

Note on Latin Grammar

Jinyuan Wu

December 7, 2024

Chapter 1

Introduction

1.1 The language and the speakers

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 nineteenth century, Latin was still fluently used by scholars and in the Catholic Mass. A decline in the popularity of Latin was observed after that, consistent with the rapid development of English (initially also French and German and sometimes Russian) as the language of rapidly developing natural sciences, largely replacing 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.

Latin today is still revered as *a* classical language that has deeply shaped at least the Western world, but it has long lost the prestige of *the* classical language that every decent individual is to learn once in their life. Still, as an important silhouette of what early Indo-European languages looked like, as one of the earliest languages receiving systematic grammatical analysis, and as a language that considerable deviates from its modern descendants in its grammar, Latin deserves its status as a model organism in modern linguistics.

1.2 Classification and history

1.2.1 Latin in Indo-European family

Latin belongs to the Indo-European (IE) family. Languages most closely genetically related to Latin include Faliscan, a language attested from seventh-sixth century BC, Lanuvian, and Praenestinian. These languages form the Latino-Faliscan family.

The Sabellic family, which consists of Umbrian, Oscan, and South Picene, was also spoken in ancient Italy. The exact relation between Sabellian and Latino-Faliscan is

not completely settled: many postulate an Italic language family with Latino-Faliscan and Sabellian as its two branches.

Above the level of the Italic family, it has been proposed that Latin is genetically related to the Celtic branch.

1.2.2 Old Latin

Pre-Classical Latin had gone dramatic evolution. The earliest texts are still not fully deciphered today and the Romans themselves find it hard to read old agreements they signed with the Carthaginians (Leonhardt, 2013, p. 52). Indeed, late Old Latin texts had already shown changes that are also present in Vulgar Latin, which are however absent in Classical texts, suggesting archaisms in the standardization process which makes the Classical texts looking *older* than they actually are, while late Old Latin texts are *later* in the sense that they reflect innovations in the living language (§ 1.2.3).

Box 1.1: List of Old Latin traits

- -os as ending
- texts?

1.2.3 Classical Latin

The road to standardization The formation of what is known today as “normal” Latin, i.e. the kind of Latin usually taught in Latin classes, follows the general pattern of linguistic standardization. The motivation of the process was closely connected to the formation of new identities, which, in the case of standardization of Latin, was the awareness that the Romans should focus on their own cultures and got some sort of spiritual independence from the Greeks: thus Virgil sought to replace *Odyssey* by his *Aeneid*, and not just to translate the former (Leonhardt, 2013, p. 66). The focus on the new identity triggered the so-called “first congress” phenomenon, in which broad discussion on what variety of a language is to be used and eventually the creation of a new literary language. In Rome, this was done by leading members of the upper society: Cicero mentioned linguistic problems in his discussions on rhetorics, Caesar was worried about irregularities of Latin inflectional morphology and proposed some spelling reforms to make homophonous endings more identifiable, and several other figures composed grammars of Latin (Leonhardt, 2013, pp. 60-61).

Certain traits of Classical Latin are also recognizably deliberately refined. The situation is the most telling when a feature is present in both pre-Classical texts (especially dramas, where the language is conceivably colloquial) and post-Classical texts but absent in Classical Latin, meaning that this feature was artificially suppressed in standardization (but likely had always stayed in oral uses). Instances include missing word final *-m* and *de* + noun phrase in place of the genitive case (Alkire and Rosen, 2010, § 2.4.4, § 8.3.1).

Although Classical Latin underwent deliberate refinements, it was still a true living language in its time, namely around the first century: Cicero was equally Classical in his popular oratory and in his other works, and the fact that public oration was to be understood by as many as possible clearly shows that Cicero’s language was a real

language that was actually spoken by at least Roman elites (Leonhardt, 2013, p. 71). What happens here is comparable to

Canonization and fixing of Latin After the Classical period, Latin in official writings stopped evolution, which is the destiny of all authorial languages and also happened to Greek (Leonhardt, 2013, pp. 71-72). Indeed, if a language first develops a writing tradition and then undergoes considerable change, usually this will lead to the old literature being forgotten by most (if not nearly all) readers (Leonhardt, 2013, p. 144). The study on what happened after a Classical Latin literary canon had appeared therefore had to rely on less known and more indirect sources, as is briefly discussed in § 1.2.4.

1.2.4 Sub-elite Latin

Vulgar Latin as ancestor of Romance languages Before we discuss Roman-era sources for non-Classical forms of Latin, let us first look at the eventual daughters of Latin: Romance languages. Comparison between Romance languages reveals several interesting commonalities: neutralization of Latin long \bar{e} and short i in their modern reflexes and words commonly used in classical Latin works like *loquor* or *pulcher* being completely missing. With these facts, we conclude that there exists a proto-Romance, which is a kind of Latin but deviates from Classical Latin, in which there is no distinction between \bar{e} and i , and *loquor*, *pulcher*, etc. has been replaced (Herman, 2010, pp. 1-3).

Now we go back to earlier sources. Besides reconstruction based on modern Romance languages, data on non-standard Latin can also be found in “mistakes” or “barbarisms”, which are likely reflections of aspects of the oral language which has deviated from the classical works, in texts in late Roman Empire. These “mistakes” seem to be consistent with the aforementioned proto-Romance: for example, neutralization of \bar{e} and i is indeed constantly observed in these texts (Herman, 2010, p. 4).

Non-compliance to Classical Latin standards can be observed even in texts that are supposed to be official and formal, especially after the collapse of Western Roman Empire. In *Chronicle of Fredegar* (mid 7th century), for example, we observe aforementioned Romance-like mixing of vowels and incorrect endings which suggest they had been weakened in the oral language (as in later Romance languages). What is particularly telling is hypercorrections which reveal that the author really couldn’t tell some Classical Latin sounds apart but was under the vague impression that some sound distinctions were needed in Classical Latin (Leonhardt, 2013, pp. 158-159).

Unity of Vulgar Latin The next question is whether we can say there is one single Vulgar Latin. Regional variance of Latin can be expected to exist, according to our experiences with contemporary languages, and do seem to exist according to Latin authors’ accounts (Herman, 2010, p. 116). We are however still confident to say that the regional variance was not of very strong significance, for two reasons.

First, non-literary written texts were largely uniform across the whole Empire (Herman, 2010, p. 117): this is partly due to the existence of many (Classical-like) formulaic texts that can be used everywhere, but when deviations from Classical Latin appeared, they too appeared everywhere in the empire (Clackson and Horrocks, 2011, p. 235), though the propagation may take some time (e.g. see § 5.3.1), and the statistical frequencies of non-standard features (attested ubiquitous) vary regionally (Herman, 2010, p. 118).

Second, Roman grammarians were keen to correct “corrupt” usages of Latin and were known to record local uses, like Africans did not distinguish between long and short vowels. But they reported nothing more fundamental than that: the most notable differences seemed to be pronunciation (Herman, 2010, p. 118).

Based on the available evidence, we can define *Vulgar Latin* as the spoken variety of Latin used by people who were without very formal Classical Latin-oriented school education, which deviates from Classical Latin in a more or less homogeneous way across the Empire.

Elusiveness of Vulgar Latin Although Vulgar Latin, as is mentioned above, is relative uniform, it would be futile to attempt to compose a grammar of Vulgar Latin comparable to a grammar of Classical Latin. The first reason is the fact that although regional variance in Vulgar Latin is mostly ignorable, the language does have chronological development, and stylistic variances in different communities and spheres are to be expected. Conscious and unconscious prescriptive pressures can also cancel existing variances.

The second reason is about the nature of materials we have about Vulgar Latin: as is mentioned above, they were all “Classicalized” when written. Two extreme perceptions of Vulgar Latin are both incorrect. Vulgar Latin is *not* an unattested language that is quite different from Classical Latin and can only be reconstructed by the comparative method, as we have texts. But Vulgar Latin is also not another literary language besides Classical Latin, because these texts are strictly not truly “Vulgar Latin texts”, as Romans did not have field linguists to record Vulgar Latin speeches as they were, and all “Vulgar Latin texts” inevitably have both Classical and Vulgar traits (Herman, 2010, pp. 8,26). If a feature that’s absent in Classical Latin appears consistently, it has to be a Vulgar, but if certain deviations from Classical Latin is absent, we do not know if it’s because it’s because Vulgar Latin (in the given time period) was indeed consistent with Classical Latin or if it’s because relevant deviations have been regularized by scribes.

This is probably most obviously demonstrated in *Gaius Nouius Enus*, the permission of extension of a loan written on a wax tablet dated to 39 AD. Some consistently appearing spelling “mistakes” clearly show that the spoken language of the author is not precisely the same as Classical Latin and a large portion of the deviations are in accord with spoken Latin revealed by Pompeian graffiti. However, some sub-standard features which are observed in *earlier* periods are absent in the text, likely because they were stigmatized and the author didn’t want to sound “rural”. Despite the sound changes in the phonology, the text is still pretty comprehensible within the grammatical framework of Classical Latin, which is not surprising given its formulaic nature. Interestingly, some archaisms can be observed in the text, clearly demonstrating that the author was very serious about it, so any spelling mistake has to be because the author did not have access to the “right” pronunciations, not because he had a sloppy attitude. So now what we have is a mixture of some intended archaic spellings, the Classical grammar and a large portion of Classical words, and some sound variances cross validated by other sub-standard texts, while other changes, attested elsewhere, are not present (Clackson and Horrocks, 2011, § 7.4.1).

A final remark is that saying “Romans spoke Vulgar Latin and wrote Classical Latin” still does not accurately reflect the language situation of Roman Empire: besides Latin, we have Greek in the eastern regions, and bilingualism and language mixing are common (Clackson and Horrocks, 2011, § 7.2).

Christian Latin Now we turn to various sources of Vulgar Latin texts. In Christian texts, a tendency is that in the Empire period, early texts reflect *more* spoken features than later texts do: early Christians were often poorly educated, while later Church Fathers had had a good literary education. The same pattern can also be seen in how Neo-Latin replaced Medieval Latin. Many works of the latter however are still Vulgar in the sense that their purpose was to reach the widest audience possible, and definitely some features of spoken syntax and vocabulary were mixed into the texts (Herman, 2010, p. 24).

1.2.5 End of Latin as a living language

Evolution into Romance languages At some point, deviation of Vulgar Latin described in § 1.2.4 finally reached a breaking point such that it's better to stop calling it Vulgar Latin at all: if speakers wanted to learn Classical Latin, they could only learn it as something brand new, not as some prescriptive rules that “purify” their mother tongues (“don't forget to use the ablative here!”), and a speaker without any literary training would be unable to comprehend a Classical Latin article read to them. The problem is when was that point.

It seems that until the 7th century AD, the spoken language was still some sort of Latin: saints' lives were read directly to the congregations in Latin, with the expectation that they could understand them. In Gaul, the first piece of evidence of lack of intelligibility was King Charlemagne's observation that even the clergy could not always understand the meaning of Lord's Prayer. Later we have the famous phrase *lingua romana rustica* mentioned at the council at Tour in 813 (Herman, 2010, p. 114). The Romance part of Strasbourg oaths give us a glance into what *lingua romana rustica* looked like: the text can still be mapped to Classical Latin word-to-word, but the forms of the words had changed so much that intelligibility between the language reflected in the oath and Classical Latin had been greatly weakened (Clackson and Horrocks, 2011, pp. 300-301).

In other areas, evidence for evolution of Latin into Romance was one or two centuries later. And even after Classical Latin became largely unrecognizable to speakers of Romance languages, the memory that the local vernacular and Classical Latin have a shared origin may still be maintained. In Italy, as late as the fifteen century, Latin and Italian were respectively known as *grammatica* and *volgare*, i.e. two varieties of the same language (Leonhardt, 2013, p. 146).

The fate of spoken (Classical-like) Latin The birth of Romance languages does not necessarily mean the death of (Classical-like) Latin as an oral language. In Rome this was preserved at least as late as 1000 AD, as is exemplified by the epitaph of Pope Gregory V, which mentions that he spoke Latin besides a possible Romance language (Clackson and Horrocks, 2011, p. 268).

1.2.6 Coverage of this note

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.

1.3 Sociolinguistic status

1.3.1 Latin before and during early Republic

Etruscan: an early supraregional language The historical and contemporary importance of Latin of course doesn't endorse it as an inherently superior language. Indeed, we only have a handful of Latin texts before 600 BC; as a comparison, there are about 150 pre-600 BC Etruscan texts. Even in the period between 600 BC and 100 BC, during which we have around 3000 Latin texts, we have about 9000 Etruscan texts, which is three times as many as their Latin counterparts. Early Roman elites were partly educated in Etruscan, which lasted to as late as 4th century BC (Leonhardt, 2013, p. 43).

Oscan: a competitor of Latin Oscan was a supraregional language before the rise of Latin, and it seems people speaking the language were recognized by Romans as *Osci*, a concept based not on ethnicity but on language. It's possible that Oscan had a literary tradition (see the discussion below about Greek), which was not preserved in inscriptions, probably for the same reason the great works in Latin were not carved into stones either and that this tradition was possibly semi-oral (Leonhardt, 2013, p. 51).

Greek: the model of literature language The most notable language that was more prestigious than Latin and highly influenced Latin literature was Greek. It is well known that the Roman written culture derived its form from the Greek.

Box 1.2: What type of Greek?

Koine, or Attic, or something else?

Perhaps Romans were not the only people impressed by Greek culture: Greek tragedies were translated into Etruscan and played, and Lucanians – likely Oscan speakers – appropriated Greek art for tomb paintings. The whole Italy was deeply influenced by Greek literature. Indeed, many early Latin poets were not native speakers of Latin: they were multilingualistic and lived on adapting Greek themes and genres and mixing them into Latin. It's more than likely that they would do the same for various Italic languages. Ennius, for example, explicitly said he had three hearts, a Greek, an Oscan, and a Latin. Maybe Oscan had its own literature as well, which unfortunately wasn't pass to us because only Oscan inscriptions survived. The influence of Greek culture to Romans was therefore both direct and indirect: some Greek influences were adapted in the whole ancient Italy, and then Romans adopted them (Leonhardt, 2013, pp. 49-50).

1.3.2 Latin in late Republic and early Empire

The rise of Latin seems closely related to the rise of Romans as a politic superpower in Italian Peninsula. The superiority of Greek however had lasted at least to the first century AC, when Latin literature only had its audience in Rome but Greek works was welcome everywhere (Leonhardt, 2013, p. 52).

Box 1.3: When did the status of Greek die away in Western Europe?

When the empire collapsed?

1.3.3 Latin in Medieval times

Before Carolingian Renaissance The prestige of Latin never vanished during the Middle Ages: in former Roman areas, since “barbarians” had long been incorporated into Roman society, the Latin heritage was just accepted as a part of their culture (Leonhardt, 2013, p. 137), and since missionaries played an important role in the culture of non-Roman areas, Latin works were highly regarded because they were essential for the training of future Roman Catholic clergies (Leonhardt, 2013, p. 138).

This doesn’t mean authors were always capable to produce Ciceronian Latin works. After the collapse of Western Roman Empire follows a period when truly Classical Latin works ceased to be produced: writings from this age were usually deeply influenced by Vulgar features (§ 1.2.4). It is likely that there was almost no language communities that can use Latin for some sort of communication and maintain the diglossia status (Leonhardt, 2013, p. 160).

Carolingian Renaissance

Box 1.4: Carolingian Renaissance

- Existing Latin grammars
- Irish scholars
- Virgil and Cicero, for example, were read solely for the purpose of learning what’s good Latin (Leonhardt, 2013, p. 170): can we say the way Carolingian Renaissance revives Latin depended on available Latin texts at that time?

A living second language In the rest of the Middle Ages, the status of Latin was somewhere between a truly dead language and a normal living language (Leonhardt, 2013, p. 145-150). Latin was still live in the sense that people learned Latin the same way modern people learn languages like Spanish or Mandarin, that’s to say, they learned Latin by memorizing a limited set of essential grammatical markers (in the case of Mandarin, they are “empty words”, and in the case of Latin they are inflectional paradigms) and engaging with an existing community under practical situations: the overwhelmingly rule-based learning style was a much later invention. On the other hand, Latin differed from a true living language because what was acquired by the learner was always a “learner’s language”, i.e. a possible mental grammar that is *one* explanation of the Latin text that the learner encountered, but probably not *the* explanation that a Roman had in their mind. When learning a true living language, on the other hand, the learner’s language is only auxiliary and is always subject to correction by native speakers. Latin therefore was a living second language, a language that was supposed to be living but no one spoke as their mother tongue. The situation is not unlike English outside the core Anglosphere, where people use English, the primary goal is often not to communicate with native speakers (often there is none) (Leonhardt, 2013, pp. 150-154).

