grammar.* PhD thesis, Master's thesis, University of Zürich, 1998.

Daniel Siddiqi. *Syntax within the word: Economy, allomorphy, and argument selection in Distributed Morphology*, volume 138. John Benjamins Publishing, 2009.