Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki in auditorium XII, on the 15th of January 2016, at 12 o'clock.

Timo Korkiakangas

SUBJECT CASE IN THE LATIN OF TUSCAN CHARTERS OF THE 8TH AND 9TH CENTURIES

Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki in auditorium XII, on the 15th of January 2016, at 12 o'clock.

Supervisors: Hilla Halla-aho, Anneli Luhtala

Professor: Olli Salomies

Latin Language and Roman Literature

Department of World Cultures

Faculty of Arts

University of Helsinki

Contents

Acknowledgements

language studies. Three years of peer support from fellow postgraduate students and feedback from senior scholars at Langnet courses and seminars contributed greatly to my linguistic expertise and self-understanding. Other important reference groups for me were the ever-changing community of Finnish researchers at Villa Lante, Rome, as well as the Late Latin-oriented researchers around two international conferences, the Latin Vulgaire – Latin Tardif conference and the International Colloquium of Latin Linguistics conference.

Last but not least, I want to thank my parents who have always supported me wholeheartedly and, as teachers of languages and history, are largely responsible for what I am today.

Helsinki, 16 December 2015

Timo Korkiakangas

1. Introduction

1.1. The objective of the study

The object of this study is the case marking of the subject in early medieval Latin. The study is based on a treebank of Latin charters from Italy. The purpose is to examine whether and how the nominative/accusative-type morphosyntactic alignment changed into a semantically-motivated (active/inactive) alignment in Late Latin before the disappearance of the case system. This kind of evolutionary development has been suggested by previous research.[1] The study will be carried out by analysing the distributions of the two syntactic cases, the nominative and the accusative. The central question to be answered is: which semantic and syntactic factors determine the selection of the subject case in each subject/finite verb combination? This question will be transformed into more exact research questions below.

The proposed study will be implemented applying quantitative corpus methods to the Late Latin Charter Treebank (LLCT). LLCT is a parsed corpus of Latin charter texts (c.200,000 words) written in Tuscany between AD 714 and 869, a time period which I still consider to be Late Latin. The charters are mainly private documents concerning selling, buying, and donating landed property between clergymen and laypersons. The documentary language of the early Middle Ages is formulaic and conservative. However, it has been shown that several spoken language innovations succeeded in infiltrating into the less formulaic parts of charters.[2] As a consequence, charters constitute a unique possibility of gaining a grasp on the language utilised outside literary texts in a period with little surviving evidence. Due to the special textual quality of charter language, it is obvious that philological and linguistic approaches must go hand in hand when examining morphosyntactic alignment in a charter corpus.

The concept of morphosyntactic alignment derives from typological studies on grammatical encoding systems of the nuclear arguments of verb. In Classical Latin, the subject of the finite verb is in the nominative and the direct object in the accusative. In Late Latin, however, the role of the accusative changes. This is visible in the so-called 'extended accusative' which is attested occasionally in non-literary sources. On the basis of these, it has been postulated in previous research that in Late Latin the nominative/accusative contrast was (re)semanticised so that the nominative came to en-

---

[1] E.g. Plank 1985; La Fauci 1988; La Fauci 1997; Cennamo 2009; Rovai 2012.
[2] Fiorentino 1994; Sornicola 2008, 2012, 2013; Sabatini 1965; Larson 1988.

code all the agent-like arguments and the accusative all the patient-like arguments.[3] This semantic alignment is considered a transitory stage in the process that led to the neutralisation of the case contrast in most Romance languages. The primary goal of this study is to determine how often and in which contexts the primary argument of the finite verb, i.e. the subject, is marked in LLCT with the accusative rather than the nominative.

Late Latin charters are the only coherent non-standard material that is large enough to provide the possibility of studying case alignment with corpus-linguistic and, hence, statistical methods. That is why the LLCT treebank was built on these texts. Technically, the research is realised with the computer-assisted procedure of treebanking. Treebanking provides each word of the corpus with lemmatic, morphological, and syntactic analyses. These are then queried in a query program. Finally, the resulting distributions are processed in order to detect statistically significant dependences between the subject case and relevant semantic and syntactic variables.

My aim is to answer the following questions: Does the evidence of LLCT reflect a semantically-based morphosyntactic alignment? If so, is the subject case selection dependent on semantic factors or do syntactic factors also play a role in it? What are the most important semantic and syntactic factors? To what extent do these factors interact with each other? Can transitivity degree be used as a measure of subject case selection? Are there any extra-linguistic factors at play? Are these extra-linguistic factors based on the formulaicity of the charters or are they psycho-syntactic by nature? As a byproduct, I shall outline an overview on how charter Latin can be used to study linguistic variation and change. This includes creating a verb type classification of charter Latin, discussing the transitivity degree in relation to what I call the charter genre, and proposing guidelines for a new model of assessing transitivity in (charter) Latin.

This study uses charter Latin as evidence for the cross-linguistically interesting realignment that is assumed to have taken place in Late Latin. In my estimation, the chosen approach is desirable in Latin linguistics that, for the present, exploits only partly the advances of typologically-inspired syntactic theory. As a consequence, this study is not a comprehensive analysis of all aspects of charter Latin, not even its case system.[4] The focus of the present research is explicitly on the encoding of the subject argument. Although the study will not carry out a chronological comparison within

---

[3] Plank 1985, 291; Plank 1995, 1193. For a criticism of the plain resemanticisation view, see Cennamo 2001b, 18.
[4] There still seem to be functional residues of a genitive/dative case in LLCT.

the corpus, it will provide new evidence concerning the extension of accusative subjects to transitive verbs, i.e. the final stages of realignment process – a topic called for in previous studies.[5]

The beginning of systematic linguistic research on charters dates back to 1927 when Vielliard published a study on the Merovingian charters of Gaul. She was soon followed by Pei (1932), who concentrated on the Northern French documents of the 8th century. Studying Italian charters became possible when Schiaparelli had published the charters of the Lombard kingdom (up to AD 774) in a modern edition in 1929 and 1933. The first, rather schematic, study is Funcke 1938, whose approach was completed in Politzer & Politzer 1953, where the authors conducted a basic quantitative study on Lombard charters from the 8th century. Because the Politzers left their results underinterpreted, Larson (1988) saw it necessary in his detailed study to revise almost everything that was written about the Latin of Lombard charters until then. The present study has found the previous research greatly helpful even though not one of the studies is specifically about the syntax/semantic interface.[6]

The following subchapters of chapter 1 provide the necessary information on the source material, i.e. the charters used as the basis of LLCT. Chapter 2 describes how LLCT is operationalised in order to be subjected to linguistic study. Chapter 3 sets the theoretical context for studying (Latin) case system and morphosyntactic alignment. Chapter 4 and chapter 5 are the analysis chapters proper: the former discusses the syntactic variables and the latter mostly the syntactic variables, whereas chapter 6 examines the interaction of the previously discussed semantic and syntactic variables. Chapter 7 is the conclusion.

1.2. Description of the material

Chapter 1.2. describes the charter material and discusses the advantages and problems involved in using it in linguistic study.

---

[5] Cennamo 2009, 319, 323–324; see also Adams 2013, 249.
[6] There are also other comprehensive linguistic descriptions of spatio-temporally defined charter corpora: medieval Latin documents of the Portuguese territory (Sacks 1941), North Italian charters of the 8th century (Politzer 1949), Spanish charters from the 8th to 11th centuries (Bastardas Parera 1953), and the *Tablettes Albertini*, i.e. Vandal African documents from the late 5th century (Väänänen 1965). Studies that discuss only some minor linguistic phenomenon of charter Latin include Sabatini's (1965) seminal study on the different linguistic parts of charters, Tekavčić's article about the relation of Lombard charter Latin and Old Italian (Tekavčić 1975), a study of the spoken traits of the Ravenna papyri (Lazard 1993; see also Lazard's other articles), and, more recently, a study on the plural forms of the Lombard charters (Faraoni 2014a). I have earlier studied the neuter in the Ravenna papyri (Korkiakangas 2010a).

1.2.1. What are charters?

Medieval documents of the 8th and 9th centuries were written by hand on parchment. The parchment was then folded and stored as proof of the completed legal act. The prevailing document type in Lombard and Early Carolingian Italy was the charter (*charta*, *chartula*). The same also applies to documents written elsewhere in Italy, i.e. in the Byzantine exarchate, but this tradition is not studied here.[7] I include in my corpus all the charter types found in the *Codice diplomatico longobardo* 1–2 of Luigi Schiaparelli.[8] They are mostly private documents, such as sales contracts, donations, exchanges, pledges, and confirmations (*chartae venditionis*, *chartae offersionis/dotis*, *chartae viganationis*, *chartae pignoris*, *chartae confirmationis*), but there are also some public documents, such as judgements (*notitiae iudicati*). I also include *notitiae* and *breves*, which are more informal lists and registers written for private use.[9] For a typical *charta venditionis*, see Appendix 1.1.

A factor common to all these charters is the interest of the local church: they are all related to ecclesiastical businesses, which is also the reason for their conservation in the archives of bishoprics. Indeed, no charters from private archives or collections survive. Charters were not legally authenticated by chanceries or public authorities but by the authority of the church, symptomatic of which is the ecclesiastical status of several scribes (e.g. *notarius et presbiter* or *diaconus*). In the case of non-ecclesiastical scribes, the charters drew authentication from the old notarial tradition and from the credibility of the witnesses.[10]

From the linguistic point of view, it is important to note two facts: first, all document types are highly formulaic by nature and feature more or less the same formulae, except for the judgements, *notitiae*, and *breves*; second, in spite of their formulaicity, the documents are likely to contain parts that may reflect traits of spoken language (see section 1.2.3.). The judgements, which have their own particular formulae, usually also embody much non-formulaic text, as lawsuits typically are not reducible to formulaic expressions. It is specifically the linguistic perspective that does not allow the inclusion of public document types other than judgements. Other public document types, such as royal diplomas or *placita*, which are certainly also juridically and diplomatically far from private documents, tend to present a more elevated linguistic register and, more importantly, were

---

[7] For charters of the exarchate, see Tjäder 1982. In the exarchate, the documents were written increasingly in Greek. On Greek charters, see Zilliacus 1941, 79–107.
[8] Schiaparelli 1929, vii–ix.
[9] Bresslau 1958, 46–48.
[10] Pratesi 1979, 44, 47; Guyotjeannin & al. 1993, 88–89.

often written by the notaries of the royal or ducal court, not by the local scribes even if dated in Tuscany.

In general, charters present the only major possibility to study the Latin of the Lombard and Carolingian Italy because other written products of the period are scarce. Actually, charters, along with law texts, have sometimes even been considered the 'national literature' of the Lombards.[11] This image arises from the relatively vast number of original charters surviving from Italy from the 8th and 9th centuries. In the archives of Tuscany alone, more than 1,100 charters are available from between 685–899. The majority of the charters, over 700, come from the 9th century. Many more Italian and Tuscan charters are transmitted as later copies or in cartularies,[12] but this type of material is not reliable enough for linguistic study, as the original wording and spelling cannot be determined with certainty.

1.2.2. Editing charter texts

This study is based on the Late Latin Charter Treebank (LLCT). The treebank, i.e. parsed corpus, has been built on a digital edition of Tuscan charters that I prepared on the basis of three printed editions.[13] These out-of-copyright editions can be found digitised online. The nature of my edition is dictated by its use in linguistic study. According to the ideal of open data, the charter corpus is available in the Perseus Latin Dependency Treebanks under the public domain licence while the LLCT can be freely downloaded from my GitHub repository (https://github.com/timokorkiakangas/ LLCT/blob/LLCT/LLCT-with-new-attributes.pml.xml).[14] Therefore, only copyright-free material was used. This means that I am not able to utilise the most recent editions, such as Kurze (1974) or *Chartae Latinae Antiquiores* (ChLA) series 1–2 (1954–), but have to rely on older editions which are:

1) *Codice diplomatico longobardo* (CDL) 1–2, edited by Luigi Schiaparelli (1929–1933) for the Tuscan charters between the years 714–774;

---

[11] Bartoli Langeli 2006, 26; Sanga 1995, 84, 86.

[12] See the charters of the Duchies of Spoleto and Benevento in CDL 5 and CDL 4:2.

[13] By 'text edition' or 'edition', I mean the text form that I use as the basis of the corpus (and, thus, for the treebank). By 'corpus', I denote the corpus-linguistically motivated collection of texts apt for the linguistic study pursued here. 'Treebank' is then created by covering the morphologically annotated corpus with syntactic markup.

[14] My digital editions of the texts can be viewed and downloaded at the Perseus Latin Texts site (see Bibliography for the links).

2) *Codice diplomatico toscano* (CDT) 2:1, edited by Filippo Brunetti (1833) for the non-Luccan charters between the years 775–813;

3) *Memorie e documenti per servire all'istoria del Ducato di Lucca* (MED) 5:2, edited by Domenico Barsocchini (1837) for the Luccan charters between the years 775–869.

CDL 1–2 are available digitised and proof-read on the site hosted by the Institut für Mittelalterforschung of the Austrian Academy of Sciences. CDT 2:1 and MED 5:2 have been scanned and optical-character-recognised (OCR) by Google and can be uploaded at the Google Books Library.[15] Before starting the annotation, it was necessary to proof-read carefully the OCR digitalisations and to edit the texts in order to render them suitable for syntactic research. The text of my corpus is meant to serve mainly linguistic analysis, not diplomatic or historical research.

Because the two 19th-century editions are lacking in precision and abound with incorrect readings, I checked all case-endings, as far as it was possible, against the facsimiles of the charters available in ChLA. In this way, it was possible to expand the insufficiently indicated abbreviations in Brunetti's and Barsocchini's editions and to verify the extent of some lacunae in the text. Without the facsimiles, it would also have been impossible to tell which words or passages really were abbreviated in the originals and which ones were abbreviations added by the editors. Especially Barsocchini has left out several repetitive formulaic passages which had to be recuperated. For ergonomic reasons, however, I left unexpanded the 688 standard-form subscriptions that Barsocchini has truncated.[16] Using facsimiles also helped to detect some further errors caused by the optical character recognition of the scanned texts. The last 21 MED charters from the years between 865 and 869 were not yet published in ChLA at the time of creating the corpus, so I checked them directly against the original charters at the Archivio Storico Diocesano in Lucca in February–March 2012.

As a result, the text of the corpus is not that of Schiaparelli, Brunetti, and Barsocchini as such, but a morphologically and syntactically valid hybrid based on the three scholars' work by mediation of my editorial activity. In the name of economy, those original errors that do not affect morphosyntactic study, e.g. several proper names, were often left uncorrected. The graphical choices of the original editions were retained regarding the writing of letters *u*, *v*, and *w*.

---

[15] For the web addresses of the mentioned material, see Bibliography.

[16] E.g. "*Ego Andreas rogatus* ec." instead of "*Ego Andreas rogatus ab Aloni me teste subs*(*cripsi*)".

1.2.3. Role of formulaicity

A distinctive feature of legal documents throughout the history has been their tendency towards standardisation of both contents and language. The legal value of the document was seen essentially to depend on its form. The regular, repetitive parts of the charters are called formulae. By the end of the Late Antiquity, the Roman chancery tradition had developed a highly articulated repertoire of formulae to be used in several types of documents.[17] Many of these old formulae also appear in Lombard and Carolingian Italy along with new types. Indeed, the Italian charters of the 8th and 9th centuries are almost wholly constituted of formulae with only little space for variation. It is to be noticed, however, that at that time the formulae were produced by memory or by comparison to earlier charters at hand, not copied from formularies, as was done later in the Middle Ages.[18] Writing by memory, of course, resulted in a considerable amount of variation.

Charter texts can be roughly divided into formulaic and non-formulaic parts as far as their language is concerned. The less formulaic sections are often called 'free' parts. The least formulaic part of a charter is the so-called dispositive part or *dispositio* that contains the case-specific details of the legal act that made the composition of the charter necessary. The *dispositio* often describes the transferred property, the measures and boundaries of the plot of land, or the sum of money paid. Of course, even the extremely formulaic parts contain changeable items, e.g. the name and the year of reign of the king in the *datatio* clause, but it is the *dispositio* that is the only part in which the scribe could not rely on formulae but had to improvise. In *notitiae*, *breves*, and *iudicata*, which are probative, not dispositive documents, the distribution of formulaic and non-formulaic parts is different, and, in general, they provide more space for improvisation than the normal *chartae*.[19]

It is important not to treat equally the formulaic and the less formulaic parts in linguistic analysis. Their linguistic difference was first observed by Ludovico Muratori in the 18th century.[20] Later, Francesco Sabatini pointed out how crucial it is for linguistic study to analyse separately the 'free'

---

[17] Bresslau 1958, 46–49. This variety can be perceived in document collections, such as *Tablettes Albertini*, Ravenna papyri, and Merovingian charters of Gaul.

[18] Schiaparelli 1933, 3; Pratesi 1979, 88; Amelotti & Costamagna 1975, 215–217. Even a rapid scan of the charters makes it obvious that they were not directly copied from written exemplars: the variation within the charters of one scribe, not to say one town, is striking. Mistakes based on recall cut-offs are frequent, whereas copying-based errors, such as *propter homoioteleuton*, are absent. In Lombard Italy, there seem to have been no notarial schools nor was the documentary practice regulated by law, except for certain general mentions about those *qui cartolas scribent* (*Leges Liutprandi* 91; see Caprioli 1978, 206–208; Azzara & Gasparri 2005, 240–241). See Larson 2000, 162–163, for a possible fragment of a Lombard formulary. For medieval formularies, see Bresslau & Klewitz 1958, 226–235; Rio 2009.

[19] On dispositive and probative parts, see Pratesi 1979, 25.