Latin’s competition with vernaculars In some senses, the status of Latin as *the* true literary language protected – not eliminated – our understanding of early European vernaculars, because if one vernacular – say Frankish – was chosen by Charlemagne as the official language, it soon would undergo the first congress phenomenon aforementioned and then developed a high variety, and the low varieties would then

be poorly documented, as in the case of Vulgar Latin (Leonhardt, 2013, pp. 143-144).

Of course, if the status of Latin was so strong, no written materials would be produced in vernaculars at all. The interaction between Latin and the vernaculars therefore is an interesting topic. A general observation is that written vernaculars appeared first in non-Roman areas, and then in formerly Roman areas. This could partially be explained by a residue of diglossia in formerly Roman areas: since Classical Latin could still be partially (although highly limitedly) comprehended by speakers of local vernaculars, no attempts would be made to write anything using local vernaculars, just like how modern Greeks once rejected translating the Bible into Demotic Greek.

This mechanism however can only explain the *absence* of vernacular literature in formerly Roman areas, but the *presence* of vernacular literature in non-Roman areas does not automatically come into being because of lack of Latin capacity: it's definitely possible that these areas were so culturally impoverished that no texts – vernacular or Latin – were published at all. This however is definitely not the case because we actually observe a *positive* correlation between vernacular literature and Latin literature (Leonhardt, 2013, pp. 167-168).

Latin's competition with Greek and Arabic Unlike Greek, Latin had never gained much interest outside Western Europe.

Box 1.5: Latin literature outside Western Europe

- Arabs' attitude
- Byzantine

1.4 Previous studies

Box 1.6: Topics in previous studies of Latin

- Grammars in ancient Rome
- In Medieval times
- Latin and functional-typological tradition

1.5 Texts

Box 1.7: List of Latin texts

- Old Latin texts: archaeology
- Old Latin texts: Roman authors
- The road to standardization
- The Classical canon: Cicero, etc.

- No one read Classical Latin after Christianization, except in Ireland
- Latin inscriptions, curse tablets, etc. in Empire period
- Medieval Latin, Neo Latin

Chapter 2

Phonology and the writing system

2.1 Phonemes and the alphabet

The writing system Although the phoneme inventory of a language often is not accurately reflected by its preferred writing system, since Latin is a classical language and no ancient Roman is alive today, its phonology has to be inferred from known texts.

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. Note that ancient Romans only used the big letters; the small letters was invented during the era of Charlemange.

The letter *j* was not used by ancient Romans, although we sometimes see *I* appearing at the start of a word and therefore possibly represents the semivowel /j/. The letters *U* and *W* are also not used. Similar to the case of *I*, the letter *V* is used to represent what appears to be the semivowel /w/ as well as the vowel /u/. The letter *K* is an archaic one and only appears before *A* in a small number of words (Oniga, 2014, chap. 2). The letters *Y* and *Z* are used to spell Greek words that include *Y* and the voiced dental affricate, respectively.

Consonant inventory The Latin consonant inventory is therefore given by Oniga (2014, Table 3.1). Note that the letter *X* is a double consonant: it means /ks/ or /gs/.

Vowel inventory Two semivowels – /j/ and /w/ – can be recognized, which appear as *i* or *u*.

The vowels are given by Oniga (2014, Table 3.2). Each vowel has a long variety and a short variety.

Historical evolution of Latin vowels

2.2 Prosody

2.3 Morphophonological rules

Some phonological rules in Latin are sensitive to morpheme boundaries. We can therefore assert that at least in a historical stage, Latin speakers had a clear sense of mor-

phemes as real phonological objects, instead of mere theoretical models.

2.3.1 Vowel deletion

Short vowels *a*, *o* and *e* become zero before a morpheme boundary or another vowel (Oniga, 2014, § 8.3). This rule is exemplified by the absence of the thematic vowel in both declension (TODO: rosis) and conjugation (TODO).

2.3.2 Vowel shortening

A long vowel before another vowel or morpheme boundary is not deleted, but shortened. Again this is exemplified (§ 7.3.1, TODO: ref)

Also, in the final syllable of a phonological word, a long vowel before a consonant except *s* is also shortened. Counterexamples when the vowel is not in the final syllable exist, like *bāris* ‘a type of flat-bottomed freighter used on the Nile in Ancient Egypt’.

A long vowel is generally shortened before a sequence containing a liquid or nasal and a following stop consonant, like *nt*. This rule comes from an older Indo-European sound law: the Osthoff’s Law (Oniga, 2014, p. 55).

2.3.3 Vowel weakening

When a short syllable is in a medial, open syllable, and a morpheme boundary occurs immediately before, within or after the syllable, it becomes *i* (Oniga 2014, p. 55; TODO: ref).

2.3.4 Vowel lengthening

A vowel is always *lengthened* before *nf* and *ns* (Oniga, 2014, p. 55).

Chapter 3

Syntactic overview

We carry out this report round by round. In the first round we distinguish words and morphemes (§ 3.1), sketching how abstract linguistic structures are linearized into utterances in Latin.

In the second round, this note takes a syntax-first approach in the second round. Unlike traditional Latin grammars that start with morphological properties of words and then introduce syntax to impose some constraints on how words are bound together by constituency and dependency, this note starts with structural possibilities of clauses (§ 3.2) and noun phrases (§ 3.3), and modifications to them that aren't themselves noun phrases or clauses (§ 3.4). Then, according to how roots fit into these environment, we define parts of speech (§ 3.5).

This means, kindly of strangely, we first define *noun phrases*, and then define *nouns*. Note that there are several senses of terms like *noun*: their meanings as purely syntactic concepts, as morphological concepts and as parts of speech labels in dictionaries may not be identical § 3.5.1.

One advantage of the traditional approach is that it's parsing oriented; suggestions for parsing Latin texts are given in § 3.6.

The third round of our description consists of chapters following this chapter focus on different topics touched in this chapter.

3.1 Morphological typology

Wordhood Latin is well known for its rich morphology, which enables a rather free – but still not completely arbitrary (chap. 9) – constituent order. The grammatical categories seen in nominal and verbal morphology already reflect the most salient grammatical categories in NPs and clauses; thus the concept of morphological wordhood can be defined according to the boundaries of inflectional templates; and the traditional wordhood definition largely agrees with this definition. Thus, in this note I follow the traditional definition of Latin wordhood, which is easily done using phonological criteria, or, to be more accurate, orthographical criteria: what was documented by ancient Romans as a word is recognized as a word. For words without inflectional morphology,¹ wordhood can still be defined according to phonological wordhood. The two criteria of wordhood also largely agrees with the traditional word.

¹The traditional name is *particles*; see below.

Also we can talk about various purely syntactic ways to define wordhood. For example a noun phrase minus large complements or adjuncts should be recognized as a noun, or a clause minus large complements or adjuncts should be recognized as verb. In this regard Latin has split verbs in periphrastic conjugation, and in the preverb construction, the prefixed verb, a morphologically single word, contains the verb proper and an adverbial root that syntactically should be a part of the locative complement of the verb. This line of wordhood definition helps us understand the structure of Latin noun phrases or clauses but goes against any conventional definition of wordhood.

The only subtlety in wordhood definition seems to be due to clitics, the most important one being the coordination *-que*: they are not a part of inflectional morphology of any stem but have to be attached after another word. For them, the orthographical standard simply dictates that they are to be recognized as a part of the words they follow. In conclusion, we can say the traditional notion of Latin words is (expectedly) linguistically sound and should be kept in use.

Stems and endings in nominal and verbal morphology A stem can be well defined for both verbal and nominal morphology in Latin, enabling a clear derivation-inflection distinction: prototypical derivational processes that are considered to be a part of the head noun/verb and not the surrounding NP/clause are morphologically realized strictly before prototypical inflection processes, forming the stem (§ 5.1.1, § 7.1.1) to which inflection processes are applied, with possible contextual allomorphs (§ 2.3). Latin inflection is always suffixal, while derivation is predominantly prefixal. Despite its richness, a large portion of instances of Latin derivation are historical, with meanings of derived forms having significantly shifted and no longer regularly inferrable.

A general tendency in the stem structure is modification affixes usually are prefixes, while affixes bearing grammatical information (like change of part of speech) are usually suffixes (§ 5.1, TODO: ref). Apart from affixation, Latin does use compounding, as in *aequilibrium* (*aequ-i-libr-ium*, ‘even-balance-SG.NOM, equilibrium’; Smith 2016, § 92), but compounding was already less productive – if not completely obsolete – in the Classical period.

Non-concatenative morphology Concatenative morphology (affixation and compounding) is prominent but isn’t the only morphological device. Reduplication is attested in formation of the perfect stem (TODO: ref); this however is largely historical and is no longer productive. Dropping of first-conjugation stem-final vowel (§ 7.3.1) may be analyzed as subtraction, although it can be seen as due to morphophonological rules. 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.

Residues of the ablaut system in PIE can be observed in Latin. In the third declension, for example, we have the following alternation: *gen-us* ‘family-NOM.SG’/ *gen-er-is* ‘family-GEN.SG’, and since *-us* is *-os* in earlier stages of the language, we observe an alternation between *o* and *e*. Another instance is the alternation between *toa* and *tegō* ‘cover’. It’s also possible that no vowel exists at the place where we find *o* and *e*. We therefore find a vowel gradation system with three variants: \emptyset , *o*, and *e*. Note that what controls the vowel gradation system is grammatical categories like case, etc., and therefore ablaut is a part of morphology, not just phonology (Weiss, 2009, p. 46).

Box 3.1: Ablaut example in Latin

- In noun inflection
- In verb inflection?

3.2 Clausal syntax

3.2.1 Introduction

Clauses in utterances We start our systematic discussion on Latin syntax from the verbal system, i.e. the clause. A clause may appear as a single utterance called a **sentence**,² which is either declarative or interrogative. An embedded clause may be a complement of a verb (complement clause construction), or a high-level modifier in the verbal system (subordination),³ and as a relative clause (i.e. modifier in the nominal system).

Box 3.2: Distribution of clauses

List where clauses can be found in other structures.

Clause types Clauses types in Latin can be roughly divided into finite and non-finite ones, which can be told from verbal morphology.

In sentential clauses, the category of speech force is marked by e.g. fronting of the interrogative pronoun. No other grammatical markers are dedicated to clause type: there are, for example, no sentence-final particles that can only be observed in *sentential* clauses and not embedded finite clauses.

Non-finite constructions include INFINITIVES and PARTICIPLES, and the GERUND. These constructions can be distinguished by morphology only: in English we have infinitive clauses, but strictly speaking, there is no such thing as “infinitive verb”, as the head verb of an infinitive clause has exactly the same form of a non-third person singular present tense verb; in Latin however, the head verb of an infinitive clause in Latin indeed has a separate position in the paradigm.

Box 3.3: Implicit marking of speech force categories

Are there any e.g. obligatory fronting or contraction that may be related to the speech force of a clause?

We also need to investigate the licensing environment and the inside *wh*-forms if the clause is a relative clause.

Box 3.4: Clause-level subordination

- Coordination
- Conditional

²Some authors (including some Latinists) use the term *sentence* as a synonym of *utterance*; it is not followed in this note.

³The term *subordination* sometimes is used as a synonym of *clause combining*. Here we use *subordination* in a narrower meaning.

- Concessive

Internal makeup of a clause The external grammatical function of the clause may be reflected by conjunctions and/or the morphological MOOD. The complexity of the clausal structure is strongly influenced by the clause type, which, as is said above, is reflected in the verbal morphology. The maximal complexity possibility of a clause consists of a nucleus clause with possible information structure-related order alternations and/or clause-level subordination constructions.

At the surface level, the nucleus clause contains a (omissible) subject (§ 3.2.2), a main verb with TENSE marker and a possible auxiliary verb (chap. 7), adjuncts and internal arguments.⁴ The main verb shows regular agreement with the subject. The inflectional ending of the main verb, together with some adjuncts and an optional auxiliary verb, marks tense, aspect, mood (TAM) grammatical categories (§ 3.2.3). The distinction between what are known as peripheral arguments – another type of adjuncts – and core arguments is discussed in § 3.2.4; we expect most so-called peripheral arguments to in some sense have scopes over core arguments.

The core arguments – and to some extents the peripheral arguments – are ultimately determined by the **argument structure** of the main verb. Arguments have “deep” and “surface” positions. The deep positions are defined by properties like how reflexives work, mapping of semantics to argument slots,⁵ the semantics difference between S=A and S=P valency alternations (§ 3.2.5), etc. The mapping from deep positions to surface positions (like *subject*) is influenced by

- *Voice*. Latin has a passive **voice**, which is marked on verbal morphology (chap. 7) and seems to work solely on the concept of **object**; no other highly regular voice constructions of such kind are observed (§ 3.2.6). It’s still possible to have verb frame alternations that highly depends on the properties of the verb stem.
- *Alignment typology*. Regarding **alignment typology**, Latin is a nominative-accusative language with clear S=A neutralization and therefore a well-defined **subject** (§ 3.2.2).

Details of deep argument structures and their reflections in the surface argument structure are specified in possible **verb frames** in a language. The argument structure is often related to the **event structure** of the verb, and therefore the verb frame may also influence the TAM marking (chap. 8).

⁴In some works on English syntax, e.g. Huddleston and Pullum (2002), the term *complement* means a dependent position which, according to various standards, is more closely related to the lexical head; in the case of clauses, a complement is just an argument.

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 *predicative complement* in Huddleston and Pullum (2002).

A further confusion, as is seen in § 8.1.2, is due to the term *complement-taking verb*, i.e. a verb that take a complement clause or something semantically equivalent to a complement clause as one of its arguments.

⁵We should however note that two arguments being denoted by the same semantic role label (e.g. *stimulus*) do not mean they are truly semantically equivalent, and clearly they are necessarily syntactically comparable. In Dixon (2005, § 4.2), for example, what bears the label *target* is almost prototypically patient-like in verb frame I and is the goal of a directional construction in verb frame II, and of course there are syntactic as well as semantic differences between the two frames.

Constituent order The flexible constituent orders of Latin lead someone to characterize it as a free-order language. However, a closer examination reveals that dependency relations in Latin clauses do control the constituent order and a neutral order roughly described by Fig. 3.1 can be established. This order is due to both default information structure and argument structure, and order alternations on top of it due to information packaging are also well attested (§ 3.2.7). Prosody also plays a role in determining the constituent order.

Subject - direct object - oblique argument - adjuncts - source/goal - cognate object -
verb

Figure 3.1: Latin neutral constituent order

3.2.2 Subject

Clausal pivot Latin is a nominative-accusative language. This means we have a well-defined clause-level pivot which, in active constructions, can be identified with the structurally most agent-like argument⁶ in the argument structure. This pivot is called the **subject**.

Existence of a clause-level pivot can be clearly identified from the following facts.

- *Extraction in coordination.* When two clauses with a shared subject are coordinated, the subject usually appears at the beginning of the coordination construction (Pinkster, 2021, § 19.5). This demonstrates that the subject is somehow structurally promoted to a higher position in this kind of coordination.⁷
- *Case marking.* Subjects are always nominative for finite clauses, whenever case marking is available (i.e. whenever the subject is an NP or a gerund). On the other hand, nonfinite clauses are either subjectless or have an accusative clausal dependent functioning as the subject (§ 10.1.1): we may argue that by being nonfinite, they have deficient nucleus clause structures and cannot afford a canonical subject.
- *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.
- *Constituent order.* In clauses with neutral information structure, the subject usually appears at the beginning of the clause. Violations to this generalization are countless but they can all be analyzed as focalization or topicalization of other constituents and postponing of the subject (§ 9.6).

Latin subjects also have a quirky features. A subject is an NP or a complement clause, usually an infinitive but never a gerund (§ 7.5.2.1). This constraint isn't seen in any other clausal complement types.

Since we have already identified a clausal pivot in Latin, we can call the rest of the nucleus clause the **verb phrase**. We will in turn separate TAM markers from it

⁶What it takes to be agent-like is discussed below.

⁷Omission of the object in coordination, on the other hand, shows more diversity in constituent orders (Pinkster, 2021, § 19.6), suggesting that the exact structural analysis may also be more complicated.

(§ 3.2.3), getting the extended argument structure, and then separate peripheral arguments from it (§ 3.2.4) and get the core argument structure (§ 3.2.5). Because of the flexible constituent order, however, the term *verb phrase* doesn't have much value for surface-orientated analysis.

Neutralization of argument structure pivot and clausal pivot In Latin we can also recognize an argument structure-level pivot, or in other words, an “agent-like” argument in each clause (§ 3.2.5). The fact that in active clauses, an agent-like argument is always the subject means Latin is indeed a nominative-accusative language.

Impersonal constructions In certain cases, the subject position can be left empty. The verb then takes the default 3sg agreement marking.

3.2.3 Tense, aspect, modality, and so on

Tense, aspect and modality is most evidently marked by verbal inflection in Latin (chap. 7). Other TAM categories however are regularly marked by adverbials.

Box 3.5: Latin nonfinite constructions

Nonfinite constructions lack *morphological TAM marking*, but are TAME adverbials possible for them? Is nonfiniteness a morphological or syntactic property?

Speaker-oriented categories Sentential clauses, i.e. clauses appearing as a complete utterance (§ 3.2.1), can first be classified into declarative, interrogative, and imperative ones: the interrogative clause is marked by fronting of WH-words, and the imperative clause is marked by verbal morphology.

Latin does not have morphological marking of evidentiality. There are however hypotheses that the alternation between *Accusativus cum Participio* and *Accusativus cum Infinitivo* marks evidentiality (Greco, 2013).

Epistemic modality is marked by the MOOD in declarative and interrogative clauses (Pinkster, 2015, p. 388).

Primary tense The primary tense establishes a *reference time* and compares its relation with the *speech time*. The “speech time”, of course, can also be the time of writing, or even the time of reading: *now readers may wonder why ...* Latin has a past/present/future tense system: this can be clearly observed by staring at the finite paradigm. The past future tense seen in English (*he [would] take part in in if he knew it*) is absent in Latin.

The reference time sets the stage of the situation of affair being described; the exact time when the situation happens is not necessarily identical to the reference time: see the category of anteriority.

Denontic modality The morphological MOOD in the imperative construction seems to mark the denontic modality: there is no epistemic modality to mark anyway (Pinkster, 2015, p. 388).

Habituality and similar categories In Latin, just like in English, the habitual aspectuality is not marked by verbal morphology, but often the PRESENT category has a habitual reading.

We note that crosslinguistically, habituality and anteriority and point of view aspectuality (see below) can co-exist: in English, for example, *already* can be seen as a weak version of anteriority, and we have examples like *what we feel has [always]_{habitual} [already]_{anteriority} been felt, again and again*. We can take a step further and observe

co-existence of habituality, anteriority, and imperfectivity: *this life has [always]_{habitual} [already]_{anteriority} [been happening]_{progressive}*. The scopes of the three categories seem to follow the order of habituality > anteriority > point of view aspectuality.

Box 3.6: Co-existence of habituality and anteriority

In Latin is it possible?

Anteriority Anteriority, i.e. whether the state of affair being described happens before or after a reference time (whose relation with the speech time is determined by primary tense), is sometimes recognized as tense and sometimes aspect. In Latin we can recognize a clear perfect/simultaneous distinction from verbal inflection because we have a PERFECT stem and a IMPERFECT stem and also different endings for the two system.

It seems we also have a posteriority value in Latin, because even when the primary tense is absent (as in the two FUTURE participles) or marked otherwise (as in two periphrastic verb forms), a meaning of “future” – not necessarily defined according to the speech time – is still available, and the FUTURE participles rarely appear with *sum* in PERFECT (which encodes anteriority) forms (Pinkster, 2015, Table 7.3).

Point of view aspectuality A category that is often known as the grammatical aspect is the *point of view* of the speaker to the situation being described, or, in other words, the position of the “event time” (which is compared with the reference time in anteriority) within the whole situation: the situation may be described as a whole (**perfective**), or the clause only describes a fraction of a continuous event (**imperfective** or **progressive**). The distinction between the two can be subjective: closing a door is an almost instant event, but we still have *I’m closing the door*.

In Latin we do not have a clear marker of the perfective/imperfective distinction. It is however generally believed that at least the IMPERFECT verb form regularly has an imperfective meaning.

The relation between the point of view and the situation can be more complicated: for example we have terminative, inchoative or retrospective aspects. The Latin *-sc-* infix is an inchoative marker.

Lexical aspect All the categories above are either about the temporal location of a situation or the speaker’s subject judgments of the situation. A situation also has inherent properties, known as its lexical aspect.

3.2.4 Peripheral arguments

The term **peripheral arguments** is used in Dixon (2009) to refer to what others call (non-TAM) adjuncts and specify information like the spatial and temporal location of the event, the instrument, the manner, etc. The distinction between peripheral arguments and **oblique arguments** – core arguments which are also known as instruments, manners, locations, etc. – is inherently not clear; grammatical phenomena traditionally attached to the distinction involves (McInnerney, 2022):

- *Structural closeness to the main verb*. The prototypical core argument structure is a structure expressing ‘someone did something to ...’. Constituents like so-called circumstantial constituents modify the core argument structure as a whole, so they are prototypically peripheral. The syntactic status of other constituents are not so clear (see below).

- *Idiomization and subcategorization*, i.e. whether lexical properties of the main verb blocks removal of the argument. These factors are syntactic aspects of verb frames. English prepositional verbs are a good example of this. A argument licensed by the subcategorization pattern may still be omitted because of pro-drop, but in this case usually it has a definite reading because when omitted, it's expected refers to an entity specified in the context. Indefinite implicit pronouns are still possible in some cases; whether this is recognized as valency alternation or pro-drop is purely terminological.
- *Semantic and pragmatic requirement*. A clause should “say something”, and hence we have semantic obligatoriness of certain arguments. This does not necessarily dictate subcategorization. An argument that's semantically not obligatory may still be syntactically obligatory: *slog*. A semantically obligatory argument is usually syntactically projected, but it can be filled by an indefinite null pro-form: *I ate* (§ 3.2.5).

Distribution patterns of arguments may sometimes be explainable by both the second and the third factors. For example, the clause *it rains* will almost never have an instrument phrase, which could be explained by semantic incompatibility between a weather event and an instrument phrase, or by subcategorization. Whether in extreme semantic cases an instrument argument can be licensed may vary across speakers, and often subcategorization patterns may just be syntactic fossilization of semantic factors.

Despite the theoretical independence of the three parameters, for clear functional reasons, they are often correlated, and most of the time the core/peripheral distinction is still meaningful.

In Latin, arguments that are not structurally close the the main verb seem to appear at the center of a clause, indicating that they do not undergo any information packaging operations by default (Fig. 3.1, § 3.2.7).

Circumstantial adjuncts Circumstantial adjuncts are clausal constituents that modify the event as a whole and specify when and where it happened. They likely are the most peripheral constituents in the extended argument structure, modifying the core argument structure about ‘someone did something to ...’ (Cinque, 1999, p. 29).

Manner There seems to be two manner positions: one has a position similar to that of the instrument and mean phrases, and in the broad focus constituent order resides *after* the direct object (Devine and Stephens, 2006, p. 71-75). Another position seems to scope over the core argument structure as a whole, and therefore is more peripheral and appears before the object in the default order (Devine and Stephens, 2006, pp. 101-109).

Instrument and mean The exact syntactic status of the instrument and the like is controversial. Crosslinguistically, it seems some instrument phrases are a part of the core argument structure (Pascual Pou, 2001), and some instruments, even the pure instrumental ones, can be regularly promoted to the subject position under proper pragmatic environments (Alexiadou and Schäfer, 2008), indicating that they belong to ‘do something to ...’ part of the core argument structure, and they are not as peripheral as circumstantial adjuncts.

Box 3.7: Instrument argument

In Latin it seems an argument/adjunct distinction exists for instrumental arguments: (Devine and Stephens, 2006, pp. 57-58, p. 65-67). Does this mean that when both appear, the argument instrument appears before the adjunct instrument?

If we want to decide the position of English *with*:

- *he covered the hole with a blanket* is not the same as *he covered the hole with his toolkit*: the first can be transformed to *a blanket covered the hole* while the second can't. See The Reversible Core of ObjExp, Location, and Govern-Type Verbs
- Instrument-subject alternation: *the knife killed him*. This also leads to an agent-causer distinction. See Instrument Subjects Are Agents or Causers

Internal makeup of peripheral arguments in Latin The internal makeup of the argument – case marking and prepositions – is weakly correlated to the core/peripheral distinction in Latin. Nominative arguments are never peripheral, and accusative arguments are rarely peripheral, but nothing could be said for the rest of the cases and prepositions.

3.2.5 Core arguments

Grammatical relations, deep and shallow As is said in § 3.2.1, core arguments have deep and surface (i.e. clause-level) status in the clause. In this section, we list several prototypes of the deep argument structure and discuss how the arguments are assigned surface grammatical relations. The most uncontroversial surface grammatical relation, the subject, has already been described in § 3.2.2, and here we consider other surface grammatical relations.

The **direct object** is the surface grammatical function of the P argument in the prototypical transitive construction. Here we list its properties and see whether arguments in other verb frames can also be described as direct objects:

- *Case marking*: Direct objects are always accusative when it makes sense to talk about case – but not all accusative arguments are direct objects.
- *Passivization*: If an argument is coded as the direct object, then it can regularly be promoted to the subject position in a passive clause (§ 3.2.6). Secondary objects are less frequently promoted in passivization.

As is said in § 3.2.4, oblique constituents in the clause are not necessarily peripheral. The most prototypical examples are seen in directional constructions and dative constructions: if we acknowledge that the object is a core argument, then the locational/dative phrase has to be a core argument as well, because the two are licensed together. It's also possible to have manner and instrument phrases as core arguments. Traditionally the recipient argument of a verb with a meaning of giving is known as the **indirect object**. We do note that this “indirect object” isn't very different from other oblique constituents in the core argument structure. The term *indirect object* therefore is used as a synonym of *recipient argument* in this note.

The instrument expression is structurally further from the main verb but not as peripheral as circumstantial phrases (§ 3.2.4). In some verb phrases it can also be considered as a

Box 3.8: *full of*

Is it necessary to stipulate an argument structure of *full of*? consider a three-place verb where the third argument is genitive. This seems to be related to be judicial verbs: *accuse of*.

Agent and patient in the prototypical transitive construction The most prototypical transitive construction involves an agent (A), which initiates the event and exists independently of the event, and a patient (P), which is affected by the event, sometimes would not exist without the event. In active clauses, the A becomes the subject and the P becomes In Latin often the transitive construction can be detransitivized by suppressing the P argument (we can also understand this as filling the P position by an empty indefinite pronoun), resulting in what is known as the **absolute use** of the verb (Pinkster, 2015, § 9.16). This is also observed in English: *the dogs bit me* v.s. *the dog bites*. This leads to a S=A valency alternation.

There exists another class of transitive constructions, in which the internal argument participates in a situation of affair, and the external argument causes this. We can say internal argument is P-like and the external argument is A-like. It's possible that the external argument can be removed and we still have a well-formed core argument structure: this leads to S=P valency alternation. In principle, the argument structure that contains only the internal argument may be unaccusative or unergative (see below) and the transitive use of the verb can be viewed as a causative construction.

Box 3.9: Detransitive derivation

In Latin, do all S=P verbs have unaccusative intransitive counterparts?

The sole argument of intransitive verbs The sole argument of intransitive verbs – referred to as S in the typological literature – either starts an event, or passively undergoes an event: consider *Simon jumped off* and *Simon fell*. That's to say, the status of S in the deep argument structure may be comparable to either A or P in the prototypical transitive construction; the intransitive verbs are correspondingly known as *unergative* and *unaccusative* verbs, although the names are misleading as we are discussing about the deep argument structure and not alignment yet.

In Latin, it seems that unaccusative intransitive verbs are all deponent (Oniga, 2014, pp. 308-309). The passive voice therefore is used whenever an agent is absent. unergative intransitive verbs do not have PERFECT PASSIVE PARTICIPLES, because the passive voice is not compatible with them (Giusti and Iovino, 2019).

Psych-verbs Verbs about psychological activities may display unusual behaviors. The argument structure may be a causative construction where the stimulus causes the experiencer to have psychological activities, and the stimulus is A-like, or a construction where the experiencer directs their emotion to a target, where the experiencer is A-like. Plus, for at least some psych-verb, the stimulus-causers originate from the stimulus-target, so that at least for some speakers, reverse binding as in *this rumor about herself frightened Mary* is acceptable (Hornstein and Motomura, 2002).

The above two verb frames also have their intransitive versions, where the internal target argument or the internal experiencer argument is oblique. The intransitive

constructions can be unaccusative (*I'm anxious about ...*) or unergative (*I rejoice because of ...*). We further have a possible impersonal construction where the experiencer and the stimulus form a small clause, and the subject is filled by a dummy pronoun. All the constructions have been attested in Latin (Giusti and Iovino, 2019).

Box 3.10: Passivization in verbs about teaching

Can verbs taking accusative argument in (Pinkster, 2015, § 4.36) be passivized?

Comitative in argument structure Sometimes, the argument structure licenses a comitative construction, whose head becomes the subject or the object, and the rest of the construction becomes an oblique argument (Zhang, 2007). In Latin this is observed, with the comitative marker being the preposition *cum*.

Dative and benefactive constructions The verb frame of verbs meaning ‘giving’ in Latin regularly contains a dative construction, which is quite similar to a directional construction (see below), where the thing being given – the *theme* – goes to the *recipient*. The case of the latter is always dative. In Latin, the dative construction comes together with the causer of the event of giving, which expectedly functions as the subject of the whole verb frame, and the theme, without any semantically loaded case marking, becomes the object.

Box 3.11: Benefactive ditransitive

Do we also have *buy a bottle for me* construction in Latin? It seems Latin only has benefactive adjuncts.

It's possible to keep only the dative recipient in the benefactive construction: this leads to two-place verbs governing a dative argument (Pinkster, 2015, § 4.24). We note that the benefactive construction differs from the prototypical dative construction of giving, and the latter is never attested in two-place verb frames (Pinkster, 2015, § 12.7).

Directional constructions Some Latin verbs with a preverb seem to have a directional small clause (§ 3.2.9) hidden in their frames: the case marking of the argument(s) are sometimes never attested without the preverb, and the meaning of the verb may also vary depending on whether the preposition is attached to the verb, indicating that it's the preverb that licenses the argument (Mare, 2018).

The preverb always describes the *path* in the directional construction: the type of the *location* is either marked by the case of the location argument (often hence dictated by the preverb as if the preverb is used as a standalone preposition and the location is governed by it), or is in the default accusative case. Passivization of accusative location arguments is sometimes possible (Pinkster, 2015, § 4.22). These phenomena seem to be comparable to English preposition verbs and verb-particle constructions, where the preposition may be somehow defunct and the location argument it governs syntactically is the object of the verb, but it's also possible that the prepositional phrase is sealed and nothing can be extracted from it.

It is also possible for the location argument to appear in dative (Pinkster, 2015, § 4.25). The location argument in this case can always be understood as the goal of the action, and the dative case seems to mark this fact. This indicates that the requirement that the NP governed by prepositions has to be accusative or ablative is likely a quirky realizational feature of Latin: the fact that an argument is the target in a directional construction can only be marked by the accusative case when the

directional construction is realized as an independent prepositional phrase, but the dative case is available in verb frames.

Box 3.12: The subject of the directional construction

The subject of the directional construction may appear as the subject of the clause or the object of the clause?

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.

Box 3.13: Oblique copular complements

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* (§ 5.3).

Cognate objects A **cognate object** is a clausal dependent that is obligatorily selected by the main verb and usually overlaps semantically with the content of the main verb. In English we have *Ben died [a long and painful death]* and *Joe smiled [a smile]_i and it_i is a creepy one*.

3.2.6 Voice

Prototypical passive Latin has a prototypical passive voice: in this construction, the A-like argument – usually the most external argument – is either suppressed or becomes an instrument-like argument (we may say that when it's suppressed, the instrument-like argument is replaced by an empty indefinite pronoun). After this, the second highest argument that can go to the subject position – the direct object – is promoted to the subject position.

Deponent verbs and impersonal passive In Latin, the passive morphological marking of the verb also generalizes to all verbs without a prototypical A. This happens to unaccusative verbs (§ 3.2.5), and also in so-called impersonal constructions, where the A is suppressed but nothing is promoted to the subject position.