[20] Muratori 1751, 93–96, especially 95.

and the formulaic parts of the charters, as they represent different linguistic realities and, thus, display different types of errors. The 'free' parts are generally written in a relatively simple language that is thought to reflect, in its deviations from the standard language, the developments of spoken language. The formulaic parts, instead, abound with errors and contaminations arising from the poor command of standard Latin and from misinterpretations of complex juridical phraseology.[21] How all this is taken into account in this study is described in section 2.2.2.

In general, charters have been avoided as a material of linguistic study. This is not only because the editions were lacking or out-dated until very recently, but also because formulaicity is thought to inhibit any reliable linguistic study. It is true that formulaicity poses problems for analysis since the texts are composed of building blocks of different origin and, thus, certain linguistic phenomena are overrepresented and others underrepresented. This is, however, something that simply needs to be carefully taken into account in the analysis. One does not have to confine oneself to the non-formulaic parts as Sabatini does. The formulaic parts can also be studied, but these different parts must be treated separately and yet, on demand, they must be comparable with each other.

The formulae used in the charters not only pose challenges to the linguistic analysis but also offer solutions to some of its problems. Knowledge of the formulae often helps to understand the fragmentary parts of the charters and, thus, makes it possible to determine the syntactic and, rarely, even the morphological analyses of otherwise isolated words or sequences with considerable reliability (see section 2.1.3.). On the other hand, the existence of certain formulae does not justify forcing deviant structures into the mould of formula-based normativity. Wherever something remains that cannot be understood properly, it is excluded from the analysis by specific tags in the treebank (morphological tag "unknown" and syntactic tag "undefined").

The formulae do not only vary according to document type, but also to time and place. Although no written formularies were likely to be in use, the scribes of a certain town or area tended to utilise similar types of formulae in similar document types, which must be seen as an indication of a local scribal tradition. The differences between, say, Tuscany and the Duchy of Spoleto are obvious, as are those between Lucca and the other Tuscan towns – and even between Lucca and some of its

---

[21] Sabatini 1963–1964, 149–150; Sabatini 1965, 975–976; Sabatini 1968, 340; see also Uddholm 1953, 231–232; Petrucci & Romeo 1992, 118; Sanga 1995, 86; Larson 2000, 152–153. Larson (2003, 130) speaks about a *diglossia grafica* between the languages of the two parts of charters. Also Fiorentini (1994, 46–47) and Lazard (2007) demonstrate how charter language reflects linguistic change.

nearby villages.[22] In general, it can be concluded that the Luccan charters give a more polished impression than the charters written in smaller localities, perhaps by less experienced scribes. The situation might also have been the same between the other centres of Tuscany, such as Pisa and Chiusi, and their peripheries, but too little material survives to render this visible.

In this corpus, the chronological variation of the formulae is not considerable: the formulae of Tuscany remain practically the same throughout the study period, except for substantial changes in the *datatio* clause. Neither did the Carolingian reforms profoundly affect Italian documentary practice.[23] A certain formal and linguistic uniformisation, however, can be perceived even with attentive reading especially in the Luccan charters from the 770's onwards. This standardisation may be partly due to the Carolingian reforms, but is mainly a consequence of a further consolidation and organisation of the local political and economic power of the Luccan clergy.[24]

2. Late Latin Charter Treebank

This chapter presents a framework for using early medieval Latin charter texts in linguistic study. The study at hand is based on LLCT, a treebank composed for this particular purpose. A treebank, or parsed corpus, is a text corpus that is endowed with both morphological analysis of each word and syntactic analysis of each sentence. These analyses are realised in the form of machine-readable annotations. In order to evaluate properly the results of the linguistic analysis, their significance and their generalisability, it is necessary to understand the composition of the corpus as well as the treebanking method that technically determines the possibilities of the study. The following three sections describe how the corpus has been constructed (section 2.1.) and which kind of annotation is utilised: textual annotation (section 2.2.) and linguistic annotation (section 2.3.). Section 2.4. explains how the two annotation layers have been connected, and section 2.5. defines the conditions under which LLCT can be utilised to study the subject category.

2.1. Structure of the text corpus

The composition of the corpus is always defined by the purpose of its use and by the amount of material available. The quantitative study of the Latin nominal declension requires a very large cor-

---

[22] Schiaparelli 1933, 3–5.
[23] Bartoli Langeli 2006, 26, 30–33; cf. Guyotjeannin & al. 1993, 96.
[24] Standardisation development will be examined using numerical methods in a future study.

pus to be representative enough. Treebanking proved to be the only realistic method to manage a sufficient mass of text that is accompanied by an extensive linguistic markup. Treebanking also makes it possible to cope digitally with the formulaic nature of charters. LLCT is chronologically balanced in order to be utilised for studying diachronic variation when needed. Since no comparative areal linguistics is intended in this study, the charters were chosen from a single linguistic area (Tuscany) in order to suppress the diatopic variable.

2.1.1. Size of the corpus

The LLCT corpus amounts to a total of 198,696 words or word fragments.[25] The size of the corpus was decided on the basis of a calculation of individual case endings for a certain declension, gender, and number extracted from the approximately 20,000 words of the so-called Ravenna papyri. The Ravenna papyri that I studied previously are a corpus of 59 Latin charters from 5th to 8th century Italy.[26] Assuming that the form category distribution of the Ravenna papyri may be generalised to other Late Latin charter corpora, tenfolding the size of the corpus seems to result in such numbers of occurrences of single case endings that enable detecting statistically significant dependences with a chi-square test when studying subject case selection. In the ten times smaller corpus of the Ravenna papyri, the numbers of some rare case endings were too low to permit reasonable conclusions. Moreover, the corpus size is limited by practicality: in spite of the rare situation where more material would be available, to expand the corpus further would have taken too much time and work.

Table 2.1. Charter distribution inside LLCT.

	Date	Principal writing locations	Charter numbers in original edition	Number of charters	%
CDL	714–786	Lucca, Siena, Pisa, etc.	16–295, App.	180	34
CDT	775–813	Siena, Pisa	4–89	39	8
MED	776–869	Lucca	158–806	300	58
				519	100

---

[25] By word fragments, I mean the remnants of words with a case ending truncated by a lacuna. The word fragments are relatively few and are never included in the morphological analysis.
[26] Korkiakangas 2010a, 114–116.

Table 2.2. Numeric description of LLCT.

	Charters	Sentences	Items[(*]	Words	Abbreviated words	Fragmentary words	Analysable words[(**]	
							N	%
CDL	180	3,398	77,670	67,387	4,158	922	62,307	34
CDT	39	701	18,288	16,342	1,231	60	15,051	8
MED	300	5,128	129,876	114,967	7,745	541	106,681	58
	519	9,227	225,834	198,696	13,134	1,523	184,039	100

[*] words including punctuation   [**] words excluding abbreviated and fragmentary words (see section 2.3.1.)

The number of sentences equals the number of technical annotation units in the Latin Dependency Treebank annotation environment (see section 2.3.1.). In the annotation environment, the texts are automatically split into units by full stops. The punctuation of my edition is chiefly based on the three original editions which differ from each other in their conventions. Even though I split several long sequences into smaller annotation units in order to facilitate the annotation procedure, some units still contain more than one (non-subordinated and non-coordinated) verbal predication.[27] Due to these restrictions, the number of sentences cannot be considered a descriptive feature of LLCT in comparison with other corpora.

Similarly, the total number of words is misleading because all the charters repeat several formulaic expressions, thus multiplying the number of certain technical words. For example, typical documentary word forms that scarcely appear in other text types, such as *manifestum*, *indictione*, or *signum*, appear 133, 576, and 1,384 times, respectively. For the list of the most common words and lemmas, see Appendix 2.1.

The total number of words suitable for linguistic analysis is reduced by the fact that a considerable proportion of them are abbreviated or partly damaged. Once those abbreviated and fragmentary words are excluded, 184,039 words remain for morphological analysis. In the last column of Table 2.2., the relative sizes of the corpora are presented as percentages calculated on analysable words.

---

[27] The Perseus *Guidelines for the Syntactic Annotation of Latin Treebanks* require every complete sentence, i.e. annotation unit, to have one and only one predicate that is attached to the root of the sentence with the relation PRED. See, Bamman & al. 2007b, 4.

2.1.2. Temporal and geographical distribution

The LLCT corpus is intended to be chronologically representative although this study does not aim to conduct diachronic research.[28] However, the uneven supply of the material as well as the availability of the editions renders the composition of the corpus less systematic than would have been desired. The temporal backline was set at the 860's because, at the time of composing the corpus, no facsimiles of charters were available after 865, end year of ChLA 81. In order to enable chronological comparison of complete decades, the last years of the 860's were provided with sufficient material by using charters that needed to be checked against the originals in Archivio Storico Diocesano, Lucca.

Charters were selected from the whole time range of nearly 160 years as evenly as possible, except for the CDL charters that are all included because of the superiority of their edition (Schiaparelli) in comparison to CDT and MED. This results in overrepresentation of the charters of the 760's and 770's, as can be seen in Figure 2.1. At the same time, the first 40 to 50 years of the corpus are underrepresented due to lack of material. The imperfect chronological balance of the corpus, a problem commonly encountered in historical linguistics, is alleviated by the fact that studying diachronic variation is not included in the objectives of this study. As the last 11 decades (760–869) offer enough material for a balanced comparison, diachronic research can be plausibly conducted concerning the decades in question or by comparing longer subperiods, such as 710–789 and 790–869, to each other. The decades between 780 and 869 consist of 14,000–15,000 words each.

Figure 2.1. Number of charters per decade.

---
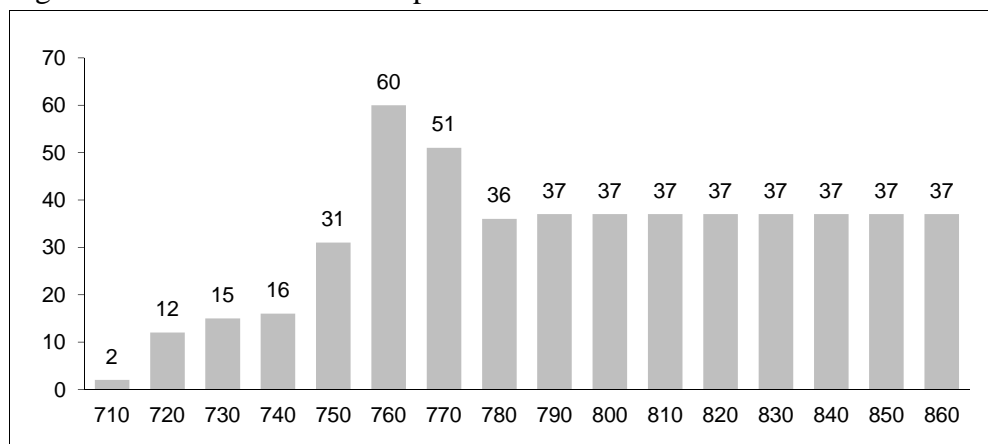
[28] Diachronic examination of case marking is one of the most important future perspectives opened up by this study.

Whenever possible, the charters from each subperiod were chosen evenly from the whole period of ten years in order to minimise the possibility of arriving by chance at interrelated charters written by the same scribe in the same circumstances. In general, the dominating influence of a single scribe (such as Adalfridi, writer of dozens of charters in the 840's to 860's) was avoided when other parameters allowed. In the case of doublet charters, only one of the originals was accepted. Severely fragmented charters were left out, and slightly fragmentary ones were included only if required by the scarcity of material in that specific subperiod.

The aim of the LLCT corpus is to make it possible to study variation and change in Tuscan Latin, so the diatopic variable is studied only inside Tuscany. All the charters come from the historical region of Tuscia, which comprises modern Tuscany plus the northern parts of Lazio and the western parts of Umbria. This areal framing is motivated on dialectological grounds: roughly put, the main isoglosses, i.e. modern dialect boundaries, between the northern and southern Romance varieties, seem to demarcate Tuscia as a 'no-man's land'.[29] The overwhelming majority of Tuscan charters come from Lucca, and it would have been possible to concentrate only on them. I found it, however, important to balance the corpus geographically by incorporating as many CDT charters as possible from outside Lucca in order not to allow the relative uniformity of the Luccan scribal tradition[30] to obscure the degree of variation necessarily caused by differing local documentary conventions even within the clearly limited area of Tuscia.

Unfortunately, all 39 non-Luccan CDT charters (only 16,342 words) date back to before 813. After 813, the areal representativeness is lost because in MED there are available charters only from Lucca and its neighbourhood. All the other editions (Kurze 1974, ChLA[2]) of the 9th century charters from the archives of Siena, Pisa, and Firenze are under copyright. This, along with the presumable aspirations for standardisation in the 9th century, is perceived in the latter half of the corpus as a visible reduction of linguistic variation. Nonetheless, the comparatively large total of 198,696 words is thought to compensate for many an irregularity in the composition of the corpus.

Most of the charters (73%) of the corpus have been written in Lucca or its sphere of influence (see Table 2.3.). Presumably, only charters important to the Luccan episcopal see were deposited in its archives, nowadays Archivio Storico Diocesano di Lucca. If the charters are grouped on the basis of

---

[29] Maiden & Parry 1997, 225–232, 297–302; Devoto & Giacomelli 1971, 65–66. For a geographical definition of the historical Tuscia, see Conti 1973, 108–109.
[30] On handwriting, see Petrucci & Romeo 1992, 82–88.

their archival origin, the prominence of Lucca is even more striking: 87% (see Table 2.4.). The disparity between the two percentages reveals that although only 73% of the charters are indicated to have been written in Lucca or in the neighbouring villages under the Luccan jurisdiction, a much larger number of charters has something to do with the Luccan church. In these cases, the scribe could have been sent from Lucca even though he composed the charter in a far-off village or town. The same applies, of course, to other scribal centres of Tuscany, such as Pisa. In general, the writing place indicated in a charter does not necessarily mean, not even on a theoretical level, that the language of that charter might display special local characteristics. Therefore, I have contented myself with a rather rough geographical presentation of the material. In Table 2.3., only locations occurring more than twice are listed.

Table 2.3. Writing places mentioned in charters.

Place	N	%	Place	N	%
Lucca and territory	380	73	Grosseto	3	0.5
Pisa and territory	20	4.0	Monte Amiata	3	0.5
Chiusi	15	3.0	Montepulciano	3	0.5
Val di Cornia/Cornino	8	1.5	Paterno (maiore)	3	0.5
Sovana	6	1.0	other locations	61	12
Luni/Lunata	4	1.0	unknown locations	6	1.0
San Regolo di Gualdo	4	1.0	∑	519	100
Castronovo	3	0.5			

Table 2.4. Current repositories.

Repository	CDL	CDT	MED	∑	%
Archivio Storico Diocesano, Lucca	152	1	300	453	87
Archivio di Stato, Siena	17	33	0	50	10
Archivio di Stato, Pisa	11	4	0	15	3
Archivio di Stato, Firenze	0	1	0	1	0.2
∑	180	39	300	519	100

Further study on diatopic variation in the charters or on the local scribal traditions would require an in-depth topographical and historical research, which is beyond the scope of this study.

2.1.3. Fragmentary charters and charters surviving as copies

Using charters in linguistic study requires that the texts are in a sufficiently good physical state of conservation. There is no doubt that the majority of the medieval charters have been lost over the

centuries. The parchment has been destroyed by fire, humidity, or attrition, or the charters have been discarded as useless and out-dated. Those charters that survive also sometimes suffer from the damages of the these factors. In some charters, parts of the text are illegible because the ink has faded due to humidity. Originally, the charters were usually stored folded in a letter format and, therefore, the folds of the letters were the first to wear out. This results in damage all along the sharp edges. Those charters that were wrapped in rolls display attrition especially at the ends of the roll.

The physical damage directly corresponds to the fragmentation of the texts and limits the usability of several charters. Only one missing word or a word with a missing case ending often makes a sentence unclear or at least invalid for use in morphosyntactic analysis. If the fragmentary parts are expansive, the syntactic cohesion of the sentences is completely lost and only unconnected syntagms can be safely studied. Therefore, damaged charters are skipped in this corpus whenever possible. When they have been included due to a shortage of material, the fragmentary parts and words understandably cannot contribute to the linguistic analysis.

Fragmented words that can be restored reliably form a special case. It is often possible to restore fragmented words and, more roughly, even longer passages, as the charters follow certain formulae. In CDL, the editor, Schiaparelli, restores fragmentary words when possible while Barsocchini (MED) sometimes ends fragmentary words in three or five dots and Brunetti (CDT) only uses dots. Following Schiaparelli's example, I have sometimes in CDT and MED restored the reliably restorable fragmentary words but, to avoid risks, excluded them from morphological analysis by a special annotation in case the words do not retain their inflectional endings (see section 2.2.1.). Even though these restored words do not take part in the morphological analysis, they can function as heads and dependents in the syntactic dependency structure of the treebank. Those fragmentary but restorable words, on the other hand, whose inflectional endings are safe are included both in the morphological and syntactic analyses. This method is justified by the fact that syntactic relations, unlike morphological endings, can be deduced rather easily if the fragmentary part is small enough and, especially, when the passage follows a certain formula.

Most of the charters are original documents in the strictest sense of the word. In the text editions, however, copies (*exemplaria*) are also published. Copies can be divided 1) into those written immediately or soon after the redaction of the original and 2) into those written much later, for example, copies ingested in the cartularies of monasteries and compiled either in the Middle Ages or after the

birth of medieval antiquarian interest in Early Modern times. Only copies of the first type are accepted into LLCT. The few CDT and MED copies are all practically coeval while, in CDL, the copies written in the 8th and 9th centuries are accepted as well. Because the major part of the oldest charters survive only as copies, the number of copies is at its largest in CDL (25 charters). In CDT and MED, only sporadic copies (13 charters) were accepted in decades in which there was a shortage of documents.