3.2.7 Information packaging

Information packaging is observed both on the clausal level and inside the extended argument structure. The relevant information packaging devices can be roughly classified into three types:

- *Information packaging within argument structure*. Information structure-neutral clauses have the whole verb phrase as the focus (so-called **broad focus** clauses),

and within this focus we can identify several weak topic and focus positions. By default, the direct object and certain oblique arguments (like the dative recipient) are weak-topicalized, meaning that they precede adjuncts in the extended argument structure (Devine and Stephens, 2006, § 1.5). On the other hand, directional phrases like goal or source are usually not weak-topicalized, which probably is because they are not conversational participants and are usually not old information. Cognate objects are always indefinite and they can't be topicalized, and because objects are close to the main verb, they are to the right of directional phrases.

The two competing factors – the peripheral nature of peripheral arguments, and direct object and certain oblique arguments being focused while others are not – lead to the neutral order in Fig. 3.1.

- *Scrambling for presupposition.* This means fronting a constituent to imply that it is presupposed (this is a weak form of topicalization) or weakly focalized (Devine and Stephens, 2006, pp. 102,108-109). When information-neutral manner adverbs are present (whether the adverb undergoes some movement can be tested by its scope (Devine and Stephens, 2006, p. 100)), this move constituents to its left.
- *Prototypical topicalization and focalization.* Prototypical topic and focus constructions completely change the broad focus information structure, and may – but not necessarily, in-situ and scrambled constituents can still receive strong focus – move the focalized or topicalized constituent to a higher position, sometimes even before the complementizer if there is any. Without information like stress or conversational contexts, it's sometimes hard to tell prototypical focalization and topicalization from scrambling but constructions like *Y non X* or interrogative clauses undoubtedly involve the former (Devine and Stephens, 2006, chap. 3).

These operations render the constituent orders of Latin clauses extremely variable.

3.2.8 Clause combining

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

3.2.9 “Small clauses”

Some constructions may be analyzed as “small clauses”, which are smaller than usual clauses and lack real TAM marking; they are therefore not “real” clauses, but still appear frequently in Latin texts. (1) is a formula used frequently in Catholic Church and is an example of a small clause as a single and complete utterance. The fact that *gratias* is not in the nominative or the vocative case means (1) is neither a finite construction nor a single-NP utterance.

- (1) De-o grati-as
 God-SG.DAT thank-PL.ACC
 ‘(May) thanks be to God.’

Since English – and a lot of other languages – bans small clauses appearing as complete utterances, translation of these constructions may map them into full clauses in the target language.

Box 3.14: Why small clauses

Alternatively, small clause constructions can be analyzed as a full clause (for example, a copular clause) with the verb deleted. The main problem of this analysis is if this is true, we need to explain why prototypical transitive clauses never see their verbs omitted. It's therefore better to say that a copular clause is a small clause *plus* the copular, and that a giving or receiving ditransitive verb takes a small clause as its internal complement, which doesn't involve a questionable verb deleting process and also captures the intuition that a small clause is a full clause minus the verb.

3.3 The noun phrase

3.3.1 Introduction

Noun phrases in utterances

Box 3.15: NP in utterance

- Response as a piece of utterance
- Argument and adjunct

Internal makeup of noun phrases The hierarchical structure of the noun phrase is often ignored in traditional grammar, and yet it has delicate formal parallelism with the structure of the clause. As Latin is a classical language, detailed tests about these structures are not always available and are likely subject to individual variances. In this section we list some points worth noting that are repeatedly attested in other languages.

First, noun phrases are labeled for their grammatical functions, which in Latin is mostly done by the morphological case system and also sometimes the preposition system (§ 3.3.2).

The noun phrase can then be divided into determiners (§ 3.3.3) and the inner parts (the nucleus noun phrase), just like how we first note that a clause is a nucleus clause plus information packaging and/or marking of clause types and the grammatical function of the clause (§ 3.2.1).

The nucleus noun phrase⁸ can further be divided into a core argument structure surrounding the head noun (§ 3.3.6), and a series of more peripheral dependents modifying the former.

Some languages, like English, have a genitive construction in which a genitive phrase is raised out of the nucleus noun phrase and acts just like the subject in a clause; Latin doesn't have this kind of constructions.

⁸This is known as a *nominal* in Huddleston and Pullum (2002); the term however is usually used to refer to morphologically nominal parts of speech in the Latin grammatical tradition.

Box 3.16: The “subject” in Latin NP

Does Latin have constructions like *his play of the national anthem*?

In the modifier series surrounding the core argument structure, modifiers that are “direct modifications” and structurally close to the head noun modify it as a whole, just like clausal peripheral arguments and lexical aspect do, and are often semantically non-restrictive⁹ and more about eternal properties of the things referred to. Direct modifications usually form a hierarchy in terms of scopes and frequently constituent orders (Huddleston and Pullum, 2002, p. 453, [10]).

Structurally higher modifiers, on the other hand, restrict the reference of the noun phrase, just as structurally higher TAM categories do (Cinque, 2010, pp. 23,25), and they generally look like reduced relative clauses, and are therefore predictive (Cinque 2010, chap. 5)¹⁰ are usually semantically intersective.¹¹ We note that the innermost direct modification constituents are also intersective in some sense (Devine and Stephens, 2006, p. 476).

Core/peripheral distinction in the argument structure of a noun phrase can be determined by the three parameters listed in § 3.2.4, and again the three parameters, despite being occasionally independent to each other, are often correlated, giving us an impression of a unified parameter of core/peripheral distinction (Huddleston and Pullum, 2002, pp. 439-443). The internal makeup of these arguments may vary cross-linguistically: in English they can be noun phrases, nominals,¹² or even adjectives considering (Huddleston and Pullum, 2002, pp. 439-440): thus we have a *[London]_{location} direction modification [legal]_{argument} expert*. Just as the manner adverb is closer to the main verb than TAM adverbs, so are these arguments of the head noun closer to the head noun than the direct modification constituents (Huddleston and Pullum, 2002, p. 452, [2]).¹³

Finally, at the center of the noun phrase lies the noun stem, which may be a result of derivational morphology. Compounding in nominal derivation should be distinguished from phrasal nominal modification/complementation by tests like coordination and subcategorization – a nominal compound should be more restricted in co-

⁹Semantically, some interpretations are also restrictive, as *the beautiful dancers were admitted into the academy* can be understood as ‘the aforementioned dancers who were beautiful in their dancing were admitted’. This restrictive meaning however can also be derived without postulating more structural complexity: *beautiful dancer* is a property, and *the beautiful dancers* simply invites the listener to recall a group of people with the property of being very well-qualified dancers from the conversational context. If the conversational context contains a group of dancers who were not all good at dancing, then *the beautiful dancers* gains a restrictive meaning, but this is not dictated by the syntactic structure. On the other hand, the interpretation of *the dancers that were beautiful* has to be restrictive (‘recall the aforementioned dancers – now I’m talking about some of them who were beautiful’).

¹⁰Note that there may be several types of *predicative* dependents, for example in English postnominal and prenominal predictive constituents have subtle meaning differences.

¹¹For example, *beautiful* in *a beautiful dancer* may refer to the manner of the dancer dances, or merely be a property of the specific dancer being referred to (in the latter case, its meaning is close to the noun *beauty*). The latter reading is intersective in the sense that the reference of the noun phrase is given by the intersection of the property of being a dancer and the property of being beautiful. The second property has nothing to do with the first, and is equivalent to what a relative clause would contribute to the meaning of the noun phrase.

¹²In the sense in Huddleston and Pullum (2002), not in the Latinist tradition.

¹³It seems Huddleston and Pullum (2002) doesn’t consider the possibility of peripheral adjective arguments, which are still to the right of direct modification adjectives.

ordination (*back-* and *tooth-ache* for example is marginal in English) and may have a subcategorization frame that's totally not predictable from its branches. In Classical Latin real compounding is no longer productive.

Long noun phrases that fully demonstrate the aforementioned complexity are hard to find in nature texts. In English we can construct this example: [*the company's*]_{subject-genitive} [*three*]_{num} [*EXTREMELY AWFUL*]_{prenominal RC} [*“attractive*_{evaluative} *young*_{age}]_{direct modification} [*movie*]_{argument} [*superstar*”]_{head: compound} [*aboard*]_{postnominal RC}. In Latin such elicitation is not possible.

Constituent order Similar to the case in clause structure (chap. 9), the internal constituent order of NPs reflect the hierarchical closeness of the dependents to the head noun, but are also strongly influenced by information structure (Pinkster 2015, § 11.75; Devine and Stephens 2006, p. 484). How these factors are reflected in the exact constituent order varies from author to author and from time to time (Devine and Stephens, 2006, p. 484). We further note that even in English, constituent orders do not transparently reflect the relative scopes of noun phrase dependents: participle modifiers, for example, seem to have higher scopes, but they regularly appear after the head noun, without reflecting their syntactic positions in the pre-head linear order (Cinque 2010, chap. 5).

The genitive argument of a deverbalized noun corresponding to the subject usually appears to the left of the genitive argument corresponding to the object (Devine and Stephens, 2006, p. 316), and indeed we observe that among seemingly direct modifications, “inner” ones, i.e. ones corresponding to the material, etc., tend to appear to the right of “outer” ones i.e. ones that require information from the conversational context and are not themselves properties and non-intersective (Devine and Stephens, 2006, pp. 481-482). Determiners tend to appear at the leftmost position. Therefore the aforementioned noun phrase structure is indeed reflected on the surface linear order.

The head noun is frequently weakly topicalized (scrambling; c.f. § 3.2.7). We know this is an information structure operation, because after the head noun has already appeared in the utterance, it is often no longer fronted (Devine and Stephens, 2006, p. 484). Scrambling makes the head noun a “little subject” in the noun phrase, which is supposed to be more or less independent of the semantics of what follows it, and this may hinder the interpretation of non-intersective modifiers, and therefore the head noun often appears between the non-intersective direct modifiers and the intersective direct modifiers (Devine and Stephens, 2006, p. 485). This analysis is corroborated by the fact that semantically empty nouns easily appear to the left edge of the noun phrase: *beautiful* in *a beautiful person* can never have a non-intersective meaning as it has in *a beautiful dancer*, and hence nothing blocks the leftward movement of the head noun.

The default constituent order is attributive-head noun (Allen and Greenough, 1903, p. 396).

3.3.2 Case endings and prepositions

Structural cases in clauses All cases are not born equal. In Latin we have two *structural* cases licensed in clauses: the nominative and the accusative, which are purely decided by the syntactic environment and don't have much semantic significance. 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 structural use of the accusative case labels the object.

Structural case in noun phrases A complete overview of its distribution reveals that the genitive case is a structural case representing *any* type of nominal dependency and not just possession. Some of the nominal dependencies can be metaphorically understood as possession, but others can't (Oniga, 2014, p. 244): it's hard to explain genitive of quality in terms of possession. This is comparable to English *of* in *a man of great talent* and *it's so kind of you*, which seems to be a purely structural phenomenon and is not based on metaphor; and indeed what is considered as prototypical possession also has strong cross-linguistic variation (Dixon, 2010, pp. 262-263).

Directional and locational constructions Directional and locational constructions appear in both clauses and noun phrases, and we can observe several inherent uses of cases in them. In directional constructions, the ablative case marks the source. The accusative case has an inherent case usage: it marks the *goal* in a directional construction (Oniga, 2014, p. 238). The dative case also marks the goal, but this is only possible when the path is marked as a preverb.

In (semantically more static) locational constructions, the location argument is always ablative.

It seems we have a clear division of labor in directional constructions in Latin: morphological cases are for marking the type of the location, while prepositions are for marking the path (§ 6.4.1).

Inherent case in clauses Besides the uses of cases in directional and locational constructions in clauses, there are several other important clausal inherent cases, which marks arguments mentioned in § 3.2.5: the instrument or similar arguments is marked by the ablative, the psychological causer, i.e. the stimulus, is marked by the genitive, the target or benefactive or experiencer is marked by the dative.

Prepositions may also be used to explicitly mark the type of the argument structure. The preposition *cum* for example marks the comitative construction (§ 3.2.5, § 6.4.1). In these cases the morphological cases seem to be mechanically dictated by the preposition and may be purely realizational, without labeling any independent syntactic information.

3.3.3 Determiners

Box 3.17: Determiners

- Demonstratives
- Quantifiers
- Identity

3.3.4 Numerals

3.3.5 Attributives

The problem of distinguishing core and peripheral constituents in the noun phrase is as hard as it in the clause (§ 3.2.5). This section only discusses adjective or numeral attributives in detail. For in-depth discussion of relative clauses, see chap. 11.

3.3.6 Core argument structure

3.4 Adjectives and adverbs

Verb-like and

3.5 Parts of speech division

3.5.1 Theoretical caveats and typological information

Difference between syntactic, morphological and lexical part of speech labels

We have four prototypical syntactic positions for roots: the noun as the head of a noun phrase, the verb as the head of a clause, the adjective as the head of a direct modification to a noun phrase (Cinque, 2010, p. 23), and the adverb as the head of a direct modification to a clause. They are defined by virtue of appearing at the centers of clauses or noun phrases above and are about *syntactic positions* of roots, which are usually quite universal. We can also define other syntactic positions based on language-specific non-universal syntactic properties. A nominalization construction, for example, is a type of noun phrases that admits a root categorized as a verb at its center. and we can then say that what appears at its center is a *verb used as a noun* or something similar. These syntactic labels are literally *parts of speech*. Note however parts of speech tags defined in this way are *not* properties of *the root*: for example a root may both appear as a regular noun and as a regular verb in the sense defined above, and it's impossible to discuss whether the root is a noun or a verb without the syntactic context.

There are other ways to define *parts of speech*. A second type of way to define these concepts is realizational morphology. Adjectives (defined in the first sense) may have two ways of inflection (as in, say, Japanese), or they may inflect just like nouns (defined in the first sense; this is the case in Latin). In Latin we have nominal and verbal morphology (§ 3.5.2) as well as so-called PARTICLES, but there is no specific morphological template for adjectives in the first sense.

How a root enters grammatical constructions and is morphologically realized (specified by the first and the second sense of parts of speech above) is guided by labels in the *lexicon*. This is the third way to define *part of speech labels*: a part of speech tag in this sense should give us some *lexical* information about the possibility of the root's parts of speech in the first and the second senses; it's about how the roots are *lexically categorized*. This is the job of the lexicographer, and this is how the term *parts of speech* is used in most cases: it's essentially about how to write a dictionary so that you can look for a root and use it wisely to make sure its part of speech in the first and second senses is not messed up. In Latin, therefore, the tag FIRST CONJUGATION VERB tells us a bundle of information: the root in question here regularly appears at the centers of clauses, may participate in certain nominalization constructions, in its morphological realization agrees with the subject of the clause in number and person, has certain endings, etc.

(A fourth definition of parts of speech tags is semantic: what refer to objects are noun-like, etc. But often this is not very useful for grammatical descriptions. The distinction between nouns for objects and nouns for events may have grammatical

significance but usually it can also be established according to more syntactic criteria, like argument structures.)

Lexicon as repository of “idioms” The third sense of parts of speech can be understood as idiomization: therefore a FIRST CONJUGATION TRANSITIVE VERB is to be understood as the complex of a root categorized as a verb (in the first i.e. syntactic sense) in a certain type of clause construction (i.e. transitive subcategorization) and placed into a certain morphological template. Recording this form and *only* this form into the lexicon implicitly forbids placements of the root into other syntactic and morphological parts of speech.

In this way the lexicon may contain construction with different sizes. Larger constructions, like the verb with a directional prefix, or even a whole verb phrase (e.g. *kick the bucket* in English) may have developed their own meanings, and they are to be recorded in the lexicon as well. We will expect something like *treatment*, which is the root *treat* plus the nominalizer *-ment* and the subcategorization frame introduced by it, to be recorded as an entry in a dictionary. Note that there is no guarantee that parts of speech labels in the third senses that are defined for roots still work for these larger constructions: in Mandarin Chinese, for example, reduplicated adjectives seem to form a distinct lexical class (Paul, 2014, § 5.3).

A syntactic unit with a fixed meaning recorded in the lexicon is *lexicalized*. This sometimes may lead to unit to be *fossilized*: certain syntactic operations may hinder its non-compositional interpretation and hence be unacceptable. This kind of fossilization may eventually lead to *syntactic fossilization*, in which the internal structure of the syntactic unit collapses, altering its behaviors (a typical example being some English prepositional verbs allowing passivization), and *grammaticalization*, where the syntactic unit becomes a part of the grammar (e.g. complement-taking verbs become auxiliary verbs, which appear in parts of the conjugational paradigm and have to appear in a given order).

Openness and content/function distinction Putting together the degree of grammaticalization, the openness of a part of speech, and the syntactic position of the units, we have the following classification of what may appear in a dictionary:

1. prototypically lexical (open to new members, large in size, being able to head noun phrase, etc.), or
2. closed but still lexical (a class with no new member allowed but still large in size, being able to head noun phrases, clauses, etc.),
3. a group of pronoun-like things, which appear in syntactic environments similar to those of nouns, verbs, etc., or
4. a finite list of members that are pure markers of grammatical categories.

The first two types are usually considered lexical. The first three types are usually considered to have “real” part of speech labels: a member of the fourth type is always attached to a construction headed by something in the first three types of parts of speech and doesn’t really need its own part of speech label and is prototypically functional, not lexical.

But really, in the third sense of parts of speech, only the first two types of items have real parts of speech: grammaticalized syntactic items they are a part of the grammar by definition and strictly speaking should not be included as a part of the lexicon.

But frequently they are for convenience, and also because the lexical/functional distinction can't always be practically established, as said below.

The distinctions between the four type of course are not categorical. Regarding openness, the most important issue is that some so-called closed classes still admit new members albeit slowly. The distinction between lexical and functional categories is hard for a classical language, because this distinction needs native speakers' judgments on subtle cases and even psycholinguistic data (in acquisition, agrammatism, etc.), which is of course not available to Latinists. Additionally, grammaticalization is not synchronic among speakers, so even if we do have access to fine details of acceptability judgments, what we find will likely be author-specific and do not lead to any generalization to Latin as a whole. The distinction between the third and the fourth types may be problematic for, say, demonstratives, which can be a determinative marker as well as a whole NP on its own. Finally, as is mentioned above, prototypical lexical words can be seen as roots plus "categorizers," and in languages with lots of etymologically related nouns and verbs, we can even argue that the noun and verb endings are of the third type. This however is merely a terminological issue.

Summary of lexicographic methodology We have seen that parts of speech tags have multiple senses, and the decision to define parts of speech tags that appear *in a dictionary* should consider both syntactic and morphological properties (which may also be abbreviated into parts of speech tags), which needs language-specific analyses and require a good observation of constructions available in the language. The lexicon contains are various idioms in the language, and we place parts of speech tags in the lexical sense beside each entry to convey their syntactic and morphological properties in a brief way. Although grammatical markers in theory are not a part of the lexicon, because the content/functional distinction shows diachronic and populational variances, they are often included into dictionaries for convenience.

Depending on the peculiarities of the grammar of a language, the dictionary may be organized in different ways. Arabic doesn't have many compounds and has very regular derivations, so Arabic dictionaries are always organized according to (uncategorized) roots. Dictionaries for Mandarin Chinese are based on "characters" because Sinitic roots are mostly monosyllabic while both "true" compounding and nominal attributive constructions (c.f. English forms like *data processing*) are very, very frequent, so giving a long list of words and lexicalized noun phrases is probably not a good idea. In Classical Latin, compounding is shunned, and although affixation is frequent, the meanings of the derivatives are often obscured (for example, the meaning of the prefix *re-* is not always clear), so synchronically many derived words are nothing different from roots. Categorization of a root into several different syntactic parts of speech is indeed possible in Latin but is not as frequent as in e.g. Arabic or Chinese and derivation morphology can be reasonably well described based on notions like "this word turns into another word", without mentioning too much about roots. Therefore Latin dictionaries are never based on roots but on *words*. The same can be said for English.

In the rest of this section we discuss lexical categories (i.e. parts of speech defined in the third sense) attested in Latin.

3.5.2 Lexical categories with inflectional paradigms

Some Latin word classes can be defined easily via morphology and these classes prove to have uniform morphosyntactic behaviors. Words with large inflection paradigms

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 nouns, verbs, and adjectives are all prototypical lexical categories; there seems to be no closed lexical categories (like Japanese verbal adjectives) in Latin.

Nouns Nouns prototypically head noun phrases. They are declined for case and number (§ 5.2), and the features also spread to other nominals in the NP by agreement.

According to their meanings and ability to license NP dependents, descriptive parameters of Latin nouns include inflection class, valency, compatibility with single or plural number, and **TODO**

Verbs Classification parameters of the Latin verb include the conjugation class (chap. 7), the syntactic and semantic argument structure (chap. 8), the event structure (**TODO**: Compatibility with TAM),

3.5.3 Latin “particles”

Parts of speech without much inflectional morphology include **PREPOSITIONS**, **ADVERBS**, **INTERJECTIONS**, and **CONJUNCTIONS**, traditional known as **PARTICLES**, although in modern usage the term *particle* usually refer to grammatical items, and therefore at least some adverbs should be excluded from that class; also, some adverbs have inflection with respect to *grade* (comparative and superlative). It’s often said that adverb class and the preposition class have a large overlap: often a preposition has an intransitive counterpart, which is similar to a prototypical adverb – but see the discussion below about the lexical/functional distinction in prepositions. Conjunctions may be seen as “prepositions for clauses”. The functions and etymologies of particles are highly diverse.

Adverbs The category **ADVERB** is often posited as a catch-all category for any lexical words that are not nouns, verbs or adjectives.

Box 3.18: Definition of adverbs

How to define the **ADVERB** class in Latin?

A majority of adverbs historically originating from fossilized case forms.

Prepositions The preposition class usually shows great diversity both within a language and among different languages, and what are known as prepositions may be lexical words or function words (e.g. Garzonio et al. 2021). In certain languages, some prepositions are obligatorily selected in certain grammatical constructions and have to form a hierarchy when two or more of them appear together, and hence may be seen as a case particle: they are phonological realizations of grammatical relations, not lexical words. Other prepositions are just adverbs with complements. Several case particle-like prepositions have been discussed above; a detail lexical/functional classification can be found in § 6.4.1.

3.5.4 Latin pro-forms

For categories in the third type, Latin has pronouns and pro-adverbs; depending whether we treat prepositions as heads: 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, and these prepositions can be understood as adverbs selecting a complement. TODO: adverb modifying PP This is also the case of adverbs: some adverbs seem to be periphrastic markers of TAM categories and therefore may be considered as a part of the grammar, while others seem to carry “real” meanings.

3.5.5 Diagram of Latin word classes

Fig. 3.2 is a visualization of the classification of Latin word classes.

Parts of speech absent in Latin 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).

3.6 Reading Latin

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. According to these grammatical items, segment the sentence into large parts: noun phrases, relative or complement or adverbial clauses, verbal system, etc.
4. 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 (§ 7.2.3); *-um* may be second declension accusative, but there are other possibilities (Table 5.1).
5. Use constraints like “the preposition *in* licenses the accusative case or the ablative case” to narrow the possibilities identified above.
6. 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 § 13.3.1.

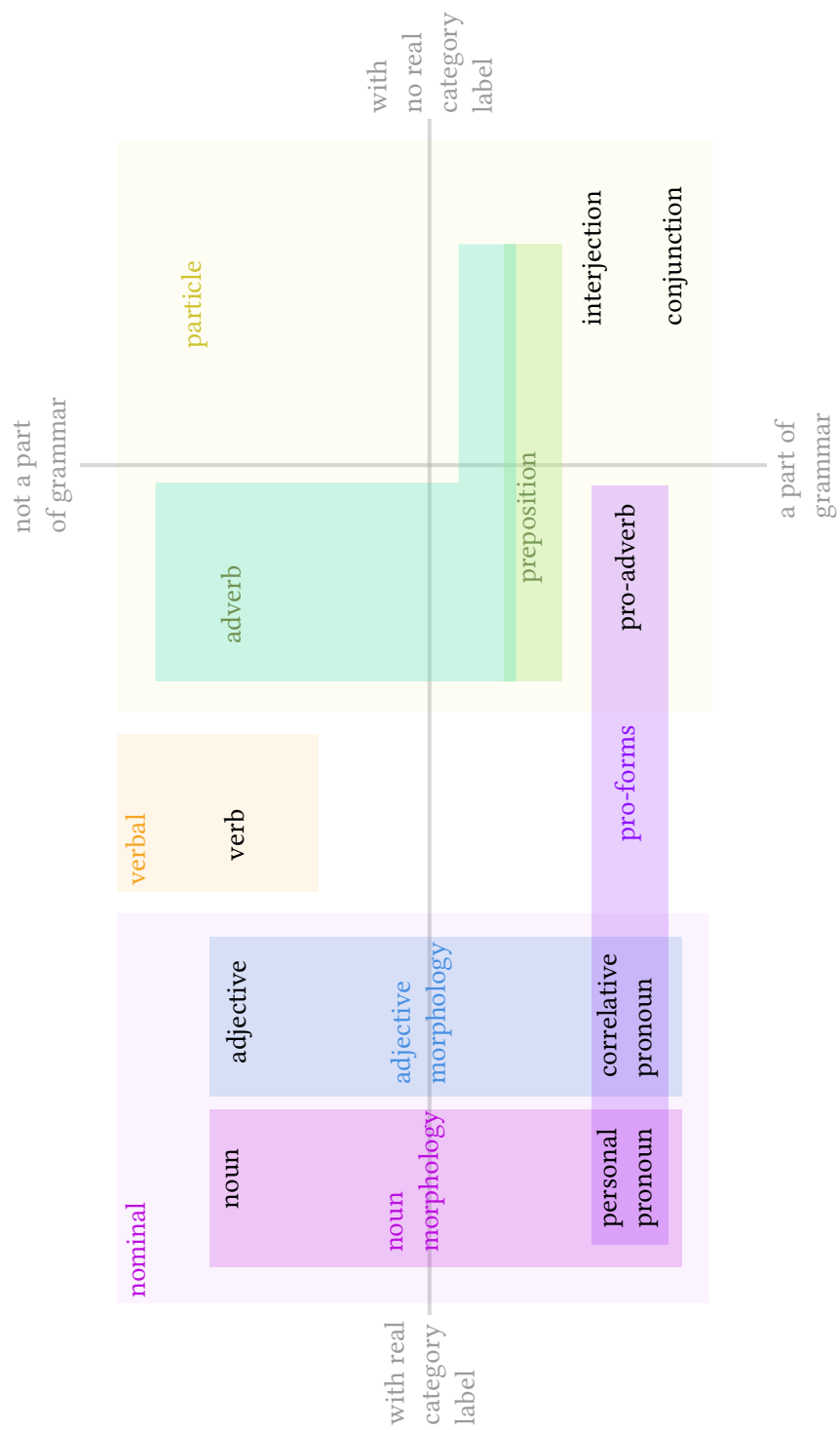


Figure 3.2: Latin word classes

Chapter 4

Pro-forms

4.1 Overview

Pro-forms can be divided into pronouns and correlatives.

4.2 Personal pronouns

Latin pronouns are complete NP themselves: no attributives should be attached to them. Pronouns are declined for case, gender and number, and they also can be governed by prepositions.

It should be noted that what counts as third person pronouns is not uncontroversial among reference books regarding Latin; some place the reflexive *sui* series into the third person column in the table of personal pronouns, citing the fact that the *is*, *ea*, *id* series mark the gender category, while the first and second person pronouns do not.

4.3 Reflexive pronouns

It can be seen that *se* lacks nominative forms. In some circumstances, the reflexive *se* looks like the subject; but these cases are more appropriately analyzed as reflexive usages of transitive verbs (1).

- (1) ...necessarium est primo investigare de ipsa sacra doctrina, qualis sit, et ad
quae [se extendat]

‘It is necessary to first investigate into this sacred doctrine, of what kind it would be, and what it would extend itself to (i.e. its nature and extent).’ (*Summa*, I q. 1 pr.)

4.4 The “basic” correlatives

Latin correlatives can be classified in Table 4.1; the rows correspond to their immediate roles in the clause (“basic” i.e. NP heads, place, source, time, etc. that are otherwise expressed by prepositional phrases), while the columns correspond to their meaning

and/or relation with a more precedent in a higher position, like whether the word is a demonstrative or maybe a relative pronoun.

The form of correlatives in the “basic” can be summarized as the follows:

- *Indefinite*. They may be homonyms of *qui* or *quis*, or they may be formed by adding *ali-* or *-dam* to *qui* or *quis* (with possible internal alternations), or maybe they are formed in the format of *ali-quis-X-quam*.
- *Indefinite relative*. Formed by reduplication of *quis* forms.
- *Other*. Always take a form similar to *alter* or *alius*.

Here is a guide for quickly determining the meaning of a “basic” pro-form:

- What starts with *ill-* is a distal demonstrative.
- What starts with *ist-* is a medial demonstrative.
- What starts with *h-* is a proximal demonstrative.
- What starts with *s-* is a reflexive pronoun, or maybe a reflexive possessive determiner like *suus*.
- What starts with *qu-* is an interrogative pronoun or a relative pronoun or an indefinite pronoun or an indefinite relative.
- What starts with *null-* is a negative pronoun.
- What starts with *omn-* is a collective pronoun.
- What starts with *ips-* is an identity pronoun (‘this exactly ...’).
- What starts with *alter-* is an “other” pronoun.
- What starts with *ali-* may be an “other” pronoun, or an indefinite pronoun.
- What ends with *-dem* and starts with a third person pronoun is an identity pronoun; sound changes are possible. For example *īdem* or *eandem*.
- What ends with *-dam* and starts with *q-* is an indefinite pronoun.

Some forms can be easily confused with correlatives. They include:

- *quod* used to introduce a finite complement clause.
- *quia* as a conjunction of a reason clause.

4.4.1 Demonstratives

Latin has proximate, medial, and distal demonstratives: the first refers to something close to the speaker, the last refers to something far away, and the second refers to something in between, probably near the listener.

Table 4.1: Classification of Latin correlatives

	"question"					
				demonstrative		
	interrogative	relative	indefinite relative	proximal	medial	distal
	identity	collective	negative	indefinite	collective	negative
nomininal head						
	basic dual					

4.4.2 Interrogative and relative pronouns

4.5 Degree

The *tam quam* pair may be seen as a pro-form that takes the place of the degree expression of a clause.

Chapter 5

Nominal morphology

5.1 The structure of noun

5.1.1 The inflectional endings

Latin nouns are declined for case and number, which is agreed upon by other nominal words in the NP. The structure of the Latin noun is just the stem plus inflectional ending with the case and number categories fused into one suffix (§ 5.2). There are five declension classes in Latin.

There exists a discrepancy in recognizing the stem of a noun. The traditional approach is to find the common part of all case forms of one nominal lexeme; thus *rosam* ‘rose-SG.ACC’ is analyzed as *ros-am*, with *ros-* being the stem. Allen and Greenough (1903, p. 17) documents the paradigms of all the five declension classes in this approach. The full list of attested noun endings is Table 5.1. The list is still not the full picture of Latin nominal inflection: stem alternation is seen in the third declension (§ 5.2.3), and this alternation is frequently directly exposed at the end of a nominative and accusative noun.

Table 5.1: 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.ABS
-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.ABS
-ē	V, SG.ABS
-ei/-ēi	V, SG.GEN, SG.DAT
-em	IIIFM, SG.ACC; V, SG.ACC
-ēbus	V, PL.DAT, PL.ABS
-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, SG.ABL
-ibus	IIIFMN, PL.DAT, PL.ABS; IVFMN, PL.DAT, PL.ABS
-is	IIIFMN, SG.GEN
-īs	I, PL.DAT, PL.ABS; IIMN, PL.DAT, PL.ABS
-ō	IIMN, SG.DAT, SG.ABS
-ōs	IIM, PL.ACC
-ōrum	IIMN, PL.GEN
-r	IIM, SG.NOM, SG.VOC
-ū	IVFM, SG.ABS; IVN, SG.NOM, SG.DAT, SG.ACC, SG.ABS, 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 5.1: Things that look like case endings but are not

There are several possible cases where something seeming to be a noun case ending turns out to be something else. To name a view:

- The ending *-as* is not always the first declension plural accusative ending; *-a-* could be a part of the stem of a third declension noun. Consider *vērītās*.
- The ending sequence *-io* can be found in Table 5.1 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).
- A word ending with *-us* could be an adverb: *-tus* is an adverb suffix.

Another approach – informed by historical comparison with other Indo-European languages – make use of the thematic vowel in declension endings that can also be found in Table 5.1 (Oniga, 2014, pp. 45, 63). Thus, the first conjugation singular ac-

cusative ending *-am* is *-a-m*, with *-a-* being the thematic vowel; we may also attach the thematic vowel back to the stem in the narrow sense *ros-* and redefine the complex *ros-a-* as the stem of *rosam*. In this way, the declension endings can be summarized in a neater way. A comparison between reconstructed PIE noun and pronoun endings and early Latin endings is given in [Clackson and Horrocks \(2011, p. 14\)](#).