All 38 accepted copies are located inside the corpus according to the date of their originals that is available in the copied protocol even if the date of their copying is known, which is the case in some of them (e.g. CDL 22, 113–115, 120). The legitimacy of ingesting copies is based on two facts: first, most of the dated copies are almost coeval with the original charters, and this may be generalised to the undated copies as well; second, in the cases where both the original and the copy survive (e.g. the text of CDL App. from AD 803, copied almost verbatim in MED 385, a court record from AD 813), the scribe has strictly adhered to the principle usually expressed at the end of a copy: *ex autentico fideliter exemplavi nec plus addedi nec minime scripsi*, "I copied faithfully from the original, nor did I add or omit anything" (CDL 28).[31] The differences are insignificant and, at most, graphic by nature. Thus, no damage is caused to the diachronic reliability of the corpus when treating the copies as originals. The copies are likely to reflect the linguistic level prevailing at the redaction date of the original charter even if the time between the redaction and the copying was long.

2.2. Textual annotation

In this study, the term 'textual annotation' covers all the annotation types apart from linguistic annotation and the archival-contextual metadata attributed to each charter (writing place, scribe, date, document type). Textual annotation describes selected diplomatic information applied to the corpus text. This textual annotation is performed through the TEI XML markup standard. Diplomatic information concerning abbreviations, lacunae, and diplomatic parts of the charters affects directly the realisation and use of the linguistic annotation. Problems concerning textual annotation are discussed in the following two chapters.[32] For linguistic annotation conventions, see section 2.3.

---

[31] According to Larson (2000, 160), Osprandus, a Luccan scribe who wrote quite correct Latin in his own charters, did not correct the non-standard language of the charters he copied, such as CDL 48 and 67. See also ChLA 702, a charter from Salerno whose draft survives as a dorsal *dictum* (Sanga & Baggio 1995, 250–251). These cases suggest that the scribes copied texts with no conscious orthographic or morphological revision in mind.

[32] An overview on the textual annotation of LLCT is found in Korkiakangas & Lassila 2013.

2.2.1. TEI XML edition principles, abbreviations, and lacunae

"The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics."[33]

The TEI Guidelines are a technical standard of presenting mainly primary source materials for online research in XML format. The textual components and concepts are encoded by way of a markup language consisting of tags and attributes specified for one or more academic disciplines. The TEI markup language defines about 500 structural (e.g. paragraph, sentence, word) and conceptual (e.g. person, place name, author) textual features, only a fraction of which is applied to this corpus.[34] My treebank is compatible with the Perseus Latin Dependency Treebank (LDT) that employs the P4 release of the TEI Guidelines.

The granularity of the structural annotation affects directly the usability of the corpus: the more information is encoded, the more variables can be included in the queries and, as a consequence, the more complex phenomena can be perceived. First, the text is provided with sufficient archival-contextual metadata in due form, i.e. in tags or annotations. These are supplied in a form of a header for each charter. In my treebank, the following seven parameters are specified in the headers:

1) document number and page number of the original edition
2) date (as indicated in the *datatio* of the original document)
3) place of writing (as indicated in the eschatocol)
4) name of the scribe plus his possible title (e.g. *notarius* or *subdiaconus*)
5) conservation status (original or copy) plus the date of the copy if known
6) document type (*donatio*, *charta dotis*, *iudicatum*, etc. as indicated in the document)
7) in CDT and MED, also the short editorial *regestae* are included

No structural inline annotation is added to encode line breaks, sentences, or paragraphs. The sentences are automatically split for linguistic annotation by full stops in the annotation environment. Extra-linguistic abbreviations, such as Roman numbers and '+' (*signum crucis*), are tagged as *abbr* in order to exclude them from linguistic analysis. In addition to these, a subtype of inline annotation

---

[33] Text Encoding Initiative, http://www.tei-c.org/index.xml (accessed 26 March 2015).
[34] Text Encoding Initiative, http://www.tei-c.org/Guidelines/P5/ (accessed 26 March 2015).

is introduced to encode expanded abbreviations and fragmentary words and sequences of words, i.e. lacunae. Both categories are extremely important because they are present in the plain TEI XML text, but cannot be included in morphological analysis, except for in certain cases.

Most of the abbreviated words of the original texts can be expanded with a considerable degree of reliability and this is, indeed, what was done when preparing the text edition for the LLCT corpus. The traditional method, i.e. round brackets, cannot be utilised to indicate the expansion, e.g. *not*(*arius*) because this would result in problems with morphological annotation. Therefore, it was decided to annotate with *expan* tag those words where the inflectional ending is involved in the abbreviation and the expansion is, thus, not completely reliable. For example, the abbreviation *not* is expanded in the TEI XML text *notarius* or *notario* etc., depending on the context (*Adaudatus not > Adaudatus notarius*; cf. *ad resedend > ad resedendum* etc.), but as the interpretation of the case ending is unreliable, the whole word *notarius* is excluded from the morphological analysis by the *expan* tag.

In spite of their exclusion from the morphological analysis of LLCT, the expanded abbreviations of this type are usually fully acknowledged in the lemmatic and syntactic analyses, provided that there is no possibility of mistaking the lemma and/or syntactic function.[35] For example, in the scope of the charters, the abbreviation *not* cannot be but an instance of a word form derived from the lemma *notarius*, and *resedend* comes surely from *residere* 'to reside'. In the major part of cases, there is no doubt about the syntactic function and the head/dependent relation of the expanded words: in *Adaudatus not*, *notarius* is clearly an attribute of *Adaudatus*, as well as, in *ad resedend*, *resedendum* is the complement of the preposition *ad*.

It is justifiable to expand abbreviated words of this kind in the TEI XML text because the text is not intended to be a normal diplomatic edition of the charters but, instead, serves as input for the linguistic annotation that was conducted on it in the next stage of the project (see section 2.3.). Whoever wants to use the TEI XML text of LLCT for diplomatic research, must be aware of its unreliability regarding the abbreviations, as the information about which particular letters the abbreviation contains is not encoded.

---

[35] I print the Leyden convention brackets only in the example sentences of chapter 2. Elsewhere, they are omitted because I utilise only example sentences where the key words are all diplomatically reliable.

Certain established morphological abbreviation types form a special case even though in the 8th and 9th centuries they were not as common as later in the Middle Ages. I consider the word-ending *-us*, *-ur* and *-Vm* reliable when they are marked with their customary abbreviation marks: words with these abbreviations can be safely included in the morphological analysis of LLCT. Their origin as abbreviations does not show in the TEI XML because they are not annotated with the *expan* tag. Even the nouns with a nasal stroke indicating *-m* over the last letter of the stem are considered relia-ble because the theme vowel is known if the lemma is known: e.g. *act > actum*.[36] This applies to several other cases as well: for example, to *ad resedend*, where a (neuter) gerund *resedendum* is expected.[37]

As for lacunae, all the words with an illegible inflectional ending are, of course, excluded from morphological analysis by tagging even though the beginning of the word would be completely preserved and the ending deducible with good probability. As was stated in section 2.1.3., some fragmented words can be restored reliably enough because they are part of well-known formulae. This is the reason why they can serve as elements in the syntactic dependency structures of the sen-tences. If a fragmentary word that preserves its inflectional ending is unquestionably restorable, the word is restored and included in both the morphological and syntactic analyses. This procedure of including only partly restorable fragmented words into dependency structures is adopted in order to ensure the maximal usability of the dependency network of the syntactic trees even in the fragmen-tarily preserved charters. Longer fragmentary passages cannot be restored, whether they are formu-laic or not.[38]

2.2.2. Encoding diplomatic parts

Along with the annotation described above, another type of diplomatic inline markup is introduced. This annotation indicates the diplomatic division of the charter by marking the non-formulaic parts of the document by specific *seg type="free"* tags. With this procedure, it is possible to analyse the language of the non-formulaic and formulaic parts separately, as described in section 1.2.3. For full diplomatic segmentation of a *charta venditionis*, see Appendix 1.1.

---

[36] It is true that scribes may sometimes have failed to write nasal strokes and other visually scant abbreviation marks merely by oversight. This, of course, affects the distributions of certain inflectional endings, such as *-am* and *-em*.
[37] Korkiakangas & Lassila 2013, 64–65.
[38] Korkiakangas & Lassila 2013, 66.

Sabatini in particular presupposes that the non-formulaic and formulaic parts can be distinguished rather easily. At the same time, both Sabatini and Larson intimate that the non-formulaic parts are the only ones worth studying and that they can be found exclusively in the *sedi di licenza* listed by Sabatini or in the dorsal notes written on the back of the charters.[39] I find this rather dichotomous stance far from unproblematic. Even the formulaic parts can tell a great deal about linguistic change and variation indirectly. Second, the only really 'free' parts of documents seem to be the lists of products or names, found inside the dispositive part or in the dorsal notes, but they are seldom of syntactic interest. Moreover, almost all written expressions seem to be based, consciously or unconsciously, on earlier written models while, at the same time, some seemingly formulaic expressions may well reflect certain phenomena of spoken language. Therefore, too sharp a separation between formulaic and non-formulaic material does not seem linguistically useful.

Even inside the *dispositio* (section 1.2.3.), for example in a *charta concambiationis* of (1), several formulaic expressions can be identified: preamble clauses (*modo vero* [--], *et recepi ego* [--]), dispositive verbs (*stetit atque convenit* [--] *dares* [--] *dedi*), confirmation clauses (*ut sint inter nos* [--]) and various formal insertions (*iam nominatae*, *in suprascripto*, *predictu*, *hoc est*).[40] In spite of these, I annotated this *dispositio* throughout as 'free' because the formulaic phrases (underlined) remain in the minority.

(1) CDT 28 (AD 787) *Modo viro boni animis inter nos stetit adque convenit, ut tu s(upra)s(crip)tu Tao dare novis Ansperto abbati sorte vestra de terra et silva in ipso Fauclanu in locu, que dicitur monte Audualdo, et illa cetina da illi noccli et ipsa petia de pratu maiure. Ita et dedit ego q(ui) s(upra) Tao cl(ericus) iste iam nominate pezze de terra in s(upra)s(crip)to locu monte Audualdo, fini via publica usque in orto da nucli, quem in ipso locu nihel novis reservabimus de parte mea. Et recipi ego Tao ad te pr(e)d(ic)tu Ansip(er)tu abb(a)te pro s(upra)s(crip)ta terra mea, hoc est terra vestra, campu et silva in campu Gaufredi et pratu cum saudo, hubi iam ante os dies nos s(upra)s(crip)to Tao casa abuemus vel genitor meus Posso, ud sint inter nos omni tempore firma et stavile commutatione ambavus partib(us).*

[39] Sabatini 1965, 975–976; Larson 2000, 152–153. See also Larson 2012, 65–66.
[40] Cf. Sanga & Baggio 1995, 250–251, for the insertions made by the scribe when preparing the definitive charter text on the basis of a draft.

The *notitiae iudicati* present, presumably, the clearest cases of non-formulaic language. For example, an accused, who ends up confessing his delict, quotes his very own treacherous words in (2). Being an aggravating confession, the utterance is likely to have been recorded verbatim. This is the motive for which meticulous verbalisations of actually spoken words can be expected to figure in the *notitiae iudicati*. A caveat, however: not everything that occurs in the *iudicata* can be deemed non-formulaic. For example, the clause of (3) is found, *mutatis mutandis*, in many a *notitia* despite its non-formulaic appearance.

(2) CDL App. (AD 786) *uade, si potest, tolle cartula*(m) *illa*(m) *quam ego feci Deusdedi pr*(es)*b*(iter)*i de eccl*(esia) *mea s*(an)*c*(t)*i Angeli et res ad ea*(m) *p*(er)*tinente, et dili* [= *dele*] *ea*
"go there if you can and fetch the charter that I made for the good of Deusdedi, the priest, concerning my church of Saint Archangel and the property therein involved, and destroy it [= the charter]"

(3) CDL App. (AD 786) *d*(um) *testimonia et notario ipsum* [= *notarius ipse*] *p*(er) *euangelia testimoniu*(m) *suu*(m) *confirmauerunt et ipse Deusdedit pr*(es)*b*(ite)*r iurauit sicut supra iudicauim*(us)
"when the witnesses and the notary confirmed their testimony by the Gospel book and Deusdedit, the priest, witnessed in the same manner as we judged previously"

Contrary to the 'free' parts, formulaic parts are rather straightforward to define. For example, invocation and date clause, such as in (4), and all the elements of the eschatocol (*rogatio*, *actum* clause, subscriptions, *traditio*) are clearly formulaic.[41] Therefore, it seems best to distinguish the clearly formulaic parts from the less formulaic 'free' parts by an approximate annotation. In a corpus of c.200,000 words, even a rough binary segmentation of this type is likely to bring forward the contrast between the formulaic and non-formulaic text types where it really exists. A more complex diplomatic segmentation would be more revealing, but hardly utilitarian in regard to the time input it would require.

---

[41] For all these diplomatic parts, see Bresslau 1958, 46–48; Pratesi 1979, 67–79; Guyotjeannin & al. 1993, 72–85. Formulaic parts usually contain more abbreviations, which can be seen e.g. in the *invocatio* and *datatio* of Appendix 1.1. In certain cases, the number of abbreviations can be exploited in defining which sentences are formulaic and which not.

(4) CDL 169 (AD 763) *in nomine Domini Dei salvatoris nostri Iesu Christi*, *regnante d(om)n(o)*
nostro Desiderio rege, *anno regni eius sexto*, *et filio eius idem d(om)n(o) nostro Adelchis rege*,
anno regni eius quarto, pridie nonas febr(uarias), *per ind(ictione) prima*
"in the name of Lord God, our Saviour, Jesus Christ, in the sixth year of reign of our lord
Desiderius and in the fourth year of reign of his son, our lord, king Adelchis, day before the
Nones of February, under the first indiction"

To sum up, the sentences of those nuclear parts of the *dispositio* that are clearly not formulae are
tagged as 'free' although they may contain several isolated formulaic items (e.g. *suprascripto* or
predictu). From outside the *dispositio*, only sporadic non-formulaic sentences can be found. These
are belated insertions to the *dispositio* (5), specifications (6), or other atypical expressions, such as
the *scripsi* clause of the Senese charters (7).

(5) CDL 16 (AD 714) *et post hanc conpleta cartula rememorauimus particellula nostra de*
oliueto in Uaccule
"and after completing this charter, we remembered our share of an olive grove in Vaccule"

(6) CDT 43 (AD 796) *i sunt Waltari cl(ericus) et Alpari barbanis meis et Walari parentes meos*
"they are Waltari, the priest, Alpari, my uncle, and Walari, all relatives of mine"

(7) CDT 44 (AD 796) *scripsi ego Ursus pr(esbiter) et not(arius) rogatus et petitus ab Rumanu*
filio quondam Rumaldo v(iro) h(onesto) et vinditores
"I, Ursus, presbyter and notary, wrote [this] upon request and appeal of Rumanu, son of the late
Rumoaldo, *vir honestus*, the seller"

Example (8) is from a *dispositio* which contains only a few non-formulaic words. In the annotation,
it is considered thoroughly formulaic even though the first lines of the *dispositio* could be annotated
as 'free' because of the name and place variables. This, however, is not done in order not to split
sentences, which would complicate the dependency analysis. The formulaic phrases are again un-
derlined.

(8) CDT 71 (AD 806) <u>*Consta nos*</u> *Amato et Susinnu et Santulu g(ermanis) filiis quondam*
Fausto de bico Spiniu Caprinu finibus Suanense <u>*ac die bindedisse et bindedi tibi*</u> *vir*
beatissimo Sabbatino abb(ati) rector monasterio sancti Salbatori sito monte Amiate, <u>*id est*</u>

omnia sortem nostra, quem abere et tenere bisi sumus, id est casa, corte, ortis, bineis, prati, silbis, cetinis, pascuis, aquis aquarumque ductibus, cultum vel incultum, omnia et in omnibus, quantum ad ipsa suprascripta sortem nostra pertenuit. Homnia in integro bindedimus tibi suprascripto Sabbatino abb(ati) *vel ad posteris vestris potestate possedend*(i) *acceptoque ad te, quod inter nobis bonam et spontanea bolumtate conbenit, pretium oc est sol*(idos) *sexsagenta, in finitum et deliberatu susceptum pretium et ad presentis absoluto*.

The *pertinentia* clause (*id est casa*, *corte*, *ortis* [--]) is completely formulaic as well as the affirmation *homnia in integro bindedimus tibi suprascripto Sabbatino abb*(ati) *vel ad posteris vestris potestate possedend*(i). In the *pretium* clause (*acceptoque ad te* [--] *pretium oc est sol*(idos) *sexsagenta*, *in finitum et deliberatu susceptum pretium et ad presentis absolute*), the only non-formulaic word is *sexaginta*.

The subscriptions are also formulaic, whether they follow the wording dictated by the scribe or are copied from the preceding subscription line – or were all written by the scribe himself. The last option was used when the subscribers were illiterate or did not want to write for some reason. In that case, the scribe composed the subscription and the subscriber signed merely a small cross (*signum manus* or *signum sanctae crucis*), e.g. (9). In the Italian private documents of the 8th and 9th centuries, the subscription was not considered as important an authentication method as in earlier or later periods.[42]

(9) CDL 169 (AD 763) *signum + manus Teuderisci presbiteri*
"sign + of the hand of Teuderiscus, the priest"

(10) CDL 164 (AD 762) + *ego Fratellus pr*(es)*b*(iter) *rogatus a Gauspert rector ecl*(esie) *sancti Fridiani in hanc pagina me testis suscripsi*
"+ I, Fratellus, the priest, upon request of Gauspert, rector of the church of Saint Frigidianus, signed this document as witness"

When the subscribers did subscribe personally, they wrote a lengthy formula, as in (10). These autograph subscriptions form a special case from the point of view of linguistic analysis, as they rep-

---

[42] Pratesi 1979, 53–54; Bresslau & Klewitz 1958, 176–177.

resent linguistic conditions different from the language used by the scribe. The scribes were expert writers but the subscribers necessarily not. It is true, however, that scribes also appear as subscribers. Nevertheless, it may be assumed that autograph subscriptions provide more linguistic variation in inflection and orthography than the parts written by the scribe.