Table 5.2: PIE nominal case endings

	singular	plural
NOM.M/F	*-s	*-es
NOM.N	*-m/∅	
ACC	*-m	*-ms
GEN	*-es/ *-os	*-om
DAT	*-ei	
ABL	*-ōd/∅	*-b ^h os/ *-ōis

Box 5.2: PIE ending

- What's the neutral plural ending?
- Stem's selection of endings in PIE

5.1.2 Derivational morphology

Inside the stem we find a list of

5.2 Declension of regular nouns

5.2.1 The first declension

5.2.2 The second declension

The stem-final thematic vowel in the second declension is historically *-o-*. Because of sound changes, the NOM.SG ending becomes *-us*,

5.2.3 The third declension

The third declension is a big tent containing several subclasses. Nominative singular endings attested in the third declension include *-s*, *-t*, *-x* (i.e. *-cs*) ([Allen and Greenough, 1903, § 53](#)).

5.3 Uses of cases

The roles of the five cases are not symmetric.

On the other hand, the rest cases are *inherent* cases: they have direct 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. Sometimes inherent cases are realized as prepositions.

Historical evolution of case morphology From PIE to Romance languages, the case system in Latin has progressively eroded. There are two ways to simplify a case system: it’s possible that two morphological cases are merged altogether, which means the relevant grammatical relations are no longer realized in nominal morphology (e.g. § 5.3.2); it’s also possible that the structure of the case paradigm is retained, although for some declension classes, some of the case forms are regularized (e.g. § 5.3.1).

5.3.1 The nominative

Historical evolution in Vulgar Latin The first declension plural nominative ending *-ae* started to be replaced by the accusative *-as* starting at no later than first century A.D. but arrived at Gaul only after the fifth century. Note that this replacement is not equivalent to elimination of morphological marking of the subject/object distinction: the *-us/-um* alternation is still kept (Herman, 2010, p. 55). The motivation may be analogy of the neutralization of the nominative and accusative endings in the third declension, and/or analogy of Oscan and Umbrian nominative plural ending *-as*.

5.3.2 The accusative

In Vulgar Latin A general tendency in Vulgar Latin is the ablative in prepositional constructions being replaced by the accusative (Herman, 2010, p. 53). Similarly, internal arguments which should be genitive or dative or ablative tend to be accusative in Vulgar Latin (Herman, 2010, p. 54). This suggests the accusative being recognized as the “default case”.

5.3.3 The dative

The dative case is an inherent case assigned to the benefactive, the experiencer, and the purpose (Oniga, 2014, p. 251); thus it’s the case assigned to the indirect object (§ 8.6.1) as well as many adjuncts TODO

5.3.4 Ablative

so its distribution is similar to the dative case: it appears in source and instrument indirect objects (TODO) as well as various adjuncts.

A source argument can be the position from which something 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.

Chapter 6

Adjectives and adverbs

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.

6.1 Declension of regular adjectives

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

6.2 Arguments of adjectives

6.2.1 Dative

- (1) ... consubstantialem Patri ...
consubstantial-SG.ACC Father-SG.DAT
'consubstantial with Father' (Nicene Creed)

6.3 Comparative construction

6.4 Prepositions

6.4.1 Are prepositions function words or content words?

As is mentioned in § 3.5.3, various degrees of grammaticalization may happen to prepositions. In this section, we discuss Latin prepositions which are better analyzed as grammatical markers instead of transitive adverbs.

Box 6.1: Latin preposition and adverbs

Can *ex* or *in* be analyzed as adverbs? If so they should be able to be modified by other adverbs?

Preposition in valency alternation English *of* is used to introduce the object in nominalization and *by* is used to introduce the deep A argument in passivization. We can therefore say that the first is an analytic genitive case marker, and the second is an analytic instrument case marker.

In Latin, one use of *ab* falls in this category: personal agents in passive clauses are always preceded by *ab*. The ablative case governed by *ab*, in this use, marks nothing: this resembles the fact that English perfect auxiliary *has* selects a past participle or the Japanese verbal conjugation suffixes add CONTINUATIVE or GERUND endings to the verbal complex before them, which can be explained by diachronic reasons.

Preposition sequence in directional constructions In directional constructions, stacking of case markers and/or prepositions in a fixed order and with a finite number of possibilities indicate that the prepositions have already been grammaticalized. This is the case in Finnish or, to a lesser degree, in English (Svenonius 2010; consider English *from behind the wall*).

In Classical Latin there is no clear example of this type: no preposition stacking is allowed in Latin, so coexistence patterns cannot be used to establish a lexical/functional distinction. We however note that directional prepositions in Latin seem to always encode the path and never the properties of the location. Semantically, *ex* encodes a path going away from a given region, while *in* encodes a path going into a given region, and the status of the location as the goal or the source is marked by its case. Tentatively we can analyze *ex* and *in* as grammaticalized markers of the type of the path in the directional construction.

In early Christian Latin, the form *abante faciem domini* ‘from in front of the face of Lord’ is attested (Herman, 2010, p. 24), and the path-figure-ground hierarchy is perfectly demonstrated here. This shows that at least in some sub-elite varieties of Latin, *ab* and *ante* are grammaticalized directional markers.

Preposition as modifier in another prepositional phrase It is also possible that a preposition is an adverbial modifier of an existing prepositional construction (Botwinik-Rotem and Terzi 2008; c.f. English *the boat drifted [down from up above the dam]*). This case seems absent in Latin.

Comitative The preposition *cum* regularly marks the comitative construction, both in verb frames and in comitative NPs.

6.4.2 Prepositions as transitive adverbs

Some Latin prepositions are likely as lexical as other lexical words. They are “transitive adverbs”.

The form *circum* appears in several contexts and means a (static) location in space or a dynamic path depending on whether a path is implied by the verb; in all these cases, however, the meaning ‘around sth.’ is always kept. This is very unusual if it is understood as a purely grammatical orientation particle: homophonous grammatical particles appearing in two contexts in principle can have rather different meanings. It is likely a root surrounded by noun phrases expressing the GROUND and the FIGURE in a spatial relation, resulting in a locational expression Mare (2018).

6.4.3 Intransitive prepositions

Many prepositions can also be used without any complements and may thus be classified as an adverb; they can also appear as preverbs.

The preposition/adverb distinction

Chapter 7

Verb morphology

7.1 The structure of the inflectional paradigm

7.1.1 The verb template of the finite paradigm

For a heavily inflected language like Latin it's an appealing idea to start the discussion on clause structure with a surface-orientated analysis of verb structure. The morphological template is shown in Fig. 7.1; some inflected Latin verbs and their parts are shown in Table 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 suffix** (also known as the **tense suffix**), and the person, number and voice marker, which is called the **personal ending** here, following the terminology in Allen and Greenough (1903, § 165).

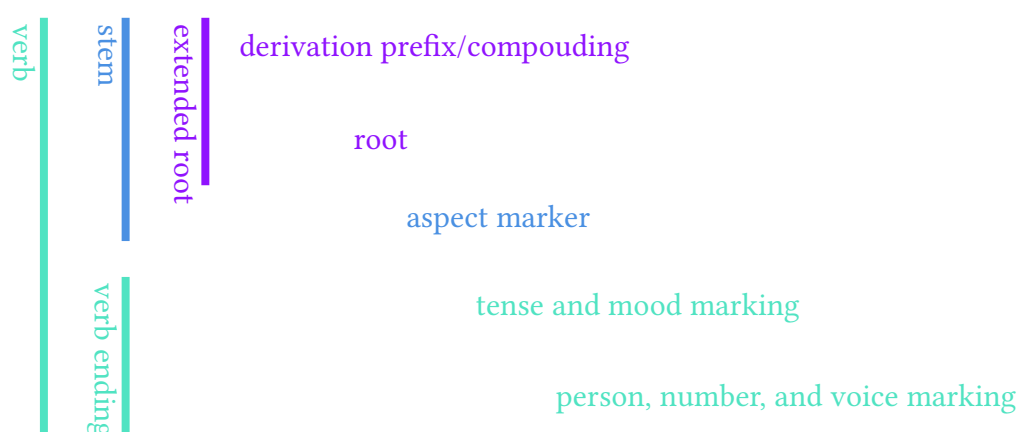


Figure 7.1: The template of Latin verbs. Indentation means linear order and not necessarily constituency structure.

The core stem assumes semi-regular alternations. The perfect marker is often but not always -v- (§ 7.2.3). Although the aspect marker is a part of the TAM marking (§ 7.1.2), 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 (§ 7.2.1).

Contextual allomorphs also exist for other morphemes in Fig. 7.1 (Embick and Halle, 2005, p. 11), as can be observed in Table 7.1. The most salient change is that an alternating vowel may appear between the root and the aspect suffix (if any), which is also known as the **thematic vowel**. It is the residue of the PIE ablaut and labels the conjugation class (§ 7.2.4), and is subject to morphophonological contextual alternation (§ 7.3.1). The tense affix is determined by “tense” in the sense of traditional Latin grammar, i.e. both tense and aspect (Table 7.3), and has dependence on the conjugation class (TODO: ref). The PERFECT tense also has its own personal endings (Table 7.4).

Box 7.1: More than one thematic vowel?

Phenomena similar to the thematic vowel – vowel alternation with no morphosyntactic significance between two morphemes – can be observed between the tense suffix and the personal ending (as in *amābuntur*; § 7.3.3) as well. We may go further and say that the *-imus* ending seen in first person active indicative perfect, as opposed to *-mus* observed in other cases, contains a thematic vowel *i* after the perfect marker *-v-*. Some generalize the concept of thematic vowel and stipulate a thematic vowel position after *every* morpheme Embick and Halle (2003). This is also motivated by evidences from some descendants of Latin (Oltra Massuet, 1999; Oltra-Massuet and Arregi, 2005). Whether this is needed for describing Latin however is

Table 7.1: Examples of Latin finite verbs

verb form	stem		tense and mood	pers. ending	
	extended root				aspect marker
	ext. rt.	thematic v.			
<i>amō</i>	<i>am</i>			<i>ō</i>	
<i>laudāmus</i>	<i>laud</i>	<i>ā</i>		<i>mus</i>	
<i>olēvimus</i>	<i>ol</i>	<i>ē</i>	<i>v</i>	<i>imus</i>	
<i>amāveris</i>	<i>am</i>	<i>ā</i>	<i>v</i>	<i>s</i>	

Below I discuss the subsystems in Fig. 7.1.

7.1.2 TAM categories

Two subsystems can be observed in the tense and mood suffix in Fig. 7.1, which are of corse, TENSE and MOOD. Six TENSE morphological categories can be recognized in total: PRESENT, IMPERFECT, PERFECT, PLUPERFECT, FUTURE, FUTURE PERFECT. There are two MOOD morphological categories: INDICATIVE and SUBJUNCTIVE; sometimes the IMPERATIVE is recognized as the third MOOD. These categories can be established based on sole morphological basis.

We need to note that the morphological marking strategies of TAM do not transparently reflect the underlying abstract TAM features (as 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 *morphological realizations* of TAM, and I use default fonts for abstract TAM values. Some other grammars, like Jacques (2021); Friesen (2017), use initial capitals for the former.

Furthermore, two constructions with *syntactically* different abstract TAM features may have comparable *semantics*. Examples include that if readers of a text know it's about historic events, using present or past tenses in the text will not influence its meaning, and that the distinction between simple past and present perfect isn't always considerable. Sometimes this leads to morphological merger of TAM values (Table 7.2). Choosing one of several constructions with almost identical meaning sometimes follow stylistic conventions (§ 7.4).

Marking of primary and secondary tenses Abstract TAM categories marked by TENSE morphology includes at least primary tense (past, present, future) and secondary tense (anterior i.e. perfect v.s. simultaneous i.e. plain): the PLUPERFECT, the PERFECT, and the FUTURE PERFECT all have a perfect meaning.

Point of view aspect and its combinatorics with secondary tense It seems the PERFECT can also be used to refer to any incident in the past, while the IMPERFECT can also describe something happening in the past. The distinction between the two is likely that point of view aspect (imperfective i.e. progressive v.s. perfective): the IMPERFECT has a progressive meaning.

After enumerating possible combinations of the imperfective-perfective distinction and the secondary tense, we find that among four possible options, three are directly morphologically marked: plain (simultaneous, perfective), imperfective (simultaneous, imperfective), perfective (anterior, perfective). The three combinations are often known as the three *aspects* in Latin linguistics. In the past paradigm they correspond to the PERFECT, the IMPERFECT, and the PLUPERFECT, respectively.

The fourth combination, the combination of the imperfective and the anterior aspectual values, i.e. “progressive perfect” (*have/had been doing* in English), does not have an independent morphological marker in Latin. This may originate from the fact that in most situations, the perfect secondary tense presumably has a perfective meaning, as observing an event from the outside (perfect i.e. anterior secondary tense) usually means the internal makeup of the event does not matter (perfective aspect); indeed, English *have/had been doing* is mostly used to convey the meaning that the situation being described is still going on at the reference time, although this is an implication and not a proper part of the meaning of the perfect imperfective (e.g. in *we've been working on this project for at least two years; we finished it today, and today, we proudly present it to you.*, the project development was already finished at the speech time).

The TAM values of morphological TENSES The composition of present, past and future tenses and plain, imperfect and perfect “aspects” gives nine options; these options are further reduced to six because of several additional mergers.

First, the past simple and the present perfect are identified with each other, possibly because of obvious semantic closeness. Also, the imperfective and simple “aspects” are fused when the primary tense is present or future, or in other words, the point-of-view imperfective/perfective distinction is not reflected in the present or future paradigm: only the spontaneous/anterior (i.e. simple/perfect) values of the secondary tense is morphologically relevant.

The six remaining TENSE forms and their TAM values are shown in Table 7.2. In the Latin grammatical tradition, the term **tense** usually refers to these six options, instead of the past/present/future system.

Table 7.2: Latin TENSES

	past	present	future
imperfective	IMPERFECT	PRESENT	FUTURE
simple	PERFECT		
perfective	PLUPERFECT	PERFECT	FUTURE PERFECT

The usage of the six TENSE categories have peculiarities that cannot be completely summarized into Table 7.2; these are discussed in § 7.4.

The function of mood markers Similar fusion between categories is shown in the category of MOOD.¹ 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.

7.1.3 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.

7.1.4 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 TAM 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

¹Dixon (2009) only calls the category of clause type *mood*. Huddleston and Pullum (2002), on the other hand, calls syntactic modality mood and uses the term *modality* for pure semantics. 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. In terms of general linguistics, the Latin *mood* is a mixture of clause type (the real mood) and modality.

verb conjugation is realized like the English passive, i.e. via a copula and the perfect passive participle.

indicative	present/ imperfect/ future/ perfect/ pluperfect/ future perfect	active/ passive	1/ 2/ 3	single/ plural
subjunctive	present/ imperfect/ perfect/ pluperfect			
imperative	present		2	
	future		2/3	
mood	tense	voice	person	number

Figure 7.2: Categories in the finite paradigm

7.1.5 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. The gerund is morphologically the singular neutral of the gerundive (§ 7.5.2.5).

I keep the traditional notion that the two supines are better recognized as separate morphological inflection forms, instead of two specific usages of the perfect participle and the past participle: although the two supines are identical to two other inflection forms in the surface form, that they don't have number agreement or case inflection

means they are products of morphological devices that differ from the participle constructions.²

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”.

7.1.6 The two periphrastic conjugations

There are two additional periphrastic constructions beside the periphrastic forms in the perfect, passive part of the paradigm. In the first construction, the copula *sum* with arbitrary person, number and TENSE inflection (*including* the perfect TENSES) is placed together with the future active participle, to express the meaning of ‘I/you/he/she/it/they are going to do something’, TODO: a inchoative feeling? This construction is always active in voice. In the second construction the future active participle is replaced by the future passive participle; the meaning is ‘something should be done’.

The present active participle is almost never used in periphrastic conjugation. It can be seen that in periphrastic conjugation, the roles of future active participle and future passive participle and the role of passive perfect participle are not comparable: the latter dictates the secondary tense, i.e. it’s perfect, as well as the voice; the former dictate voice and a “lower” aspect/modality category (inchoative or obligation), but not the secondary tense; and in particular, the term *future* is kind of (although not really) misleading: the two future participles carry aspect and modality values with them, which are semantically related to the concept of future; but the future primary tense is not involved.

7.2 Formation of stems

7.2.1 The three verb stems

Latin shows stem alternation that is not completely predictable. All verb 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.