2.3. Linguistic annotation

The linguistic annotation of LLCT will be discussed at some length in the following three sections (section 2.3.1. to section 2.3.3.). The depth and coverage of the linguistic annotation define the preconditions that this corpus-linguistic study is founded on. Indeed, to be able to pose sensible questions to the treebank, one must know what questions the annotation can answer. In corpus linguistics, a considerable part of linguistic research has already been done before the queries begin in terms of assigning to the words their linguistic analyses.

2.3.1. LDT annotation schema

Technically, a treebank is a collection of texts supplied with linguistic (and textual) metadata in the form of machine-readable markup or annotations. Each word is associated with a tag containing its linguistic analysis. In theory, the associated categories can be infinite. Provided that not only the lemma and morphology, but also the syntax of the sentences is annotated, it is customary to use the terms 'parsed corpus' or 'treebank'. In a treebank, the syntactic structure of each sentence is usually represented by a tree structure.[43] The tree structure of one LLCT sentence can be viewed in Appendix 2.2.

I chose for LLCT the annotation schema proposed by the *Guidelines for the Syntactic Annotation of Latin Treebanks* (version 1.3, 2007). The *Guidelines* were launched by the Latin Dependency Treebank (LDT) and by the Index Thomisticus Treebank (IT-TB).[44] The LDT standard is based on dependency grammar, successfully applied to the analytical layer of annotation in the pioneering Pra-

---

[43] Hundt 2008, 169–170; Nivre 2008, 225–226.
[44] Bamman & al. 2007a; Bamman & al. 2007b. The Latin and Ancient Greek Dependency Treebanks are a joint project aimed at treebanking texts in Classical Latin and Greek. The project is hosted by the Perseus Digital Library Project at Tufts University in Boston, USA (http://nlp.perseus.tufts.edu/syntax/treebank/index.html). The Index Thomisticus Treebank is an ongoing project aimed at the syntactic annotation of the Index Thomisticus, a morphologically annotated corpus of the texts of St. Thomas Aquinas. The project is hosted by the Catholic University of the Sacred Heart in Milan, Italy (http://itreebank.marginalia.it).

gue Dependency Treebank (PDT) of Czech language.[45] Dependency grammar is particularly suitable for morphologically complex languages with a relatively free word order, such as Czech, Latin, and Greek, where the linear order of constituents is broken up with elements of other constituents. The LDT *Guidelines* have been optimised for Ancient Greek and Latin also in other respects on the basis of *Latin Syntax and Semantics* of Harm Pinkster (1990).[46] Moreover, the LDT offers a practical online annotation environment in collaboration with the Alpheios Project.[47] LLCT differs from standard Latin corpora in several aspects, but first and foremost because of its non-standard language. The basic LDT annotation schema was therefore complemented with additions tailored for non-standard Latin (see section 2.3.2. and section 2.3.3.).

The dependency relation views the finite verb as the structural centre of clause structure. All other syntactic units are either directly or indirectly dependent on the verb. Dependency grammar differs from constituent-based grammars by abandoning non-terminal phrasal categories and instead linking words themselves to their immediate head. Dependency grammar considers the syntactic functions (e.g. subject, object, attribute) to be primitive and derives the constellation of the sentence from these functions, contrary to the constituency grammars, which derive the functions from the constellation. Dependency grammar is appropriate for Latin since it is theoretically close to traditional Latin grammars, where the highly inflected nature of the language is often approached by way of defining the elements that modify other elements, i.e. their heads.[48]

In Figure 2.2., an instance of the common salutation formula (11) is provided with a dependency analysis. The arrows are directed from heads to dependents, the head of the predicate (PRED) being outside the sentence (the so-called external head). The dependents or children of the predicate *d(ixit)* are the subject (SBJ), *Gaiprand*; the object (OBJ), *sal(utem)*, and the indirect object (IOBJ) *tibi*. These dependents can function as heads to their dependents: *Gaiprand* is the head of the attribute *u(ir)*, which, for its part, is the head of the attribute *d(euotus)*.

(11) CDL 117 (AD 755) *Gaiprand u(ir) d(euotus) tibi eccl(esie) beati s(an)c(t)i Frigidiani loco Griciano p(er)p(etuam) sal(utem) d(ixit)*

---

[45] For Prague Dependency Treebank, see Hajič & al. 1999; Hajič 1998; Sgall & al. 1986.
[46] Bamman & al. 2007b, 3.
[47] Since 2014, the Arethusa Annotation Framework combines morphology and dependency relation tools into one user-friendly interface (http://www.perseids.org/tools/arethusa/app/#/). The projects involved in Arethusa are the Perseus Digital Library Project (especially Perseids Project) at Tufts University and the Open Philology Project at the University of Leipzig.
[48] Bamman & al. 2007b, 3; van Valin 2001, 102–107; Mel'čuk 1988, 3, 14–24.

"Gaiprand, *vir devotus*, expressed his perpetual salutation to you, church of Saint Frigidianus in Griciano"

Figure 2.2. A dependency grammar representation of sentence (11).
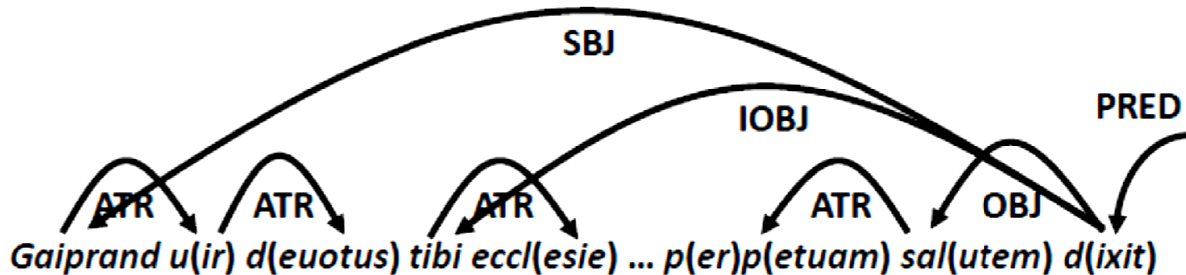
Most treebanks follow annotation schemas based on constituency grammars albeit dependency solutions are getting increasingly common, especially with languages with free word order. The dependency structure can be added either on an already existing layer of constituent structure or directly on top of the morphological annotation, as is the case in PDT and LDT.[49]

In the annotation environment, the lemmatic and morphological analyses are entered in an annotation editor where each word form is given an appropriate lemma and a nine-place morphological analysis (*pos*, i.e. part-of-speech tag) encoding the part of speech proper, person, number, tense, mood, voice, gender, case, and degree. The syntactic annotation comprises syntactic tags (e.g. PRED, SBJ, OBJ, ATR, ADV) and head-dependent relations that are defined in the tree viewer.[50] Both the morphological and syntactic annotations are managed through a graphical user interface, and the analysis data for both these annotations are saved as a single *word* element for each word in the TB.XML files. For example:

<word   id="12"   cid="43560442"   form="patribus"   lemma="pater1"   pos="n-p---md-" head="11" relation="OBJ" declension="3" animacy="0" iobj="1"/>[51]

Here, the 12th word of a sentence, *patribus*, is analysed to be a word form of the lemma *pater*, while the alphabetical code in the *pos* tag stands for 'noun plural masculine dative'. The values of the attributes 'head' and 'relation' imply that the word has a dependency relation 'object' to the preceding

---

[49] Nivre 2008, 231.

[50] Bamman & al. 2007b, 4. For the morphological tagset, see the README file for the Latin Dependency Treebank at http://nlp.perseus.tufts.edu/syntax/treebank/ldt/1.5/docs/README.txt.

[51] The XML markup language consists of elements wrapped in < >, such as *word*, that may contain attributes, such as *form* or *pos*. The values of the attributes are presented between apostrophes. I shall quote the element and attribute names in italics.

word which has the identifier number 11. Attribute 'declension' stands obviously for the declension category while the 'animacy' attribute contains even a light semantic analysis: value 0 means that the referent of the word *patribus* is animate and a proper name. The only semantic dichotomies that have been annotated are [± Animacy], [± Proper name], and [± Toponym]. So, in theory, the annotation is four-layered, but in practice, the lemmatic, morphological, syntactic, and semantic analyses are all included in the same *word* element. Finally, value 1 of the attribute 'iobj' shows that *patribus* is an 'indirect object'.

In LLCT, the linguistic annotation is related as stand-off markup to the TEI XML files that contain the edited source text provided with the textual and diplomatic inline annotations. In so doing, the linguistic markup that structures the source text constitutes an independent layer linked to the source text by specific virtual pointers (XPointers). Section 2.4. describes the technical procedure of aligning the linguistic and textual annotation layers that makes it possible to query LLCT with *ad hoc* search tools.[52]

The lemmatic and morphological annotation is performed through a semi-automatic procedure. The Morpheus Morphology Tool suggests for each word one or more lemmas and morphological analyses from the Perseus Dynamic Lexicon.[53] The user then chooses manually the correct lemma and analysis. Because several Latin lexemes have homonymic and homographic inflectional forms and because LLCT displays considerable divergences from the standard Latin orthography and, to a smaller extent, morphology, only 15–20% of the proposed lemmas and morphological analyses can be accepted as such. However, after choosing between the alternatives dispensed in drop-down menus, the percentage rises to 80–90%. The rest of the lemmas and morphological analyses must be supplied manually. Charter Latin also contains several technical words and names that do not figure in the Perseus lexicon. As a consequence, 2,724 novel lemmas (most of which are names) were introduced into the annotator-specific dynamic lexicon. The syntactic annotation was performed entirely by hand in the Alpheios Tree Viewer, as no parser could be used for the charters, due to their non-standard orthography and the often anomalous sentence structures.

---

[52] Korkiakangas & Lassila 2013.

[53] Bamman & Crane 2008. The Perseus Dynamic Lexicon is based on the existing morphosyntactic LDT treebank data. See http://nlp.perseus.tufts.edu/lexicon/.

2.3.2. Functional and formal analyses

As stated above, the annotation of the corpus is based on the LDT *Guidelines*. These *Guidelines* and the related programs supporting annotation are designed for standard Latin,[54] i.e. Latin that adheres to the Classical Latin grammar as far as morphology and syntax are concerned. The early medieval charters, however, differ from the standard in many respects concerning orthography, morphology, and syntax. In standard Latin, each syntactic function is usually encoded by a relevant case form, which makes the annotation process straightforward. With the Latin used in early medieval charters, the equivalence between form and function is often not transparent, and the existing standard Latin annotation guidelines do not manage non-standard forms or standard forms used in a non-standard way.[55]

To put it simply, it is impossible to provide charter Latin with uncontradictory annotation by applying as light an apparatus as with standard Latin. To solve this problem, several additions and modifications to the existing guidelines were introduced. The most important of these are the concepts of functional and formal analysis. I published these additions to the *Guidelines* together with Dr Marco Passarotti in the *Journal of Language Technology and Computational Linguistics* in 2011.[56] Even with the supplements, practical annotation requires highly subjective judgements on problematic cases, which is inevitable when dealing with charter texts and their language variety.

In order not to strain the annotation process, I decided that each word of the corpus can be assigned only one morphosyntactic analysis. In Korkiakangas and Passarotti's system, the analysis can be either functional or formal. To keep the treebank as queryable as possible, it would be best to label all the forms functionally, i.e. according to their (semantico-)syntactic function in standard Latin. This is not always possible, however, and some of the forms must be labelled formally.

If a word appears in its correct standard form, morphological tagging has no special relevance since form and function are matching. Functional analysis is applied when a form is language-evolutionarily deducible from the corresponding standard Latin form used in the same

---

[54] The use of term 'standard' with Latin has often been questioned, but given that the written texts of the Late Antiquity still essentially follow the Classical Latin model, a substantial consensus about 'correct' or 'accepted' morphology and syntax seems to have prevailed.

[55] See Philippart de Foy 2012 about changes in the LASLA annotation procedures to face similar problems in a medieval hagiographic corpus. LASLA, i.e. the Laboratoire d'Analyse Statistique des Langues Anciennes, University of Liège, Belgium (http://www.cipl.ulg.ac.be/Lasla/), is another big project in the field. The annotation style of syntax by LASLA concerns subordination patterns only.

[56] Korkiakangas & Passarotti 2011, 105.

(semantico-)syntactic function. For example, the non-standard plural masculine accusative form *solidus* may be used in place of the standard form *solidos* 'gold coins' as a direct object. The form *solidus* receives a functional analysis "accusative". There are, however, several anomalies where no connection to the functionally required classical form exists and formal analysis must be applied. For example, a clear linguistic error, an alleged dative/ablative form *heredibus* (CDL 192) sometimes occurs in place of *heredes* 'heirs' as subject. The form *heredibus* receives a formal analysis: either "ablative" or "dative".[57]

A few further examples clarify the procedure and give insights into some of the recurrent challenges of the annotation: for example, the following non-standard forms that occur as direct objects (although their form is not the standard accusative form) are annotated functionally as accusatives because they are meant to stand for the standard accusative forms: *solido* (standard: *solidum* 'gold coin'), *terra* (standard: *terram* 'land'), *testis* (standard: *testes* 'witnesses'). To take a longer example, in the phrase *in tua cui supra emturi* [--] *potestatem* (standard: *in tua cuius supra emptoris potestate* "in the possession of you, the above purchaser") (CDL 23), the two words of the noun phrase *tua potestatem* 'your possession' are labelled functionally as singular ablatives dependent on the preposition *in* although *potestatem* is formally an accusative singular in standard Latin. Finally, in *auris soledus trentas* (standard: *auri solidos triginta* "thirty gold coins") (CDL 45), the seemingly dative/ablative plural form *auris* is functionally labelled as a genitive singular (standard: *auri* 'of gold') showing an additional -*s*.[58]

The clear linguistic errors are always tagged according to their formal appearance. For example, if the above-mentioned alleged dative/ablative plural form *heredibus* functions as a subject of a finite verb, the form cannot be tagged functionally as a nominative because it is not possible to interpret it as a descendant of the nominative form. Thus, *heredibus* is labelled according to its form, i.e. as dative/ablative plural. The form is an error probably due to the contamination between two or more formulae, a phenomenon common in medieval charters, or to the wrong interpretation of the normal abbreviation *hhd* for *heredes*. It is of no relevance whether *heredibus* is tagged as ablative or dative; what matters is that the analysis as a subject is language-evolutionarily impossible. When running queries on the corpus, the distinction between formal and functional labelling allows one to isolate

---

[57] Korkiakangas & Passarotti 2011, 105–108.
[58] Korkiakangas & Passarotti 2011, 107.

the formally analysed forms and count the percentages of language-evolutionarily possible and impossible forms.[59]

Deciding between functional and formal analyses is seldom trivial. For example, the common subscription phrase *propria manu mea* [--] *subscripsi* "I signed in my own hand" occurs in CDL 40 as *proprias manus meas* [--] *suscribsi*. Here, the final *s*'s may be interpreted as a hypercorrection or a simple miscomprehension enabled by the general loss of word-final consonants in spoken language. Should *proprias manus meas* still be labelled functionally as ablative or could it be thought of as an intentional, albeit functionally erroneous, accusative plural?

The underlying principle must be to respect the choices taken by the scribe as far as they are traceable. As this subscription has, indeed, been written personally by a witness, the accusative-like appearance may result from an idiosyncratic misunderstanding of the formula. Moreover, the 4th declension *manus* was commonly recognised as a word with a specific inflexion, which provoked special types of mistakes in its use.[60] The subscriber may have replicated here the final <*s*>, which might have been considered typical of *manus*, even in its two attributes. As this instance shows, the problems are delicate and so must be the solutions as well: the decisions must be case-specific, and, in cases where the context is of no help, uncertain words and expressions must be left unannotated.

In *proprias manus meas* of CDL 40, I decided to analyse the NP formally (accusative plural). The use of -*s* can be considered intentional because, first, the accusative plural is likely to be a reflection of an erroneous plural interpretation ("with my own hands") and, second, the subscriber in question does not seem to overuse final -*s* in his subscription (*presbites* is a somewhat common variant of *presbyter* 'priest'):

> (12) CDL 40 (AD 728) + *Ego Gaifred pr(es)b(ite)s rogatus a Radchis pr(es)b(iter)o in hanc cartula donationis facta in Uualtprand abb(…) proprias manus meas suscribsi*
> "+ [= sign of the Holy Cross] I, Gaifred, the priest, upon request of Radichis, the priest, signed in my own hand this donation made for the good of Waltprand, the abbot"

Distinguishing between functional and formal analyses is, of course, not the only possible method for annotating non-standard Latin. In principle, one could also provide both types of annotations

---

[59] Korkiakangas & Passarotti 2011, 107.
[60] See Korkiakangas 2010, 145; Löfstedt 1961, 117–118, on the 4th declension forms.

side by side, but this sort of labour-intensive twofold annotation would often be redundant, as it would reduplicate the same information in most cases. Another possible solution would be to provide functional analysis only, and then to refine the query results according to the endings, for instance, by sorting out all the subjects ending in -*ibus*. However, this solution would result in clearly erroneous analyses: for example, the form *heredibus* would be tagged as a nominative.

The purpose of Korkiakangas and Passarotti's system is to provide morphological analyses that reflect the real language-evolutionary origin of the forms, in order to make it possible both to exploit fully the morphological tagging and to detect easily the anomalous cases, i.e. the ones whose morphological tags are incompatible with their syntactic function tags.[61] In spite of all this, the complexity of the functional/formal annotation, together with the adjustments listed in section 2.4.3., reduces the scalability of the annotation, as much post-processing is required after querying. The same degree of efficiency that is possible with the annotated corpora of standard Latin is not reached.