²Similarly, the infinitive (or the “plain form”, since the infinitive is actually a label of clauses) is recognized as a form independent from the present form in English in Huddleston and Pullum (2002, p. 74).

- 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. The reasons are the follows. First, for fluent users, recording actually attested word forms is easier compared with the morpheme-based “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).

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/>

7.2.2 Formation of the present stem

7.2.3 Formation of the perfect stem

The PERFECT stem is not completely predictable from the PRESENT stem. Frequently observed alternations are listed below.

-v- suffix The etymological source of this suffix is not clear.

-s- suffix

Prolonged vowel

7.2.4 Conjugation classes

Depending on the way realization of the paradigm for a verb, Latin verbs are traditionally divided into four conjugation classes according to the thematic vowel of the stem: if the stem ends in *a*, it's a first conjugation verb (§ 7.3.3); if it ends in *e* then we have a second conjugation verb (§ 7.3.4); TODO. The conjugation classes however can't be reduced to the thematic vowel, since there are some contextual allomorphs that can't be explained synchronically in this way (§ 7.3.1, Table 7.3). There are a handful of irregular verbs (§ 7.7) that can't be inflected using rules pertaining to the four regular conjugations.

Another aspect of the inflectional behavior of the verb is whether it's deponent; this also has some implications on the argument structure (§ 7.6).

7.3 The finite paradigm

Due to morphophonological rules, morpheme division inevitably involves controversies. This section follows the analysis in Oniga (2014, chap. 14).

7.3.1 Marking of tense and mood

The contextual alternation of the tense and mood marker is listed in Table 7.3. They are subject to phonological rules and have allomorphs in different conjugation classes. A long vowel after the suffix, if any, is shortened before *-m*, *-r*, *-t*, and *-nt* due to vowel shortening in the final syllable before a consonant, and also before *-ntur* due to Osthoff's Law (§ 2.3.2; Oniga 2014, p. 130). The present indicative doesn't add a suffix after the stem, so the thematic vowel at the end of the stem is directly exposed to the personal ending, and vowel changes in § 2.3 happen.

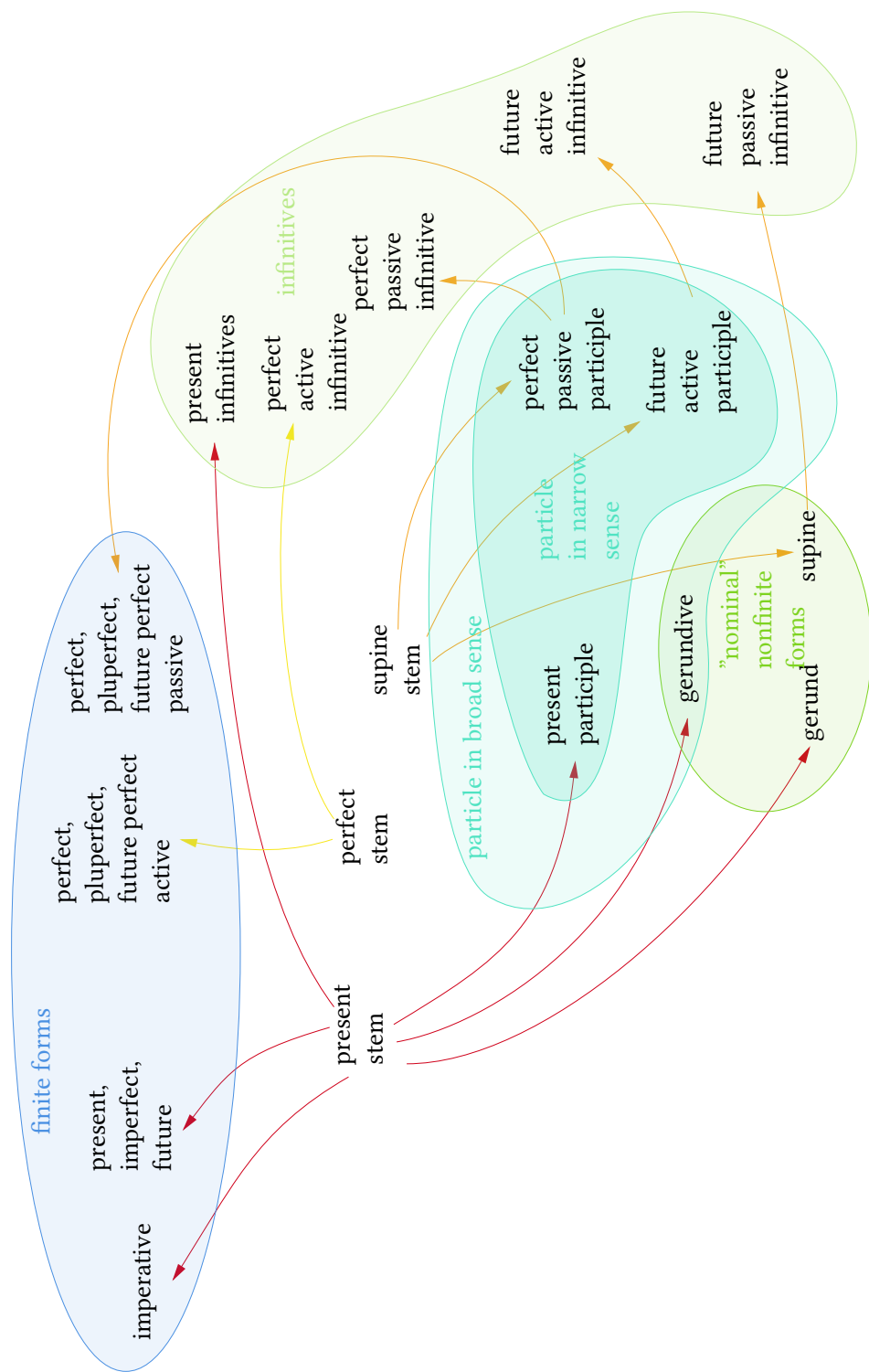


Figure 7.3: How to get all conjugation forms from the three stems

The perfect and pluperfect subjunctive suffixes are only used for the active voice. For the passive voice, periphrastic conjugation with the perfect passive participle is used.

Table 7.3: Tense and mood suffixes

Suffix	Mood and tense	Note
∅	Present indicative	With change on thematic vowel
<i>-bā-</i> , <i>-ebā-</i>	Imperfect indicative	Possibly shortened
<i>-be-</i>	Future indicative	For 1c, 2c; Allomorphs: <i>-b-</i> , <i>-bi-</i> , <i>-bu-</i>
∅	Perfect indicative	With its own personal endings
<i>-erā-</i>	Pluperfect indicative	Possibly shortened
<i>-eri-</i>	Future perfect indicative	Allomorph: <i>-er-</i> (1SG)
<i>-ē-</i>	Present subjunctive	Allomorphs: <i>-a-</i> , <i>-ā-</i> , <i>-e-</i>
<i>-rē-</i>	Imperfect subjunctive	Possibly shortened
<i>-eri-</i>	Perfect subjunctive	For active only
<i>-issē-</i>	Pluperfect subjunctive	For active only; possibly shortened

Box 7.2: Shortening or lengthening?

If we just restrict ourselves to the verbal paradigm, it may be attempting as well to consider *-ba-* as the indicative imperfect suffix, since *-bā-* does not outnumber it. But prolonging is rare in phonology: shortening, due to physiological motion control factors (some sound sequences are easier to pronounce), is more frequent. The same line of argumentation can be applied to justify the status of *-rē-* as the somehow canonical subjunctive imperfect suffix.

Perfect indicative active verb forms have their own set of personal endings (Table 7.4). The plural part of the personal endings still contains the usual *-mus*, *-tis*, *-nt* endings, and thus alternatively, we may analyze the verb endings for the PERFECT tense as the follows: the tense suffix is *-ī-* for first-person singular, third-person singular and first-person plural, *-is-* for second-person singular and second-person plural, and *-ēru-* for third-person plural; the first person singular ending is empty, and the second person singular ending is *-tī*, and the rest of personal endings are the same as those of other tenses.

Table 7.4: Personal endings for the PERFECT

Suffix	Person and number
<i>-ī</i>	1SG
<i>-istī</i>	2SG
<i>-it</i>	3SG
<i>-imus</i>	1PL
<i>-istis</i>	2PL
<i>-ērunt</i>	3PL

- The indicative:

- 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.
- 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, .
 - * 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.

7.3.2 The personal ending

Possible personal endings are listed in Table 7.5. In the active voice, *-ō* is used with present indicative, future indicative (first and second conjugations only), future perfect indicative, and *-m* is used with imperfect indicative, future indicative (third and fourth conjugations only), pluperfect indicative, and the subjunctive mood regardless of tense. The PERFECT tense has its own personal endings: in the alternative analysis outlined in the last section, the first person perfect indicative ends with *-ī*, and the second person perfect indicative ends with *-tī*.

The ending *-re* is 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 principal part (§ 7.2.1).

The *-or* version of the passive first person ending is seen in present indicative and future indicative in the first and second conjugations; the latter has the deep form *-be-or*, then then is realized as *-bor* because of vowel deletion (§ 2.3.1).

Table 7.5: Personal endings

Suffix	Person and voice	Note
<i>-ō, -m</i>	active, 1SG	Depend on tense
<i>-s</i>	active, 2SG	
<i>-t</i>	active, 3SG	
<i>-mus</i>	active, 1PL	
<i>-tis</i>	active, 2PL	
<i>-nt</i>	active, 3PL	
<i>-r, -or</i>	passive, 1SG	<i>-or</i> for present and future indicative only
<i>-ris</i>	passive, 2SG	Allomorph: <i>-re</i> (without tense suffix)
<i>-tur</i>	passive, 3SG	Non-perfect tenses
<i>-mur</i>	passive, 1PL	Non-perfect tenses
<i>-mini</i>	passive, 2PL	Non-perfect tenses
<i>-ntur</i>	passive, 3PL	Non-perfect tenses

- 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 .
 - Second-person singular:
 - * *-ris*: compatible with all non-periphrastic tenses and moods.
 - *

Box 7.3: Parsing a finite verb

Follow these steps to parse a finite verb:

- First see whether the aspect is perfect (by looking at the stem part) and the personal ending.
- Then TODO

7.3.3 The first conjugation

Attested thematic vowel alternations include:

- In first person singular present forms, the final *ā* is dropped, because the personal ending is *-ō* or *-or* and the thematic vowel is subject to vowel deletion.
- In active subjunctive *ā-* → *ē-*.

7.3.4 The second conjugation

Attested thematic vowel alternations include:

- In active indicative present forms, *ē* → *e*.

7.3.5 The third conjugation

The third conjugation has two subclasses: the thematic vowel-less case, and the *i*-thematic case.

7.3.6 Periphrastic conjugations

An auxiliary verb construction is a structure that contains one or more auxiliaries apart from the main verb, and yet is mono-clausal and is not a complement clause construction, and the auxiliaries are realizations of the verbal system surrounding the main verb, and not independent lexical verbs. Auxiliary verb constructions realizing grammatical categories that are usually realized by inflectional endings in a paradigm should be seen as a part of that paradigm, and therefore are known as periphrastic conjugations.

7.4 Uses of tenses and moods

7.4.1 The PRESENT

The

7.5 Non-finite forms

7.5.1 The infinitives

7.5.2 The gerund and participles

7.5.2.1 The gerund

The gerund is morphologically a neutral singular noun. The stem is formed by adding *-nd-* to the present stem; in other words, the accusative form of the gerund of the verb is obtained by removing the final *s* of the present active participle and adding *dum*. Note that the nominative case is missing – when a non-finite clause is required in the subject position, it's always an infinitive.

7.5.2.2 The present active participle

The present active participle (i.e. the present participle) is morphologically a third declension adjective (TODO: gender). The stem of the present active participle is obtained by adding *-nt* to the present stem. Equivalently, the nominative singular form – the citation form – is obtained by replacing the *-re* ending of the present active infinitive by *-ns* (or in other words, add *-ns* to the present stem).

7.5.2.3 The perfect passive participle

The perfect passive participle (i.e. the perfect participle or the past participle) can be found by declining the neutral accusative past participle, i.e. the fourth principal part.

7.5.2.4 The future active participle

To get the future active participle (i.e. the future participle), add *-urus* to the supine stem.

7.5.2.5 The future passive participle

The future passive participle, or in other words the gerundive, can be obtained by declining the gerund as if it's an adjective (TODO: declension class); in other words the gerund is morphologically the singular neutral of the gerundive.

7.6 Deponent verbs

Deponent verbs are verbs that are in the passive voice realizationally, either with the corresponding morphological marking (1) or periphrastic marking (2), but still have active meaning. An observed tendency is that so-called “unaccusative” verbs seem to be deponent verbs in Latin, which means the passive voice essentially is the “non-agentive” voice (§ 8.1.2).

- (1) [Confiteor] unum baptisma in
confess-IND.PRES.PASS.1SG one-SG.ACC baptism-SG.ACC into/towards/about
remissionem peccatorum ...
remission-SG.ACC sin-PL.GEN
'I confess one baptism for the remission of sins ...' (Nicene Creed)
- (2) ... qui [locutus est] per prophetas ...
who speak-APRT-SG.NOM BE.IND.PRES.ACT.3SG through prophet-PL.ACC
'...who has spoken through the prophets ...' (Nicene Creed)

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 (§ 8.4.1), 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 (§ 7.2.1) 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 (§ 7.3) 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 7.6):

- 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 § 7.3.1.
- 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 7.6: The indicative paradigm of *sum*

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

In the subjunctive paradigm (Table 7.7), 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 § 7.3.1 works.

Table 7.7: The subjunctive paradigm of *sum*

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

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 Auxiliary verb constructions

An auxiliary verb, despite morphologically being a verb, is embedded into the TAM system of a language. The following aspects can be used to identify an auxiliary verb from a lexical one:

- *Standalone or not.* A lexical verb may appear on its own; an auxiliary verb never does this.
- *Scope and hierarchy compared with other TAM marking.* The relative order of complement-taking lexical verbs can be arbitrary; this is not true for auxiliary verbs.
- *Interplay with argument structure: voice.*
- *Interplay with argument structure: complement types.* A lexical verb usually imposes semantic confinements on the complement
- *Interplay with clause type.* Some auxiliary verbs never appear in certain types of non-finite constructions, because the TAM categories labeled by them are absent in these non-finite clauses.

Latin is usually perceived as a language with few analytic properties. Traditionally only *sum* and its forms are considered auxiliary verbs when they are used with the perfect passive participle to form perfect TAM constructions in passive voice. Some however argue that there are more auxiliary verbs.

Chapter 8

Verb frames

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 (chap. 10).

8.1 Descriptive parameters

8.1.1 The argument-adjunct distinction

A clear complement-adjunct distinction – telling peripheral arguments from core arguments or oblique arguments – is hard to establish in Latin. Below I assess several parameters usually used to draw the distinction listed in [Huddleston and Pullum \(2002, § 4.1.2\)](#):

- *Content of clausal dependent*. Latin peripheral arguments do not necessarily have prepositions.
- *Easiness of topicalization, etc.* Latin is highly free-ordered and therefore all clause dependents can leave their base positions.
- *Obligatoriness*. Latin is also highly *pro*-drop, and even uncontroversial core arguments can be omitted.
- *Licensing by verb*. Oblique arguments are frequent in Latin, as is the case in English (consider *run away from* or *get into*).
- *Anaphora referring to core verb phrase*. Latin doesn't have systematic way to replace the core predicate (i.e. without adjuncts) by an anaphora.

Selection, licensing, and obligatoriness tests can still be used in more subtle ways (e.g. reference in discourse to identify omitted argument)

these criteria are however hard to use for a classical language.

Box 8.1: Argument and adjunct: directional construction

If a directional phrase obligatorily has the subject/object as its subject, then it has to be a complement and not an adjunct.

Thus, despite I'm fully aware that clausal dependents concerning place, in-

strument, 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. Following the traditional notion, the subject, several kinds of objects, the copular complement (Huddleston and Pullum (2002) calls it *predicative complement*) are identified as complements.

8.1.2 Argument as entity or situation

Verbs can be classified semantically into three subgroups, according to what their arguments refer to (Dixon 2005, Part B; Dixon 2010, § 18.5; and Dixon 2009, § 3.3):

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**¹ 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 than 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).

The valency class of the verb is strongly related to but is not determined by the semantics of the verb.