This means that LLCT is not directly comparable with other Latin treebanks. A potential user without an intimate knowledge of the annotation principles must work hard to extract valid results from LLCT because the annotation a) differs from that of LDT, b) is in part based on a rather abstract model (functional/formal analysis), c) requires a profound understanding of the changes that took place in the phonology and morphosyntax of Late Latin (e.g. *heredibus* above), d) partly depends on the subjective judgements of the annotator on the case-specific level. On the other hand, the non-standard Latin of the charters does not seem to make easier approaches possible.

A crucial point in this method is the concept of standard Latin, in relation to which the terms 'functional' and 'formal' are defined. By standard Latin, I mean the type of written language used, among others, by the Christian authors of the Late Antiquity who were seen as models for every literary work throughout the Early Middle Ages. Their Latin was still essentially Classical Latin as codified in prescriptive grammars, as far as orthography, morphology, and, for the most part, also syntax are concerned. Several non-Classical features typical of the language of the Vulgate were accepted, however. For example, *quia* or *quod* clauses replaced the Classical accusative with infinitive structure with *verba sentiendi et dicendi*. It was the variant of Latin in which writers such as St. Benedict, Gregory the Great, and later Paul the Deacon and other Carolingian savants, wrote.

---

[61] Korkiakangas & Passarotti 2011, 107–108.

It is true that charters are not literary works and their language seems to represent a peculiar genre of its own.[62] As the characteristics of this genre remain obscure for the present and no explicit new linguistic standard of charter Latin can be perceived, the only possibility is to adhere to the standard of the contemporary literary Latin described above – each study of linguistic variation and change always requires some point of reference. I am aware that using the term 'standard' when speaking of Latin has often been questioned, but given that the written texts of Late Antiquity still essentially followed the Classical Latin models, a substantial consensus about 'correct' or 'accepted' morphology and syntax seems to have prevailed.[63]

Adhering more or less directly to Classical Latin morphology is also required by the LDT annotation environment, designed for standard Latin. Therefore, both formal and functional labels of LLCT are based on standard Latin grammar. For example, *Italia* in the clause *in Italia reversus est* "he returned to Italy" (MED 394) is annotated functionally as an accusative singular because, with verbs denoting motion, such as *reverti* 'to return', the preposition *in* would govern an accusative (*Italiam*) in Classical/standard Latin. It might have been possible to reconstruct a morphological model by departing from the Italian Romance language, as Fiorentino does,[64] but assigning to *cartula* a tag, for example, 'Romance ending', would not be of much use since it does not tell anything about that Romance form or about the linguistic change behind it. Moreover, the annotation would become technically more challenging and more prone to annotator-based mistakes. Although non-standard morphology causes problems, it is, instead, easy to manage non-standard phenomena concerning syntax, such as complement clauses with *quia* or *quod* or Romance-type compound perfects with the auxiliary verb *habeo*.

2.3.3. Further adjustments to LDT *Guidelines*

Together with the LDT *Guidelines*, the principles described in the previous section are the backbone of my annotation style. The present section introduces the most important additional specifications and individual rules designed in order to treat recurrent problematic structures and morpholog-

---

[62] Bartoli Langeli 2006, 26–28; cf. Capo 1990, 208.
[63] For more about the challenges of comparing charter Latin to standard Latin, see Larson 2000, especially 161–162; cf. Lazard 1993, 391. See also Auernheimer (2003, 49–50) on her decision to choose Alcuin as the linguistic standard for her study on the Carolingian reform.
[64] Fiorentino 1994, 26–30.

ical issues consistently. A full account of the adjustments is found in Korkiakangas and Passarotti (2011). The few additions concerning semantics were mentioned earlier in section 2.3.1.

LDT does not give the annotators directions on annotating morphology because, in the case of standard Latin, morphological annotation is considered unproblematic, as it most often is. In regard to charter Latin, it is, however, the morphology that poses the greatest challenges to annotation. Some of the following annotation principles are simply technical decisions between two or more equally possible alternatives. In general, some of the decisions show clearly how difficult it is to annotate a non-standard language variety, especially within an annotation apparatus designed for the standard variety. One of the advantages of the corpus method is that the statistical tendencies that arise from large text masses verify (or falsify) the sometimes audacious annotation choices that have to be made when annotating a treebank. Thus, the linguistic analysis proper of this study is not bound to the "technical" decisions that will be presented in this chapter.

Since I study the case system, I am mainly interested in noun phrases and prepositional phrases and their relation to the verb of the sentence. Therefore, punctuation marks and non-inflectional adverbs (e.g. *modo*, *congrue*), except for negatives (e.g. *non*, *numquam*, *minime*), have been left out of the annotation.[65] In general, all the syntactic relations of the sentences have been annotated even though, in several cases, this may result in redundance in regard to the eventual need of annotational coverage. In the following, I first treat lemmatisation and syntax briefly, then morphology in more detail.

Lemmatisation

Reducing lemmas. Almost all the words in the charters have two or more graphical variants. Likewise, one single morph may have several realisations. Therefore, particular attention is paid to lemmatising all the graphical variations of a word under one common lemma in order to avoid proliferation of lemmas in the Perseus Dynamic Lexicon database.[66] For instance, nouns facing gender change, such as the masculine nominative plural *saeculi* 'centuries', as well as adjectives facing declension change, such as the 2nd declension nominative singular *inanus* 'void', are lemmatised under the standard lemmas *saeculum* (neuter) and *inanis* (3rd declension), respectively.[67]

---

[65] Bamman & al. 2007b, 30–36.
[66] Bamman & Crane 2008, 11–13.
[67] Cf. Philippart de Foy 2012, 484–489.

Proper names. Several Germanic and Latin proper names exhibit much variation. For example, the form *Delmati* (subject) is lemmatised under *Dalmatius* and the forms *Guntifrido* and *Cuntefrid* under *Guntifridus*. However, it is sometimes difficult to establish the correct lemma, as no variant seems to be more justified or more frequent than any other. In the charters, there are also several unidentified place names. Although in some cases the lemma can be reconstructed on the grounds of the modern name of the place in question, those names that remain completely opaque are labelled as "unknown".[68]

Syntax

Indirect objects. The LDT *Guidelines* use the same label OBJ for both direct and indirect objects. Even though the solution is suitable for standard Latin, where indirect objects always occur in the dative or as a prepositional phrase, it cannot be applied to non-standard charter texts, which feature a high degree of formal variation. In the phrase *terra que offerui sancti Petri cum ipsa fossa* "the plot of land, which we donated to St. Peter, with the ditch" (CDL 125), the direct object is *que* (standard: *quam*) and the indirect object is *sancti Petri* (standard: *sancto Petro*). Although they are both labelled as OBJ, *sancti Petri* must be assigned an additional tag in order to make clear its status as an indirect object (see tag *iobj ="1"* in the sample XML line in section 2.3.1.).[69]

Required arguments. The LDT *Guidelines*' notion of an object includes a wide range of phrases. Actually, LDT does not assign the label OBJ only to direct and indirect objects, but to all the complements of a verb – even to those that cannot become subjects if the verb is passivised. LDT annotates the underlined complements in the following examples as OBJ: *pater gladio utitur* "the father uses the sword", *abundat Germania fluminibus* "Germany abounds with rivers", *contendunt Romani cum Germanis* "the Romans fight with the Germans".[70] This practice is not possible in this corpus, and all the above-mentioned adverbials are labelled as ADV. The reasons for this are as follows: First, given that almost all the cases can appear with almost all syntactic functions, expanding the use of OBJ to complements would cause confusion with direct and indirect objects. Second, the vacillation of the semantic frame of verbs, together with the ambiguous sentence structure stemming from lacunae, scribal errors, and contaminated or misinterpreted formulae, does not often al-

---

[68] Korkiakangas & Passarotti 2011, 108.
[69] Korkiakangas & Passarotti 2011, 109. It is possible that the genitive construction *sancti Petri* is a crystallised elliptic version of *ecclesia sancti Petri*.
[70] Bamman & al. 2007b, 13; for the terminology, see Pinkster 1990, 12–15, 25–27.

low one to decide whether a certain (prepositional) phrase is a required or optional argument. Third, a strict distinction between complements and adjuncts is largely irrelevant in my study. I am not concerned about the valency, but about subjects and direct objects that are always required arguments and, usually, the most easily recognisable ones in a sentence. As for prepositional phrases, the ones encoding indirect objects are labelled as OBJ while the adjunctive ones receive ADV.

Morphology

Subjects. The following annotation style only applies to the subjects of clauses whose verb occurs in finite form. Therefore, the subjects of accusative with infinitive constructions and ablative absolutes are not discussed here. In standard Latin, the latter are encoded with accusative and ablative, respectively, while the standard case of a subject headed by a finite verb is the nominative. Also, the transimpersonal constructions, such as *me taedet* "I am annoyed", are excluded.[71]

I decided to tag formally as accusatives the 2nd and 4th declension masculine singular subjects ending in *<-o -u -um>*, such as the 2nd declension form *Deo* (standard: *Deum* 'God'). This was based on the commonly accepted view that these singular forms are historically derived from the accusative, not from the nominative.[72] Instead, the neuter subjects ending in *<-o -u -um>*, such as *pretio* (standard: *pretium* 'price'), are tagged functionally as nominatives. In principle, the neuter subjects could equally well be tagged as accusatives because the standard nominative and accusative forms of the 2nd declension neuters are identical.[73]

The formal tagging also applies to the 3rd declension singular subjects. In the 3rd declension, the imparisyllabic subjects, i.e. those whose stem has an additional syllable in all cases except for the nominative (e.g. nominative *potes-tas* vs. accusative *potes-ta-tem* 'possession'), are particularly reliable indicators of case because the morphological distinction is beyond the reach of the phonological levelling. Instead, the parisyllabic 3rd declension singular subjects that end in *<-e -i -em>*, i.e. those whose stem has the same number of syllables in all cases (nominative *tes-tis* vs. accusative *tes-tem* 'witness'), are not that reliable because the word-final /m/ was no longer pronounced in Late Latin and the fate of the word-final /s/ is dubious, too. Of course, even without /-m/ and /-s/, there is

---

[71] See Rovai 2012b, 118–119.
[72] E.g. Väänänen 1981, 116–117. For a critical view, see e.g. Sornicola (2011, 35) who refers to D'Ovidio and Schuchardt. The origin of ending *-o* will be discussed in more detail in section 4.2.3., where the here adopted decision to annotate the above-mentioned 2nd and 4th declension subjects as accusatives will be revised on the basis of the LLCT data.
[73] Korkiakangas & Passarotti 2011, 110.

still the distinction between the final vowels.[74] However, even the parisyllabic subjects have been tagged formally on a par with the 2nd and 4th declension singulars in LLCT.[75]

As regards the 1st declension singular subjects, their endings *<-a -am>* are always tagged functionally as nominatives. This and the preceding principles are based on the following reason: when annotating 1st declension singular subjects, no distinction can be made between the language-evolutionary outcomes of the standard nominative and accusative forms, as the endings *<-a -am>* seem to be completely intermingled. However, in the 2nd, 3rd, and 4th declension subjects, the nominative and the accusative forms may still have differed from each other because the final /s/ or its residues sort out the nominative as distinct.

It is true that the opposition between the 2nd, 3rd, or 4th declension nominative singular and accusative forms seems to have been partly neutralised in the Latin of Tuscany in the 8th and 9th centuries because of the general weakening of the final /s/. However, the pilot study I carried out before starting the annotation revealed that case opposition was at least partly preserved in the Latin of the charters: some residues of the case inflection seemed to be organised according to semantically-based alignment. Therefore, the bicasual assumption (nominative vs. accusative) was utilised as the working hypothesis for annotating all the declensions.[76] I emphasise that, as I said at the beginning of this chapter, tagging a corpus always requires hard decisions. The decisions need not be faultless, but must be observed consistently throughout the tagging process. The eventual corpus analysis will reveal afterwards whether the chosen decisions were justified or not. This is then taken into consideration in the linguistic interpretation of the corpus data, which is, thus, independent of the technical annotation decisions.

On this occasion, it has to be mentioned that I have tagged the 1st declension plural subjects in *-as* formally as accusatives. This interpretation is based on the assumption that, even though *-as* may have represented nominative morphology in some earlier texts, at the time of LLCT it was likely to be aligned with other accusative endings. Therefore the subjects in *-as* are analysed as accusatives throughout this study.[77]

---

[74] E.g. Väänänen 1981, 66. For final *-m*, see section 2.5.1.; for final *-s*, section 3.3.
[75] Cf. Korkiakangas & Passarotti 2011, 110–111, where we proposed a functional tagging for the 3rd declension parisyllabic subjects.
[76] Other studies also support the preservation of a partial case opposition. See e.g. Zamboni 2000b, 233–235, 243–244.
[77] Galdi 2012, 148; see also 140–141 for a concise synopsis of research history.

Partitive structures. In late non-standard Latin, the partitive genitive seems to be sometimes replaced by the 'Romance' form, most likely derived from the standard accusative (see (13)). These forms are tagged formally as accusatives and linked to their head via ATR.[78]

(13) CDL 267 (AD 772) *s(upra)s(crip)ta terra* ambas petias [--] *tibi p(re)d(ic)ta D(e)i eclesia offerrere uideor*

"I manifestly donate to you, the above-mentioned church of God, [--] both pieces of the above-mentioned land"

Prepositions. As a general principle, the complements of prepositions governing the accusative in standard Latin are labelled as accusatives and the complements of prepositions governing the ablative are labelled as ablatives if the case endings can be claimed to represent the original accusative and ablative forms, respectively. This requires looking at the meanings of the prepositional phrases, for example, prepositions *in* and *super* that govern the accusative when expressing motion and the ablative when expressing state.[79]

Nominal attributes. Nominal attributes occur mainly in the titles of commissioners and addressees of legal transactions, as in (14). Several problems arise when the head-dependent relations in such noun phrases are labelled. As a rule, the highest ranking member in the following hierarchy is chosen as the head of the noun phrase: personal pronouns > proper names > other nouns referring to humans. Thus, the head of the above noun phrase is *ego*, under which *Autulu* is attached as an attribute. *Vir religiosus*, *clirico*, and *filio* are then linked to *Autulu* as attributes.[80]

(14) CDL 266 (AD 722) *ego Autulu uir religiosus clirico filio quondam Bonuald de uico Turrite*

"I Autulus, *vir religiosus*, clerk and son of the late Bonualdus from the village of Turrite"

Number and person in verbs. If it is not possible to determine the number of a verb, it is tagged according to formal appearance. This phenomenon mainly occurs with the 3rd person of verbs that express actions carried out by the performing party of the charter, as it is often ambiguous whether a

---

[78] Korkiakangas & Passarotti 2011, 111; Zielinski 1972, 63; for this *cas oblique*, see Väänänen 1981, 116. However, especially when the distance between the head and its genitive modifier is long, it may rather be a question of apposition: e.g. *offero D(e)o et tibi ecclesie D(e)i et beate s(an)c(t)e Marie D(e)i genetrix* terra mea, *que habere uisu sum in loco prope Tripontio, uno foscione* (CDL 125). To my mind, *uno foscione* should be held here as an accusative-form apposition of *terra mea* (cf. Zielinski 1972, 63).

[79] Korkiakangas & Passarotti 2011, 111.

[80] Korkiakangas & Passarotti 2011, 112; for the animacy hierarchy, see section 4.1.1.

seller or donator is acting alone or with his/her heirs. However, in phrases like (15), the singular verb *uiditor* (= *videtur*) may be due to impersonalisation of the passive/middle structure.[81] The relative pronouns have become practically indeclinable by the time of early medieval Latin.

(15) CDL 23 (AD 720) *petras que iniui esse uiditor*

"the stones that are [= is] seen there"

The person of the verb is usually tagged functionally because the person is normally easier to recognise than the number. The context may be helpful: for example, in CDL 28 (AD 728) *abbas* [--] *habeas* "the abbot [--] may have", the form *habeas* "you may have" is analysed as a 3rd person singular (standard: *habeat*). In more complex cases where more letters are involved in the dubious ending, such as *donatores* [--] *habeas* (standard: *habeant* "the donators [--] may have"), the annotator has to make delicate decisions which depend on the amount of graphical variation observed in the charter. Here, the intended reading could have been even *donatorem* [--] *habeat*, the subject *donator* being used in the accusative.[82]

The above-listed additions to the LDT *Guidelines* have been utilised as the mechanical basis of the annotation of LLCT. It is good to remember that although the tagging that is followed in a corpus plays an important role in utilising that corpus, interpreting the findings is not determined by the tagging assumptions as long as the researcher is aware of them: the findings of the present study will show that several of the linguistic assumptions underlying these additions need to be revised. Having said that, consistent tagging is necessary for any corpus study.

2.4. Aligning textual and linguistic annotation

It was stated in the above chapters that the linguistic and textual XML annotation layers have been connected to each other through XPointer syntax.[83] This method makes it possible to conduct an advanced corpus-linguistic study with additional contextual data. The linguistic layer with lemmatic, morphological, and syntactic analyses is aligned with the contextual layer that contains textual information on abbreviations, fragmentary words, and diplomatic segmentation. Technically, in LLCT, the linguistic annotation of the treebank TB.XML files is linked as a stand-off markup

---

[81] Gianollo 2005, 100; Cennamo 2011, 180, 184–185.
[82] Korkiakangas & Passarotti 2011, 113.
[83] For XPointer syntax, see http://www.w3.org/TR/2003/REC-xptr-framework-20030325/ (accessed 7 May 2015).

with the TEI XML files that contain the edited source text provided with the above-described textual information as inline annotations. The technical method is developed by Matti Lassila and published in Korkiakangas & Lassila 2013, to which I refer for an in-depth description of the procedure. The method is fully scalable, i.e. it can be extended to other treebanks with similar characteristics. Indeed, I propose that, in the long run, any serious corpus-linguistic study cannot be based on texts with no textual information provided.[84]

The following three steps describe the workflow required for merging the LDT treebank markup with the LLCT-specific markup of the TEI XML files. A more detailed documentation of the workflow is available in the transformation script files themselves.[85] First, both the treebank and TEI XML files are saved in a database engine with built-in XQuery processor.[86]

- An XQuery transformation produces the XPointers that link the sentence and word identifier numbers of the treebank file with the corresponding words in the TEI XML file.
- Another XQuery transformation converts the TEI XML elements relative to diplomatic annotation into attributes for each word element in the XPointer file.
- Yet another XQuery transformation merges the newly created diplomatic attributes of the XPointer file with each word element of the TB.XML treebank file.
- An XSL transformation converts the merged TB.XML treebank file into the PML format required by the Prague Markup Language Tree Query application (PML-TQ).[87]

The XSL style sheet must be modified in advance to include the diplomatic custom attributes. The customised attributes must also be included in the XML schema that is used in the TrEd Tree Editor, the search engine that best meets the needs of the current study.[88] The TrEd Tree Editor comes with a specific PML-TQ macro add-on that makes it possible to perform queries on treebanks that are annotated with the Prague Dependency Treebank style. The PML-TQ is a potent tool that allows complex queries simultaneously on all the annotation layers of the treebank. The PML-TQ language

---

[84] An important initiative of annotating text re-uses of fragmentary authors with the CTS and CITE URN syntax has, indeed, been realised by the Perseus Project (Almas & Berti 2013).

[85] The complete scripts can be found online in Github (https://github.com/mjlassila/linguistic-annotation-merger).

[86] We have been using the BaseX XML database editor (http://basex.org).

[87] The XQuery transformation scenario is designed by Bridget Almas (Perseus Digital Library Project). The XSL transformation style sheet is modified on the basis of the Prague Dependency Treebank style sheet by Francesco Mambrini (German Archaeological Institute, Berlin). The Prague Markup Language (PML) is a data format developed for the Prague Dependency Treebank.

[88] The TrEd Tree Editor has been developed by Petr Pajas and Peter Fabian for the Prague Dependency Treebank and is aimed at querying multi-layer annotated treebanks (http://ufal.mff.cuni.cz/tred/) (Štěpánek & Pajas 2010, 1828–1830).

is based on programming language Perl, and also provides filters that can be used to filter or classify the query output. By way of example, it is possible to search with a single query all the formulaic inanimate 3rd declension singular noun and adjective subjects that are attached to their finite verb via two coordinating conjunctions on condition that the verb is in the indicative mood and the sentence belongs to a charter written in Lucca before the year 774. This imaginary query can be further provided with a filter that, for example, counts the numbers of the subject occurrences by case form and declension.[89]

It is easy to understand how much the corpus study benefits from the possibility of combining flexibly all the annotation layers that have been discussed in chapter 2. It has to be recalled that in addition to the mentioned textual and linguistic annotation layers, LLCT also contains archival-contextual metadata that come directly from the original text editions. These archival-contextual metadata (date, place, scribe, document type, number of original edition), some of which were mentioned in the above query example, have been added to each sentence element in the PML format. Encoding sufficient metadata in LLCT is important because, in so doing, the queries can be focused, for example, on certain document types, on certain geographical areas, or on certain scribes. Thus, contextual annotation is the technical means of analysing the diachronic, diatopic, and even diastratic variation within the corpus. The complete LLCT treebank in PML format can be freely downloaded from my GitHub repository.[90]

2.5. Basic query subset

2.5.1. Defining basic query subset

The normal corpus linguistic approach, in which the queries are applied equally to all the linguistic categories of the corpus, is usually not possible in the highly formulaic charter Latin of LLCT. Some lexical, morphological, and syntactic phenomena are overrepresented in the corpus due to the repetitive nature of charter texts. It is, however, not only this corpus-specific formulaicity that makes it impossible to utilise everything that the corpus contains. When studying subject case, the syncretism between two or more inflexional forms as well as the ambiguity of certain non-standard phenomena typical of Late Latin must also be considered.

---

[89] The TrEd User's Manual is found at https://ufal.mff.cuni.cz/tred/documentation/ar01-toc.html (accessed 13 May 2015).
[90] LLCT in Github: https://github.com/timokorkiakangas/LLCT/blob/LLCT/LLCT-with-new-attributes.pml.xml

To cope with these restrictions, I shall focus most queries on a subset of the treebank where certain categories that might disturb the corpus linguistic analysis have been removed. In the following, I will explain how this basic query subset is composed. I treat first case syncretism, then the problems related to certain part-of-speech classes, and, finally, overrepresentation of some linguistic categories. In order to do all this, I will have to scrutinise in detail certain, sometimes apparently marginal, phenomena of the Latin declension and their appearance in written code.

At this point, it is worth mentioning once again that I have defined as 'subject' only the head of the subject NP and not the whole NP. This is necessary in a material where the modifiers of the NP head can be in different case forms. The chosen method is also congruent with the annotation strategy of the LDT *Guidelines* that reserve the SBJ tag only for the subject NP head, its modifiers being labelled with ATR tags. For a discussion about subject NPs in LLCT, see section 5.1.

Case syncretism and other identical endings

As the basic query subset is used to examine the case distribution of subjects, i.e. mostly forms that can be interpreted as nominatives and accusatives, only those categories can be included where a formal difference between the two cases is, at least in theory, visible. Certain morphological forms of standard Latin nominal paradigms are identical in form, i.e. they are syncretistic. This is the case of, for example, the neuter nominative and accusative forms in all declensions: e.g. the following nominative/accusative pairs *verbum*/*verbum* 'word', *aedificia*/*aedificia* 'buildings', and *caput*/*caput* 'head'. In LLCT, there are also several common and proper names of Germanic origin, e.g. *morgincap* 'morning gift' and *Walfrid*, that do not adhere to the standard inflexional system but remain indeclinable in all contexts. All these cases are excluded from the basic query subset.

In late spoken Latin, the 1st declension singular nominative/accusative forms were undoubtedly identical because the word-final /m/, which had been the only difference between the two forms, was no longer pronounced. Given that the developments of spoken language infiltrated gradually into the written code, the addition of <m> became a question of education. Once the final /m/ and the possible subsequent nasalisation had disappeared, writing <m> depended entirely on how well the writer knew the standard grammar and spelling. Hence, omissions of final <m> appear in texts

that can be described as lower literacy level. Therefore, I have deemed it wise to exclude also the 1st declension singular from the query subset.[91]

In the 2nd declension singular, instead, it was theoretically possible to maintain the functional difference between nominative *-us* and accusative *-um*/*-o* longer because the difference continued to be phonologically grounded: there was still the final /s/ (or residues of it) and the vowel quality might have been different.[92] The same also applies to the 3rd declension nominative and accusative endings *-is* vs. *-e(m)*. By this, I do not claim that the phonetic erosion of word-final consonants was the only or primary reason for the loss of the Latin case system, which I see to be bound to a more general realignment of grammatical relations. Yet, a phonetic difference between two forms is required by any functional opposition.

The preceding is the basis of the well-known ancient French bicasual system where the case opposition was present only in the 2nd declension and in the imparisyllabic nouns of the 3rd declension. As for the 1st declension, no trace of case opposition was preserved in ancient French. Ancient French is not likely to have been the only instance of this type of development. It seems plausible enough to assume that there was, as a necessary stage of development, as transitory as it may have been, a case system of similar type in each local variety of Latin that would become a Romance language. In ancient French, the 2nd declension singular, the *cas sujet* and *cas régime* endings are *-s* and Ø, respectively. In the 3rd declension singular, the *cas sujet*/*cas régime* opposition appears between *-s*/Ø and Ø or, e.g. between *-eor* and *-ere*, respectively.[93]

The 3rd declension plural ending *-es* creates an exception within the basic query subset. As there is no formal distinction between the nominative and accusative *-es*, the ending ought to have been excluded from the basic query subset. In LLCT, however, the 3rd declension plural subjects, which are rather infrequent, are restricted exclusively to formulaic settings, where they occur with modifiers of other declensions. Therefore, the case form of the subject head can be deduced from the case

---

[91] Excluding the 1st declension singular is also recommended by Ledgeway (2012, 329) and Adams (2013, 240). The weakening and eventual dropping of the final /m/ and the probable nasalisation of the preceding vowel was of ancient origin and belonged apparently to all sociolects in the Later Republic (Adams 2013, 128–129; Beckmann 1963, 186–187; see *Quint*. 9.4.40 and Velius Longus in *GL* VII.54.13–15). It is highly probable that by the Late Latin period, the word-final /m/ was not pronounced in any manner in spoken language (Adams 2013, 128–129; Ledgeway 2012, 22; on the Scipio inscriptions, see Beckmann 1963, 18). The date of the disappearance is unknown (Beckmann 1963, 185–188: in pre-consonantal position, at the end of the 1st century BC; in pre-vowel position, after Quintilian; Beckmann provides different estimates for different declensions; see Väänänen 1966, 71–77, for Pompei).

[92] On the phonological development of the 2nd declension nominative and accusative singular endings, see the references in section 4.2.3.

[93] Sornicola 2011, 18–32; Buridant 2000, 63–69; Schøsler 1984, 29–63.

form of the modifier, such as *mei*/*meos* (see (16)), and the head can be included in the basic query subset. This is done in spite of the fact that the attribute and its head may sometimes be clearly in a different case (see e.g. the attributes in (14)).

(16) MED 291 (AD 800) *si* [--] *ego vel successores <u>mei</u> tibi Wiccheramo duci* [--] *intentionare aut subtragere presumserimus*

"if [--] I or <u>my</u> successors dare to contest or reverse [the property] from you, Wiccheramo, the duke"

Nouns and adjectives

As regards parts-of-speech, the basic query subset consists merely of nouns and adjectives because they have the most regular inflexion. Pronouns, numerals, and participles involve certain problems that would be difficult to handle in the analysis. Pronouns, especially demonstrative, relative, and indefinite pronouns, which are all very well represented in charter Latin, involve a large amount of ambiguity that is mainly phonologically conditioned. Certain pronouns had mixed or otherwise atypical paradigms even in standard Latin. It is likely that, in Late Latin, the pronoun system had already developed considerably towards the Romance system and, hence, it is not clear which graphical variants correspond to the real language usage, which are archaisms, which pure mistakes, and which combinations of the previous. Declinable numerals are relatively infrequent and they exhibit ambiguity comparable to the pronouns. Personal pronouns will be touched on later when the impact of animacy in case marking is examined.

Participles, gerunds, and gerundives, which are all annotated with the same tag in LLCT, rarely occur as subjects (only 19 instances in categories A1–4). As almost all of these subjects seem to be somehow exceptional or ambiguous, they have been excluded from the analysis. For example:

(16) MED 231 (AD 790) *ita ut in mea esset potestatem in omnibus secundum Dominum <u>ordinandum</u> et ipsos pauperos <u>pascendum</u>*

"so that it would be in my power to <u>arrange</u> everything according to the will of God and to <u>feed</u> those poor people"

(17) MED 77 (AD 740) *liceat ei sine aliqua taxxationem in ipsa cella uiuere et ad ipsa sancta Dei uertutem <u>seruiendum</u>*

43

"he is to be allowed to dwell in that cell and to <u>serve</u> in that holy church of God without any rent"

(18) CDL 40 (AD 728) *sicut in suprascriptas dotis a me Radchis et bone memorie Ansfred germano <u>cunfirmatum</u> legitur*
"as it says in the above-mentioned donation that it [is] <u>guaranteed</u> by me, Radchis, and by [my] late brother, Ansfred"

(19) CDL 230 (AD 769) *chui de ipsa res* [--] *uolueri dare, suptragi quesieri, dupla <u>meliorata</u> chonponatur*
"[if] he might give anyone anything from that property [or] tries to reverse [it], he shall compensate double the <u>ameliorated property</u>"

The gerunds *ordinandum* and (*pauperos*) *pascendum* in (16) are clearly coordinated subjects. As gerunds and gerundives are used instead of infinitives (17), and vice versa, it is not obvious which has been the functional status of the gerunds in (16). In (18), *cunfirmatum* seems to be part of an elliptic structure where the infinitive *esse* is left out, perhaps due to the scribe's unfamiliarity with the *legitur* passive phrase. In (19), *meliorata* has become the subject of the passive verb *chonponatur*, probably because of a contamination of the charter formula. It is to be noticed that restricting the query subset only in nouns and adjectives (including adjectival pronouns) decreases the number of analysed subjects by 66.7%: the remaining 33.3% corresponds to only 3,098 subjects.

Overrepresentation

The third and conceptually the most controversial justification of excluding categories from the basic query subset is overrepresentation. The phenomenon is related to the discussion of the previous passages: certain syncretistic categories or certain part-of-speech classes can be overrepresented compared to other classes due to the formulaicity of charter Latin. This is, indeed, the case with certain pronouns: e.g. the word forms *ego* 'I' and *qui* 'which' occur 3,252 times and 2,028 times in LLCT, respectively, holding thus the third and the fifth places of the frequency distribution (see Appendix 2.1.).[94] Both *ego* and *qui* would be excluded from the basic query subset because of their

---

[94] The high frequency of *ego* results from the fact that each autograph subscription starts with that pronoun: e.g. CDT 43 *ego Perso rogatus ad ambes parti me testi subscripsi*. The declarative part of each charter also usually contains *ego*:

overrepresentation unless the whole class of pronouns had already been eliminated on the basis of their often ambiguous inflexional paradigms.

Because of the formulaicity of charters, any linguistic class can be overrepresented if it belongs to a formulaic phrase that is present in all or nearly all the charters of the corpus. For example, the 4th declension noun *manus* 'hand' is an integral part of the subscription formulae that indicate the names of witnesses, e.g. (20) and (21). As a consequence, the lemma *manus* is the second most frequent noun in LLCT, directly after *ecclesia* 'church' (see Appendix 2.1.). Even the nominative form *manus* occurs 1,626 times with its sixth position in word form listing. Although *manus* overrepresents the otherwise infrequent 4th declension, it has no direct relevance to this particular study, which focuses on the encoding of the subject. In general, I allow this kind of overrepresentation in the basic query subset because the exclusion of all the other part-of-speech categories except for nouns and adjectives has already reduced the number of the most troublesome cases.

(20) CDT 23 (AD 785) *signum + manu Rachiperti de Cosuna testis*
"sign + [= Holy Cross] of the hand of Rachipertus of Cosuna, the witness"

(21) CDL 288 (AD 774) *ego Uuileradu presbiter testi manu mea subscripsi*
"I, Wileradu, the priest and witness, signed in my own hand"

Overrepresentation turns out to be problematic when a frequently occurring form results from an extra-linguistic contamination of charter formula. Excluding these phenomena, typical of charter Latin, requires an unquestionable justification, and I shall carry out exclusions of this kind only after a meticulous case-specific deliberation. I shall discuss a few extra-linguistic contaminations in section 2.5.1. and section 2.5.2. The fact that there are several probable extra-linguistic contaminations that cannot be proved as such reduces, of course, the reliability of the results drawn from the quantitative analysis of LLCT.

On the other hand, formulaicity does not seem to skew essentially the distributions of the most fundamental linguistic categories. For example, the part-of-speech profile of the charter Latin of LLCT resembles quite closely the profile of Caesar's *De bello Gallico* 2, a prototypical example of Classi-

---

e.g. MED 798 *manifestu sum ego Aufridi presbitero filio quondam Fridiperti quia convenit mihi*. The pronoun *qui* occurs most often in formulaic phrases, such as MED 798 *proinde ego qui supra Aufridi presbitero*, MED 214 *qui hanc cartula fieri rogavit*, or MED 228 *qui uno capite tenet*.

cal Latin prose (narrative) written 800 years earlier (see Table 2.5.).[95] It is mainly with verbs that the corpora differ from each other: the juridical language of LLCT abounds in nominal constructions at the expense of verbal expressions. I find this close resemblance a positive piece of evidence of how the trends attested in LLCT can be extrapolated to (Late) Latin in general.

Table 2.5. The frequencies of the part-of-speech classes in LLCT and in Caesar.

Part of speech	LLCT		Caesar	
	N	%	N	%
noun	57,622	31.3%	421	30.5%
adjective	22,702	12.3%	141	10.2%
verb	21,833	11.9%	228	16.5%
pronoun	21,613	11.7%	128	9.3%
preposition	20,693	11.2%	138	10.0%
conjunction	19,734	10.7%	143	10.4%
adverb	10,064	5.5%	95	6.9%
participle	9,055	4.9%	67	4.9%
numeral	723	0.4%	18	1.3%
∑	184,039	100%	1,379	100%

Dative/ablative and genitive subjects

LLCT also contains subjects that appear to have neither nominative nor accusative case marking. These subjects that are classified as genitives (1 occurrence) or dative/ablatives (61 occurrences) are obvious errors that cannot have a purely linguistic motivation. The genitive-form subject of (22) is probably a reanalysis of a (crystallised) partitive expression that is used as the base form of the noun *modius*:

(22) CDL 138 (AD 759) *simulque et inter uinea et terra iniui ad* [= *ab*] *nus offerta esse uolomus in suprascripto loco Ansulari*, *id est* <u>*modiorum*</u> *sex*
"similarly, we want it to be offered by us in the above-mentioned place Ansulari, between the vineyard and the plot, namely six <u>*modii*</u> [of land?]"