8.2 Prototypical intransitive verbs

8.2.1 The MOTION type

8.3 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

8.4 Copular verbs

8.4.1 The verb *sum*

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

TODO:

¹Here the term means "semantically equivalent to a complement clause construction".

- (1) Nono, utrum uti debeat metaphoricis vel
 tenth whether at.any.rate owe-SUBJ.PRES-3SG.ACT metaphoric-PL.DAT/ABL or
 symbolic locutionibus.
 symbolic-PL.DAT/ABL speech-PL.DAT/ABL
 ‘Whether (it) at any rate uses metaphoric or symbolic speech.’ (*Summa*, I q. 1
 pr.)

8.5 Two place verbs

The case marking of the stimulus argument has some relations to its animacy (Pinkster, 2015, § 4.34).

8.6 Preverb

Roughly speaking, the preverbs are adverbs or preposition incorporated into the verb stem. The licensing condition of the preverb is highly restricted: it only takes place when a PATH category is licensed by the verb (which in Latin requires the event to be telic, and therefore the IMPERFECT tense and the like may be a problem for prefixation)

8.6.1 The prototypical dative construction

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 (§ 5.3.3).
- *Passivization*: indirect objects are always retained in passive clauses. They are never promoted to subjects in passivization.

8.7 Teaching and secondary object

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 8.2: 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 in Huddleston and Pullum (2002). 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.

^aIn 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.

8.8 (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*:
- *Case marking*: a source argument is in the ablative case. It may come together with the prepositions *ex* or *ab*.
- *Passivization*: not available.

Chapter 9

Constituent order

9.1 Configurationality in Latin

Patterns in Latin constituent order are often overlooked in traditional grammar. Still, fine-grained constituency is demonstrated by the relation between *non* and the auxiliary (§ 9.5), radical change of VO/OV frequency when structural ambiguity is controlled (§ 9.8), and usual constituency tests (Danckaert 2017, § 1.6). The *non*-before-auxiliary condition implies a auxiliary hierarchy just like the one in English, although it's not as developed as the latter since Latin is inflectionally rich. The constituency tests hint on at least the subject-VP binary branching. The fact that superficial VO/OV orders may have structural ambiguity means it's likely that some of the constituent orders are comparable to English poetry in imitation of Latin (Allen and Greenough, 1903, § 600), Japanese scrambling and topicalization (TODO: ref).

Available evidence supports the tradition in existing secondary literatures that Latin is thus better described as a discourse-configurational language, with multiple topicalization and focalization structures (Oniga 2014, p. 189; Danckaert 2017, p. 77; Devine and Stephens 2006, among others). Initial positions in clauses clearly bear information structure functions. Constituents that are able to move to the positions include almost everything: arguments, adverbials, the negator *non*, and also the verb (TODO: aux) (Allen and Greenough, 1903, § 598). TODO: is it possible for the main verb to move to the initial point only? Note that fronting of the verb may be used to focus the verb root or the *tense* (Allen and Greenough, 1903, p. 397); this means preposing of the verb is comparable to stressing the verb in English.

Apart from the information structure, prosody is also an important factor TODO

9.2 Broad focus clauses and the neutral order

Pragmatically unmarked sentences have the arguments as pragmatic topics and the rest of the VP are the focus: 'we already know [Baebius], [his army], and [Pinarius] (topics); the piece of new information is that [a transferring process happens involving the three] (focus)'. This kind of sentences, known as "broad scope focus" sentences in Devine and Stephens (2006, p. 15) because the scope of focalization is broad and not restricted to a single argument, demonstrates a constituent order which may be referred as the **neutral order** in Latin (Devine and Stephens, 2006, p. 79), although its frequency – without controlling the information structure – isn't significant higher

than other constituent orders. This order clearly isn't a faithful representation of the argument structure, since the direct object – the argument that is supposed to be the closest one to the main verb – appears far from the main verb. The reason is likely to be that the “neutral order” also marks the aforementioned “unmarked” information structure, in which the arguments are by default topicalized.

Deviation from the “neutral” order may be divided into the following levels:

The usual typological classification of Latin as a SOV language, despite being misleading in suggesting a rigid base order, still makes sense in pointing out that indeed constituents that are “higher” are moved leftwards.

In periphrastic conjugation, the constituent order is subject + object + verb + *sum*. This may also show a mismatch between the dependency structure and the linear order, since if we consider the auxiliary *sum* to be the analytic counterpart of the inflectional suffix and has a higher position compared to the main verb (i.e. the participle), then since the subject and topicalized constituents usually appear on the left side, it also should appear on the left side compared with the participle under it. But the case may just be that the participle is focused and is moved leftward by default, just like the topicalized direct object, so the auxiliary then appears at the end of the clause in the neutral order.

9.3 Positioning of arguments

9.4 Positioning of the verb (without auxiliary)

Box 9.1: Position of the verb

The position of the verb involves a theoretical question: is its appearance away from the unmarked clause-final position due to phrasal movement (the verb root being moved to a new position, carrying all suffixal realizations of TP functional heads together with it), or is it due to being attracted by some sort of functional head (in this case the verb root is just like a head in head movement; similar mechanisms appear in, say, *on the top of the mountain lies a small village*)? Since this distinction is hard to test, I refrain from picking up one explanation.

9.5 Positioning of auxiliary and negation

One piece of evidence suggesting the grammatical status of *sum* is somehow different from a lexical verb is that its position has non-trivial interaction with the position of *non*. The negator *non* usually appears before the verb (Danckaert 2017, § 1.5, TODO: or aux?), and apparent violations seem to be constituent negation as opposed to sentential negation (Danckaert, 2017, p. 43).

9.6 On so-called postposing constructions

Whether postposing exists as an information structure marking device is still not completely clear. It's said that postposing is never used for emphasis (Allen and Gree-

nough, 1903, p. 395), and apparent counterexamples are all “afterthoughts” TODO; but

The postponed subject is likely to be an afterthought, coindexed with a zero pronominal in the rest of the sentence before it Devine and Stephens (2006, p. 87), comparable to English *it kills three people, the wandering puma*.

9.7 Notes on some typologically rare constituent orders

9.8 Historical evolution

Without sentences in which the OV/VO alternation can be alternatively analyzed as topicalization, VO frequency no longer shows significant change as time went by, indicating a well-defined extended verb phrase (Danckaert, 2017, § 1.5, p. 29).

9.9 Information packaging constructions

9.10 Existential clause

In the existential construction, the *sum* verb always appears first (Allen and Greenough, 1903, p. 396).

9.11 Cleft construction

nequitia est quae te non sinit esse senem

9.12 Question formation

Chapter 10

Complement clause constructions

10.1 Infinitives

10.1.1 *Accusativus cum infinitivo*, or the autonomous infinitive

Despite the superficial resemblance to the English object raising constructions, Latin complement infinitives with accusative subjects are licensed even after complete nominalization of the complement-taking verb. Therefore the accusative subject of the complement clause can't be seen as an object of the complement-taking verb, since the nominalized verb no longer takes object in Latin. Therefore, the *accusativus cum infinitivo* construction is comparable to English *for sb. to do sth.*, where the subject of the complement clause is autonomous; the accusative case here is the case assigned to the subject of a non-finite clause where the nominative case is not available. This is not completely unexpected, since even absence of an explicit complementizer is observed in the *ut* clause as well (Oniga, 2014, pp. 290-292).

Chapter 11

Relative constructions

11.1 General comments

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 12

Coordination and subordination

Latin coordination in the nominal system and the verbal system shows strong correspondence, with most conjunction words being shared by the two systems.

Chapter 13

Texts

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

13.1 *Aeneid*

13.1.1 Introduction

- (1) Arma virumque cano, Troiae
weapon(N)-PL.ACC man(M)-SG.ACC=and sing-IND.PRES.1SG Troy-SC.GEN
qui primus ab oris Italiam,
REL.M.SG.NOM first-M.SG.NOM from shore(F)-PL.ABL Italy(F)-SG.ACC
fato profugus, Laviniaque venit
fate(N)-SG.ABL exiled-M.SG.NOM Lavinia-TODO=and go.to-IND.PRES.3SG
litora, multum ille et terris iactatus et alto vi superum saevae
shore(N)-PL.ACC
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 (1), 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 (§ 11.1). The copula is omitted in the *qui* clause.

13.2 Liturgy texts

13.2.1 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 (2, 3), *nomine* and *patris* are third declension nouns, while *spiritus*

is a fourth declension noun.

- (2) In Nomine Patris, et Filii, et Spiritus
 in name-SG.ABL Father-SG.GEN and Son-SG.GEN and SPIRIT(M)-SG.GEN
 Sancti.
 holy-M.SG.GEN
 ‘In the name of the Father, and of the Son, and of the Holy Spirit.’
- (3) – 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.’
- (4)

13.2.2 Nicene Creed

- (5) Credo in unum Deum, Patrem
 believe-IND.PRES.1SG in one-M.SG.ACC God(M)-SG.ACC father(M)-SG.ACC
 omnipotentem,
 omnipotent-M.SG.ACC
 ‘I believe in one God, (the) omnipotent Father,’
- (6) factorem caeli et terrae, visibilium omnium et
 maker-
 ‘maker of’

13.3 Vulgate bible

13.3.1 Excerpts in John 1

- (7) in principio erat Verbum et Verbum erat apud Deum et Deus erat Verbum
 in be.IMPF
 ‘In the beginning’ (John 1:1)
- (8) omnia per ipsum facta sunt et
 all-N.PL.NOM through DEM-ACC make.PPRT-N.PL.NOM be.IND.PRES.3PL and
 sine ipso factum est nihil
 without DEM.ABL make.PPRT-N.SG.NOM be.IND.PRES.3SG nothing.NOM
 quod factum est
 REL.N.3SG make.PPRT-N.SG.NOM 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.’ (John 1:3)

As an example, below I show how (8) 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 5.1 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 5.1: 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*]_{subject} [*per ipsum*]_{instrument:PP} [*facta sunt*]_{verbal complex}]_{coord} et [[*sine ipso*]_{adverbial:PP} [*factum est*]_{verbal complex} [*nihil [quod factum est]*]_{rel}]_{subject}]_{coord}

13.4 *Summa Theologiae*

13.4.1 Introduction

Thomas Aquinas, a theologian and philosopher, was recognized by the Catholic Church as one of the Doctors of the Church, usually known as *Doctor Angelicus* (doctor-SG.NOM angle-ADJ-SG.NOM, ‘Angelic Doctor’). *Summa Theologiae* (summary-SG.NOM theology-SG.GEN, ‘Summary of Theology’) or *Summa Theologica* (summary-SG.NOM theology-ADJ-SG.NOM.F), usually abbreviated as *Summa*, is probably his most known work. As its name implies, the book is a summary of Catholic theology, containing necessary information for beginning theological students and for arguing against frequently seen heresies. Apart from its great value within and out of Christianity, the book is a good example of what Medieval Latin looks like.

One salient feature of *Summa* is it’s written in a rather strict and even dull format. The work is divided into three Parts, each containing many Questions. Each Question (i.e. a topic, like “the nature and extent of this sacred doctrine”) is divided into several Articles (i.e. a specific question, like whether the sacred doctrine is a science). Each Article follows the following scheme:

1. First, a sentence like (9) indicating index of the Article within the Question it belongs to.
2. Then several Objections are raised to support an unorthodox idea, usually in SUBJUNCTIVE.

3. Then the accepted doctrine contrary to the above heretical claims is given (10).
 4. Finally, a comment by Thomas on this Article (11).
- (9) Ad primum sic proceditur.
to first thus proceed-3SG.PASS
‘(Our topic) goes to the first issue. (lit. It proceeds to the first.)’
 - (10) Sed contra est quod dicitur II
but contrarily be.IND.PRES.ACT.3SG what.R.SG.N say-IND.PRES-3SG.PASS two
ad Tim. III ...
to(TODO) NAME three
‘But on the contrary there is what is said in 2 Timothy 3 ...’ (*Summa*, I q. 1 s.c.)
 - (11) Respondeo dicendum quod ...
respond-IND.PRES-1SG.ACT
‘I respond to what is said that ...’

Here we present a glossed Article in *Summa*: I q. 1 a. 9.

References

- Artemis Alexiadou and Florian Schäfer. *Instrument subjects are agents or causers*. Universitätsbibliothek Johann Christian Senckenberg, 2008.
- Ti Alkire and Carol Rosen. *Romance languages: A historical introduction*. Cambridge University Press, 2010.
- Joseph Henry Allen and James Bradstreet Greenough. *Allen and Greenough's New Latin grammar for schools and colleges: Founded on comparative grammar*. Ginn, 1903.
- Irena Botwinik-Rotem and Arhonto Terzi. Greek and hebrew locative prepositional phrases: A unified case-driven account. *Lingua*, 118(3):399–424, 2008.
- Guglielmo Cinque. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press, 1999.
- Guglielmo Cinque. *The syntax of adjectives: A comparative study*, volume 57. MIT press, 2010.
- James Clackson and Geoffrey Horrocks. *The Blackwell history of the Latin language*. John Wiley & Sons, 2011.
- Lieven Danckaert. *The development of Latin clause structure: A study of the extended verb phrase*, volume 24. Oxford University Press, 2017.
- Andrew M Devine and Laurence D Stephens. *Latin word order: Structured meaning and information*. Oxford University Press, 2006.
- 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 and Morris Halle. Latin inflections. *Paper delivered at Going Romance*, 2003.
- 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.
- Dianne Friesen. *A grammar of Moloko*. Language Science Press, 2017.

- Jacopo Garzonio, Silvia Rossi, et al. Functional and lexical prepositions across germanic and romance. In *Continuity and Variation in Germanic and Romance*, pages 450–471. Oxford University Press, 2021.
- 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.
- Giuliana Giusti and Rossella Iovino. Psychological verbs as a vulnerable syntactic domain: A comparative study of latin and italian. *Lingua*, 223:29–45, 2019.
- Paolo Greco. Latin accusativus cum participio: syntactic description, evidential values, and diachronic development. *Journal of Latin Linguistics*, 12(2):173–198, 2013.
- Nadine Grimm. *A grammar of Gyeli*. Language Science Press, 2021.
- József Herman. *Vulgar latin*. Penn State Press, 2010.
- Norbert Hornstein and Mitsue Motomura. Psych verbs, theta roles and reconstruction. In *Proceedings of 2002 LSK International Summer Conference*, volume 2, pages 39–58. Linguistics Society of Korea, 2002.
- 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.
- Jürgen Leonhardt. *Latin: Story of a world language*. Harvard University Press, 2013.
- María Mare. Issues on word formation. the case of latin circum. *The Linguistic Review*, 35(1):121–159, 2018. This paper assumes that the main predicate in a clause whose main verb has *circum-* preverb is the preverb instead of the verb root, which is assumed to be adjoined to the light verb; this is not necessary for the conclusions of the paper; also, saying that the verb root is first adjoined to the light verb and then the complex is merged with the locative construction is not very different from saying that a “locationality” functional head selects the verb root as its complement and the locative construction as its specifier.
- Andrew McInnerney. *The argument adjunct distinction and the structure of prepositional phrases*. PhD thesis, University of Michigan, 2022. The paper claims that in English the post-verbal constituents have a flat structure. I’m not too sure about that, considering all the work done in Cartography.
- Isabel Oltra-Massuet and Karlos Arregi. Stress-by-structure in spanish. *Linguistic Inquiry*, 36(1):43–84, 2005.
- Maria Isabel Oltra Massuet. *On the notion of theme vowel: A new approach to Catalan verbal morphology*. PhD thesis, Massachusetts Institute of Technology, 1999.
- Renato Oniga. *Latin: a linguistic introduction*. OUP Oxford, 2014.
- Montserrat Pascual Pou. A syntactic analysis of instrumental prepositional phrases. *Catalan working papers in linguistics*, 9:053–68, 2001.

- Waltraud Paul. *New perspectives on Chinese syntax*. De Gruyter Mouton, 2014.
- 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.
- Harm Pinkster. *The Oxford Latin syntax: Volume II: The complex sentence and discourse*. Oxford University Press, 2021.
- Peter L Smith. *Greek and Latin roots: part I-Latin*. University of Victoria, 2016.
- Peter Svenonius. Spatial P in English. In *Mapping Spatial PPs: The Cartography of Syntactic Structures, Volume 6*. Oxford University Press, 08 2010. ISBN 9780195393675. doi: 10.1093/acprof:oso/9780195393675.003.0004.
- Michael L Weiss. *Outline of the historical and comparative grammar of Latin*. Beech Stave Press, 2009.
- Niina Ning Zhang. *The syntax of english comitative constructions*. 2007.