The dative/ablative-form subjects, such as in (23), are more common (61 occurrences) and can be claimed to have a clearly extra-linguistic role in charter Latin. Due to their extra-linguistic

---

[95] The Caesar data are excerpts from the 2nd book of *De Bello Gallico* from the Perseus Latin Dependency Treebank (http://nlp.perseus.tufts.edu/syntax/treebank/latin.html, accessed 9 May 2015).

contaminational nature, these dative/ablative- and genitive-form subjects are left out of the basic query subset.[96]

(23) CDL 149 (AD 761) *si aliquando ego uel* <u>*successoribus meis*</u> *tibi uel ad tuis heredis suprascripta rem* [--] *subtragere aut intentionare presumpserimus*

"if I or <u>my successors</u> sometime dare to dispossess or contest [--] the above-defined property from you or your heirs"

It is true that excluding the extra-linguistically motivated dative/ablative or genitive subjects from the analysis results in problems of a methodological nature. It is questionable to exclude only those instances that stick out from the horizon of expectation, but leave all the conformant instances without further investigation. It must be remembered that there can be several cases of extra-linguistic contamination that result, for example, in accusative subjects that, in this study, end up being interpreted as attestations of morphosyntactic realignment because they are imperceptible. What ought to be done would be a) to inspect further the distribution of lemmas in all the structural categories, not only in the seemingly deviant ones, and b) to examine then all the top-frequent lemmas with the same criteria. This is what section 2.5.2. and section 2.5.3. seek to do. Unfortunately, the investigation often proves to be impossible, as in most cases no information can be found about the motivation of a certain linguistic form.

In this chapter, I have charted several linguistic categories that are not appropriate for a corpus-linguistic analysis of substandard and formulaic Late Latin material. Those categories have been excluded from what I call the basic query subset. It is still necessary to investigate whether the remaining categories contain some formulaicity-related overrepresented phenomena that may skew the results of the future queries. In the next two subchapters, I shall have a look at the most frequent word forms in LLCT. Based on this examination, I shall exclude two word forms, *heredes* and *misso*, from the basic query subset. It is important that the most frequent lexical items are not extra-linguistically motivated because they easily lead the quantitative analysis astray due to their frequency.

---

[96] Note that although in non-standard Latin texts, a noun, such as the personal name *Candido* (CDL 45), can be interpreted as an accusative or a dative/ablative depending on the context, the annotation style of LLCT does not allow dative/ablative-form subjects in the singular because subjects in *-o* and *-e* are always annotated as accusatives.

To conclude this subchapter, I summarise briefly the structure of the basic query subset as described above. Table 2.6. lists the categories that have been excluded from the analysis of case alignment. Many of the categories overlap: for example, there are neuters and 1st declension singulars in the categories of pronouns and participles. Therefore, each of the numbers of eliminated subjects in the last column of the table is calculated after eliminating the antecedent category. This is to say that the number of neuter singulars and plurals does not contain any word that belongs to the previously eliminated category of pronouns, numerals, participles, and verbs. Similarly, the number of indeclinable words does not contain neuters and so forth.

Table 2.6. Basic query subset: numbers of eliminated subjects.

Category		N
Total of subjects in LLCT		9,308
Eliminated categories	pronouns, numerals, participles, and verbs	6,210
	neuter singular and plural	574
	indeclinable words of Germanic origin	171
	1st declension singular	534
	other cases than nominative/accusative	62
	heredes and *misso*	174
Total of basic query subset		1,583

The numbers in the last column indicate that 7,725 subjects of the total in LLCT (9,308) are left out when performing queries on the basic query set. Thus, 1,583, i.e. only 17.0% of all the subjects, are included in the basic query subset. Actually, even this number needs to be reduced for certain analyses of section 5.1. because 10 of the subjects are related to two or more coordinated verbs that have different semantic values. Thus, the queryable number of subjects may be as low as 1,573. Moreover, a subset of 3rd declension imparisyllabic subjects (392 subjects) is sometimes used in chapter 4 and chapter 5 (see 4.2.2.). In this connection, it is worth repeating that all fragmented and ambiguously abbreviated words have been excluded from the basic query set as well as from all the corpus analysis of LLCT (see 2.2.1.).

2.5.2. *Heredes*

This and the next chapter discuss at some length a few possibly contaminational high-frequency words and their exclusion from the analysis. The following 'case study approach' was chosen not only to find out the skewing word forms, but also to show what kind of challenges charter Latin poses to research and how they can be successfully met. The two most frequent word forms of the

non-personal-name accusative subjects, i.e. *heredes* 'heirs' and *misso* 'envoy', seem to be related to a contamination and, consequently, will be removed from the basic query subset. The analysis uses as the point of departure the list of the ten most frequent word forms of non-personal-name nominative and accusative subjects in LLCT (Table 2.7.). The bottom line indicates the totals of the accusative and nominative subjects after eliminating all the other categories mentioned in Table 2.6. except for *heredes* and *misso*.

Table 2.7. The ten most frequent word forms of non-personal-name nominative and accusative subjects.

Accusative subject	N	Nominative subject	N
(h)eredes/(h)erid(e/i)s/heredi	68	*(h)eredes/(h)erid(e/i)s/heredi*	78
misso	28	*re(s/m)/re* (singular)	69
portione(m)	17	*utilitas/hutilitas/autilitas*	39
res/ris/rem (plural)	12	*homo*	36
medietate(m)	9	*filii(s)/fili*	26
dom(o/um)	8	*sacerd(o/u)(s)/sacertdus*	19
homines	6	*pars*	18
un(o/um)	6	*vol(u/o)ntas/volum(p)tas/bolumtas*	17
(h)offersionem	5	*domus*	15
breve(m)	5	*nullus*	13
Total of accusative subjects	642	Total of nominative subjects	1,115

These frequencies are heavily affected by the formulae that determine the abundance of several lemmas typical of the text type: for example, *heres* 'heir' and *res* 'property/thing' are archetypal representatives of charter Latin. However, the top ten of the nominative and accusative subjects is not equally distributed, which suggests that there can be different motivations for the occurrence of certain words/word forms as nominative or accusative subjects.

I start with *heredes*. High frequency alone does not make any word form suspicious, but when a suspicion arises of a linguistic atypicality, the word form requires further investigation. Of the 68 occurrences of accusative plural *heredes* + adjectival personal pronoun *meos/meus*, *tuos*, or *nostros*, 42 are found within a single coordination structure type (structural category A2–4).[97] This is quite a lot, as the total of plural subjects in that category is 199. The phrase *heredes meos/meus* is the most frequent with 54 occurrences. The phrase appears in the *sanctio*[98] part of charters and, depending on the formulae chosen, there can be more than one occurrence of *heredes meos*, *tuos*, or *nostros* in a

---

[97] For the structural categories, see section 5.1.
[98] Pratesi 1979, 75–76.

single charter. In the other structural categories, there are an additional 26 instances of *heredes meos* (2 in A1 and 24 in B2). The following sentences present some common contexts of *heredes*, both in the nominative and in the accusative:

(24) CDL 244 (AD 770) *spondeo me et <u>heredes meos</u> esse conponituros tibi et successoribus tuis praefatam rem*
"I guarantee that I and <u>my heirs</u> will recompense you and your successors for the aforesaid property"

(25) CDL 238 (AD 770) *tam ego quam <u>heredes mei</u> in ipsa casa uestra resedere et habitare debeamus*
"I as well as <u>my heirs</u> must reside and dwell in that house"

(26) CDL 85 (AD 746) *et si ego qui supra Auselmi uel <u>meus eredes</u> ipsa suprascripta casa et res eius bene non guernaremus et angaria* [--] *minime perexolserimus*
"and if I above-mentioned Auselmi or <u>my heirs</u> do not govern well the above-defined house and its property and if we [--] do not perform the corvée"

There seems to be considerable uncertainty about the correct use of this *sanctio* clause. Examining how these non-standard structures may have arisen helps to decide whether they are to be excluded from the analysis of subject case selection or not. This will also be a lesson on how the extra-linguistic factors relative to formulaicity influence charter Latin.

First, it has to be asked whether the accusative *heredes meos* subject could be a genuine reflection of semantically-motivated case alignment change. This is not very likely, as this clear preference for the accusative subject is not attested with any other animate subject lemma and, moreover, *heredes meos* appears with all types of verbs. The exceptional frequency of *heredes* (*meos*) suggests that there must be something else behind it. Indeed, several dative/ablative *heredibus meis* subjects (see section 2.5.1.) are attested in LLCT in place of *heredes meos/mei*, which implies that there could be an extra-linguistic connection of the two non-standard uses. In the following, I examine the possible origin of *heredes meos* (and *heredibus meis*) used as subjects in the whole of LLCT, not only in the structural group A2–4. In general, *heres* is a very frequent noun in LLCT with 909 occurrences, 779 of which are with *meus*, *tuus*, or *noster*.

50

Heredes meos, or *meus eredes* as in (26), may have originated from the confusion of an accusative and infinitive structure (24) with a finite verb clause (25), which are both found in LLCT. This is, however, far from indisputable, as the accusative and infinitive alternative is rarely attested and, in the corpus of LLCT, later than the finite and contaminational alternatives: the first accusative instance, *meus eridis*, is in CDL 26 from AD 720, i.e. right at the beginning of the corpus. It is useful to compare with each other the chronological dispersions of the three case forms that are attested in the plural of *heres* when used as a subject. In the dispersion plot of Figure 2.3., each instance of such subjects is marked with a cross on the line that indicates time in years.

Figure 2.3. Chronological dispersion of the plural of *heres* (nominative, accusative, dative/ablative) used as the subject.
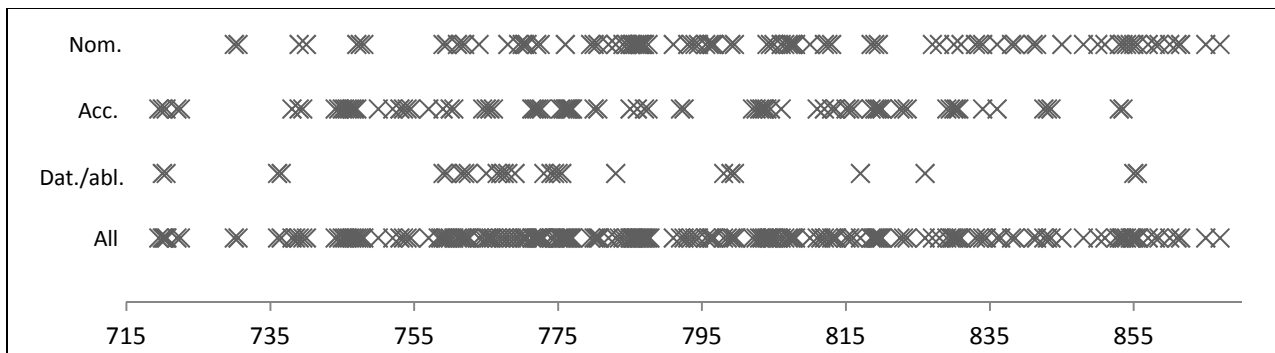
The bottom line shows the dispersion of all the plural *heres* subjects, i.e. the nominative, accusative, and dative/ablative-form subjects all added up. The plural *heres* formula is very common, though it does not figure in every charter. The scarcity of the material during the first decades of the corpus results in gaps at the beginning of the plot. The plural *heres* subject type seems to be at its most frequent during the last decades of CDL (the 760–770's), which are overrepresented within the corpus (see section 2.1.2.). It is also to be noticed that its frequency seems to decline towards the end of the time span, which could indicate an incipient change in formula usage.

What is striking is that the accusative-form *heredes* (*meos*) subject is attested mainly at the beginning and in the middle of the corpus but only occurs a couple of times in the last thirty years under examination. Equally remarkable is that the 'correct' nominative-form *heredes mei*(*s*) does not appear at all at the beginning and only becomes common after AD 755. It remains the predominant form until the end of the corpus. The dative/ablative *heredibus meis* subject is less frequent. It is found scattered almost over the whole time span, with the major conglomeration during the last decades of the Lombard kingdom (c. AD 755–775).

51

Figure 2.3. cannot resolve, as such, the problem about the origin of the *heredes meos* subjects, even though the dispersion of the accusative and nominative subjects can be seen to favour the interpretation according to which the accusative form had been borrowed incorrectly from the original correct structure (whether it be an accusative and infinitive or not) to a novel finite-verb structure. Only later on would the scribes have normalised the accusative into the nominative, hence its later proliferation in respect to the accusative variant. This would imply that the contamination had occurred before the time span observed here. It would imply as well that there was a rise in awareness concerning certain standard Latin features in the last part of the period studied in LLCT.

It was noticed that the accusative and infinitive structure is unlikely to be the point of departure of *heredes meos/tuos/nostros* subjects. It is more reasonable to suggest that the new finite-verb construction was coined on the basis of prepositional phrases. This is probable because the prepositional phrases with *heredes* and *heredibus* abound in the corpus from very early on (441 occurrences) even though they do not seem to occur so frequently during the first decades of the corpus – this is what was observed with the plural *heres* subjects as well.[99] The fact that a considerable proportion of dative/ablative *heredibus* complements is closely related to the subject argument of the sentence may have been the point of departure of the contamination: *ab heredibus* constructions are agent terms in sentences such as (27), and *cum heredibus/heredes* constructions express partnership in action with the subject (28).

(27) CDL 219 (AD 768) *et neque a me neque <u>ab heredibus meis</u> aliquando hanc meam offersionem posse* [= *possit*?] *disrumpi*
"this my donation should be never violated by me or <u>by my heirs</u>"

(28) MED 236 (AD 792) *unde repromitto ego q(ue) s(upra) Georgiperta abbatissa una <u>cum heredibus</u> et successatricibus meis tibi Gumprando diac(ono) vel heredibus tuis*
"therefore, I, the above-mentioned Georgiperta, the abbess, <u>with my heirs</u> and successors promise to you Gumprando, the deacon, and to your heirs"

The transition of the phrase *heredibus meis/heredes meos* into the new finite-verb structure seems to have been straightforward. No adaptation of case form has taken place, probably because these for-

---

[99] Only 8 of the occurrences of plural *heres* are without the personal pronoun attributives *meus*, *tuus*, *noster*, and *vester*.

mulaic high-frequency phrases had already crystallised as unchangeable structural blocks of charter Latin. The study of hypercorrect pronoun constructions in modern spoken English has shown that frequency is one of the factors determining the occurrence of hypercorrection. Indeed, the cognitive process of structural priming can describe how recurrent structural patterns are echoed in other unrelated structures, thus resulting in hypercorrection. The concept of priming can also be applied to the generalisation of contaminations in written language. The more frequently a construction is encountered, the more activated it is and the more likely it is to be used subsequently by other writers, in other utterances, and in other clauses.[100]

Table 2.8. Distribution of the prepositions used with dative/ablative and accusative plural forms of *heres*.

Dative/ablative complements			Accusative complements		
	N	%		N	%
ab	46	27	*ab*	4	1
ad	17	10	*ad*	77	29
cum	70	41	*cum*	158	59
da	2	1	*de*	27	10
de	36	21	*apud, contra, per, super*	4	1
	171	100		270	100

The prepositional phrase explanation of the origin of the *heredes meos* subjects seems plausible, as it also explains the presence of linguistically completely unmotivated dative/ablative *heredibus* subjects. Table 2.8. shows that the agentive *ab* and *cum* are together responsible for 68% of the dative/ablative prepositions and about 60% of the accusative prepositions. In this case, it is not important whether the dative/ablative is used classically with dative/ablative prepositions or the accusative with accusative prepositions. What counts is the fact that both accusative *heredes meos* and dative/ablative *heredibus meis* have been recycled as subjects. More generally, Table 2.8. reflects the common picture known from other Late Latin texts as well: the accusative is often associated with the classical ablative prepositions (*cum* and *de*).[101] I shall not discuss here the other patterns arising from Table 2.8.

Further evidence in favour of the prepositional origin hypothesis is that an apparently dative/ablative form *meis* (or less often *tuis*) of the personal adjective often occurs with the seemingly

---

[100] Boyland 2001, 384, 401–402. On the mechanisms of structural priming, see Loebell & Bock 2003, 791–795, 812–814.

[101] On the accusative replacing the ablative in prepositional phrases, see Galdi 2013, 159–164; Adams 2013, 235, 258; Väänänen 1981, 112.

nominative subject *heredes* (29). The final *-s* of *meis/tuis* can be a reminiscence of the original dative/ablative status of the (prepositional) phrase. In LLCT, both parts of the *heredes meis* type subjects are labelled nominatives. If the 3rd declension plural forms already ended in *-i* in the spoken language and if the final *-s* was not clearly distinguished anymore, the forms *mei* and *meis* could not be strictly separated. *Tui heridi* (30) represents the fully developed 'Romance' or 'Italian' form.

(29) MED 247 (AD 794) *et si* [--] *ego Ellaru vel <u>heredes meis</u> molestare aut subtragere quesierimus*

"and if [--] I Ellaru or <u>my heirs</u> attempt to degrade or dispossess"

(30) CDL 106 (AD 752) *tiui Crispino uel at* [= *ad*] <u>*tui heridi*</u> *ispondeo me esset* [= *esse*] *conponiturus duplas tales terra*

"I guarantee to you Crispino or to <u>your heirs</u> that I will recompense doubly the parcel"

The prepositional origin hypothesis supposes that the subjects of the finite-verb constructions were innovated only after the emergence of this specific prepositional phrase in the formula arsenal of charter Latin.[102] It can be seen from Figure 2.4. that the dative/ablative *heredibus* complement is in decline, its heyday being roughly between AD 755 and 820, while the accusative *heredes* complement gains ground especially from c. AD 820 onwards. This is likely to imply that the accusative was recognised as the appropriate case for all prepositional complements by that time, whereas the temporary increase of *heredibus* could be seen as a (sometimes hypercorrect) aspiration towards correction. This increase is obviously connected to the high use of the dative/ablative *heredibus* as the subject as well (see Figure 2.3.).

---

[102] Related formulaic phrases with *heredes* are found since AD 504 in the sales contracts of the Ravenna papyri: e.g. *PItal* 29.10 *tradidit Rustico acolyto ecclesiae catholicae Romanae iure directo <u>h(eredi)bus</u> posterisque eius*, *id est spatium agri*. This *heredes + posteris/successores* formula survives in LLCT: e.g. CDL 73 (AD 740) *et nunquam me <u>heridis, successoris</u> meis auersus* [= *adversus*] *ipsa Dei eclesia* [--] *ire quandoque presumat*. The prepositional recipient construction (30) of LLCT is likely to have developed out of the earlier dative *heredibus* construction.

Figure 2.4. Chronological dispersion of the accusative and dative/ablative plural forms of *heres* used as complements of prepositions.
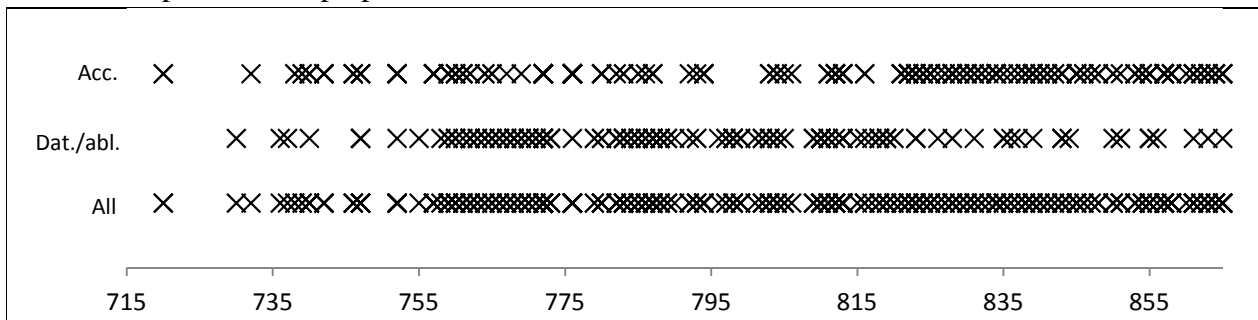
By now, it has been shown that *heredes meos* (and *heredibus meis*) subjects are quite probably of contaminational origin. Therefore, I deem it wise to exclude *heredes meos* subjects from the basic query subset from now on. In the name of impartiality, even nominative *heredes mei(s)* subjects have been removed: they are part of the same phenomenon.

2.5.3. *Res*, *portione*(*m*), *misso*, and other frequent words

In this chapter, I discuss a few other frequent word forms that seem to call for a separate discussion. When used in the subject position, some of these forms involve morphological peculiarities (*res*) or are likely to be at least partly formulaicity-motivated (*misso*). I will start with the case of *res*. In order to facilitate the discussion, I reprint here the topmost lines of Table 2.7.

Accusative subject	N	Nominative subject	N
(*h*)*eredes*/(*h*)*erid*(*e*/*i*)*s*/*heredi*	68	(*h*)*eredes*/(*h*)*erid*(*e*/*i*)*s*/*heredi*	78
misso	28	*re*(*s*/*m*)/*re* (singular)	69
portione(*m*)	17	*utilitas*/*hutilitas*/*autilitas*	39
res/*ris*/*rem* (plural)	12	*homo*	36
medietate(*m*)	9	*filii*(*s*)/*fili*	26

The most frequent nominative-subject word form after *heredes* is the singular *res* with 69 occurrences. The plural form of *res* also figures at the top end of the accusative subjects with 12 occurrences while there are only 5 occurrences of singular *res* in the accusative subjects. In the context of charter Latin, *res* 'thing' usually has a technical meaning '(piece of) property' or, more specifically, 'piece of the movables belonging to a farm house'. Understandably, *res* had been from very early on an important term in charter Latin, where transactions involving land and farm houses, including

their tenants, are usually dealt with.[103] The following phrases illustrate the common contexts of *res* in LLCT:

(31) CDL 182 (AD 764) *ut amplius de hac <u>re</u> nullo tempore aliqua oriatur intentio*
"so that no further contention should ever arise concerning this <u>property</u>"

(32) CDL 73 (AD 740) *nulla aueas potestatem <u>re</u> subtrahendi*
"you do not have any right to dispossess that <u>property</u>"

(33) CDL 40 (AD 728) *omnes <u>res</u> iuidem pertenente in tua defensionem et dominio ualeat permanere*
"all the <u>thing/things</u> that belong there should remain in your command and dominion"

(34) CDL 100 (AD 750) *ut [--] ecclesiam et omnia ea ipsas <u>res</u> suprascriptas in mea sit potestatem*
"so that [--] the church and all that above-defined <u>property</u> is in my possession"

The examples show that *res* is used in several different phrases that may be formulaic but that do not seem to limit the use of *res* only to certain kinds of phraseological contexts, as was the case with *heredes* and as is the case with *misso*, too, as will be seen below. The prepositional phrases, such as (31), are frequent and they often preserve the seemingly standard dative/ablative form, which can be held as evidence of formulaicity-driven crystallisation. The object (e.g. (32)) is also often *rem* or *re,* although *res* is gaining ground all the time (see below).

The phrases exemplify the problems involved in the interpretation of the word forms *res*/*rem*/*re*. Due to the weakening of the word-final sounds and to the merger into the 3rd declension of the 5th declension, the monosyllable *res* seems to have lost most of its case inflection. For example, phrase (33) shows how it is sometimes impossible to tell whether *res* should be interpreted as singular or plural. The verb does not necessarily agree with the number of the subject. Loss of agreement, however, is very common in charter Latin. By far the most frequent form is *res*, both in the nominative and accusative function, in the singular and in the plural (1,016 instances of all the 1,671 occurrences of the lemma *res*). This is presumably a (hypercorrect) reminiscence of the Classical Latin 5th

---

[103] Both correct and crystallised forms of *res* are found in the Ravenna papyri. The *Tablettes Albertini* contain only correctly inflected forms.

declension syncretism of the nominative singular/plural and accusative plural forms (all are *res*). Diachronically, the form *res* seems to crystallise into an indeclinable form at the end of the time period involved in LLCT.[104]

As far as can be perceived from the above examples, the behaviour of *res* does not result from the influence of formulaicity, but is an independent morpho-phonological development. It cannot be compared to the case of *heredes meos/tuos/nostros* because it is not the entire NP that is crystallised, but only the morphophonological unit of *res*: the attributes of *res* remain declinable and serve to indicate the syntactic function of their head.[105] The most common of the attributes is the article-like *ipse* (*ipsa*, *ipsas*), but *hic*, *omnis*, *pertinens*, *is*, *suprascriptus*, *iam dictus*, and many more occur as well, as can be seen from the above examples. What further supports the relative independence of *res* is the fact that even the most formulaic instances of the phrases where *res* is involved, such as (34), are not 'closed' in the sense that *res* can be replaced by other lexical items when required: compare (34) to (35) and (36), where *case* and *portionem* are used in place of *res*.

(35) MED 582 (AD 843) *ut* [--] *iam dicte <u>case</u> cum rebus suis in tua sint potestatem*
"so that [--] the already mentioned <u>houses</u> with their belongings are in your possession"

(36) CDL 239 (AD 770) *ut <u>portionem</u> eius de omnibus rebus nostris non sit in potestatem ipsius ecclesiae*
"so that his <u>share</u> of all our property is not in the possession of that church"

On the basis of all that has been said above, I shall not exclude *res* from the analysis of subject case selection. *Res* is a predilect word in charter Latin and, moreover, it seems to be natural that *res* occurs so frequently as an accusative-form subject. Actually, most of what is said about *res* also applies to *portione*(*m*), the third most frequent word form of the accusative subjects: there seems to be no binding evidence against its inclusion in the analysis.

There are 217 instances of the lemma *portio* in LLCT: the great majority of them, i.e. 199, are *portione*(*m*); 7 are *portio*, and 11 *portionis*/*portiones*. The form *portione*(*m*) covers the syntactic functions of object, subject, and prepositional complement. Five of the seven instances of the stand-

---

[104] Curiously, this preference for the form *res* is not largely attested in the continuator forms of *res* in the Romance languages: e.g. Fr. *rien* derives from the accusative form *rem*.
[105] When annotating the case form of the words *res/rem/re*, the case form of the attribute is obviously decisive.

ard nominative form *portio* are used as objects and the remaining two as complements of the preposition *de*. No subject is found with the form *portio*. Thus, the longer (accusative-form) stem seems to have nearly ousted the shorter (nominative-form) stem. This is nothing to marvel at because the word is typical of low-transitivity contexts similar to those where *res* is often found: as was seen in (36), *res* and *portione*(*m*) are interchangeable in the phrase *sit* [--] *in potestate*.

It is possible that the highly frequent form *portione*(*m*) results, at least in certain cases, from a confusion of a supposedly older object/prepositional phrase structure with a supposedly novel subject structure. The situation is, however, essentially the same as that of *res* above. They both differ from the *heredes* constructions where it is only the word itself that appears to be crystallised and not the entire phrase, as in *heredes meos/tuos/nostros*. Additionally, *portione*(*m*) is attested in several differing contexts unlike *heredes*.

Even more difficult is the case of *misso* 'envoy'. In the preceding chapter, I decided to exclude *heredes* from the analysis. I consider *misso* another possible contamination peculiar to charter Latin though the case is not as obvious as that of *heredes*. The word form *misso* appears 28 times and *missus* 2 times as a singular animate subject in LLCT. All the occurrences are found in categories A1 (10) and A2–4 (18). Aside from the subject use, lemma *missus* appears 104 times in other syntactic contexts including mainly prepositional phrases and direct objects. In the following few passages, I seek to point out the parallel nature of the contamination underlying both *misso* and *heredes*.

Missus is the perfect participle of the verb *mittere* 'to send'. From Classical Latin onwards, there existed a 4th declension abstract deverbal noun *missus*, *-us* 'sending away'. The human-referent 2nd declension deverbal noun *missus*, *-i* is attested with Christian authors with the meaning 'the one who is sent', 'messenger (of God)' (Tert. *praescr*. 20, Arn. 2.73, Isid. 7.2.35) though it has more participle-like predecessors in military contexts (Caes. *Gall*. 5.40.1, Sil. 17.77). In 6th to 8th-century Gaul, the word seems to have assumed an administrative sense: *missi domnicii* were royal messengers or envoys and *missi* (without a qualifier) episcopal or secular messengers. The term was still in use during the Carolingian regime in Italy. In Tuscan charters, *missus* appears in the case of royal, episcopal, and secular messengers.[106] The modern Italian *messo* continues the Latin word and meaning.

---

[106] *TLL* VIII, 1191; Blaise 1975, 593.

(37) CDL 56 (AD 736) *constat me Lupo uirum honestum uinditorem* [--] *uindedissit et uindedit, tradedissit et tradedit uobis domno Uualpert gloriosissimo duci per* <u>*misso*</u> *uestro Fusio* [--]

"it is manifest that I, Lupo, *vir honestus*, the seller [--] have sold and handed over to you, lord Walpert, glorious duke, through the hands of our <u>envoy</u> Fusio [--]"

(38) CDL 256 (AD 771) *quando ad ipsum usumfructum ibidem uenerit rector ipsius ecclesiae uel* <u>*missus*</u> *eius*

"when the rector of the church or his <u>envoy</u> comes to the *ususfructus*"

(39) MED 452 (AD 823) *et quando ibidem tu aut* <u>*misso*</u> *tuo veneritis super vindemia mecum vendemmiandum*

"and when you or your <u>envoy</u> come there to pick the grapes with me during harvest time"

The first attestation of *missus* (with a preposition; see (37)) is from AD 736, and there are other early attestations at the beginning of the 770's. The chronological distribution in Figure 2.5. implies that although *missus* has been (a lesser) part of the pre-Carolingian Italian charter vocabulary, it has no doubt gained momentum after AD 800. My hypothesis is that this pre-existing word was turned into a technical term when Frankish administrative innovations introduced a new class of functionaries, *missi*, in Italy. When the word was raised from insignificance to active administrative use, it seems to have been borrowed, for some reason, with its non-standard ending (-*o*). This kind of crystallisation into a functionally unmotivated syntactic form is what often happens with technical terms. Unfortunately, the lending construction is not known, but it could be searched from the Frankish charters of Gaul.

The dominant form is *misso*/*missu*(*m*), which is found 37 times with prepositions (mainly *ad* and *per*), and could be the point of departure for the subsequent proliferation of the accusative form in other syntactic functions. Indeed, the two earliest occurrences are *vendere*/*tradere per misso* and *suscipere per misso* (37). The first instance of the lemma *missus* as a subject dates back to AD 771 (38), where it is correctly in the nominative. Accusative-form *misso* subjects appear only later, around and after AD 815. They occur mainly in a single formula that can be seen in (39). They look like partly infelicitous insertions of an additional participant to the original formula, where only the *rogator* or the receiver of the charter, expressed in the 1st person (plural) or in the 2nd person (singular), was allowed to collect the ground rents paid in kind after the grapes had been harvested.

59

Figure 2.5. Chronological dispersion of singular *missus* used as the complement of a preposition, as subject, and as object.[107]

Like with *heredes*, there seem to be several problems with the use of *missus* in LLCT, and most of them are related to its formulaic status. Of course, this does not imply that all the occurrences of *misso* would be formulaicity-driven contaminations. However, given the clear connection of *misso* with formulaicity, I find it advisable to exclude the singular forms of lemma *missus* from the basic query subset. This prevents skewing the statistical analysis concerning those categories in which *misso* happens to be frequent. The case of *misso* differs from that of *portione*(*m*) or *res*, which I did not exclude, insofar as *portione*(*m*) or *res* occur extensively in various contexts while *misso* is always bound with a very specific construction and future-tensed verb *venerit/veneritis/venerimus*.

I cannot exclude the possibility that there may still be other contaminational constructions in the basic query subset, but *heredes* and *missus* leap out of the data because of their high frequency and the multifold confusions involved in their use. When the occurrences are fewer, instead, it is not possible to state whether the accusative case is due to a formulaicity-based contamination or to a real linguistically-motivated variation. Within a more substantial corpus, more information could be gathered about the possible extra-linguistic motivations behind the high-frequency word forms.

The observations that were made in this and the preceding chapter about the contaminations of formulaic expressions illustrate well the peculiarities of formulaic charter language. Frequency and perceptual saliency may be potent factors when explaining how certain contaminations become common within formulaic charter language and, so to speak, how they integrate into the accepted apparatus of written language.[108] A formulaic phrase or passage is especially prone to be contami-

---

[107] Here, *missus* can be either the head of the phrase (e.g. *ad misso vestro*) or the attribute of the head (e.g. *per misso uestro Fusio* where *Fusio* is the head).

[108] On perceptual saliency, see La Fauci 1997, 41; Winter 1971, 55–61; Ledgeway 2012, 333. Winter (1971, 61) discusses especially the relation of prominence and relative frequency in the Latin case system. For a radically different,

nated if its structure is archaic, if its meaning has become obsolete, or if it is for some other reason challenging from the scribe's viewpoint. If this kind of formula occurs in every charter, it is likely that the contamination may also gain ground outside its original context. In other words, a contaminated phrase that is frequent enough can create a norm. This is probably the case of the above-discussed words and phrases. When there is uncertainty about the correct standard, the forms, such as *heredibus*, *heredes meos*, or *misso*, erroneously sanctified by the charter formula, may be felt to be justified in all contexts of written language.

3. Studying alignment change in Late Latin: theoretical setting

This chapter discusses the recent views on the simplification of the Latin case system. Section 3.1. reviews briefly the research that has been conducted thus far on the change and variation of the Latin nominal declension and presents certain theoretical frameworks and linguistic constraints that have proven important, both in Latin and cross-linguistically. Section 3.2. and section 3.3. discuss the two available sources of data that the theory and the study can be based on: the Latin data and the Romance data. Finally, section 3.4. discusses the problems and restrictions involved in these data and theories.

3.1. Alignment change and its constraints

The Classical Latin six-case system was lost during the transition to Romance languages (with the exceptions of Gallo and Daco-Romance).[109] It is usually suggested that this simplification was gradual ($6 > 3 > 2 > 1$) and happened at a different pace in different areas of the Romance speaking world.[110] The phonetic erosion of some case endings, such as *-am* and *-em* (> *-a* and *-e*), appears to be the imminent cause of the case distinction loss. Some other case endings were, however, resistant to phonetic erosion, so the fundamental motive of the disappearance must be a large-scale systemic change from syntheticity to analyticity (and from dependent to head marking) that was taking place in Late Latin. Phonetic erosion was, thus, an independent development concomitant to the systemic change that led to the Romance nominal system, where the final sounds of nouns are no longer im-

---

cognitive view on frequency and its impact on the formation of constructions, see Bybee & Hopper 2001, 14–15; Boyland 2001, 384, 401–402; Loebell & Bock 2003.

[109] On the Gallo-Romance case system, see Schøsler 1984; van Reenen & Schøsler 2000; Smith 2011, 281–289. On the Daco-Romance case system, see Sala 2005.

[110] Lausberg 1962, 13–15; Herman 1998, 18; Zamboni 1998a, 129–135; Zamboni 2000a, 93, 110–113; Seidl 1995, 100; Clackson & Horrocks 2007, 277; Banniard 1992, 517–519; Faraoni 2014b, 1; cf. de Dardel & Wüest 1993 for a radically different opinion.

portant for encoding grammatical relations (except for number).[111] In this study, I am, of course, interested mainly in how all this was related to Italian Latin/Italo-Romance.

There are at least two methodological approaches to the study of the Late Latin case system: 1) the reconstruction of probable forms and structures on the evidence of Late Latin texts, 2) the reconstruction of probable forms and structures backwards from Romance evidence. These approaches are partly dictated by the nature of the evidence and will be discussed in two sections below (section 3.2. and section 3.3.). Traditionally, the first approach involves reconstructing the logically plausible pathways of development on the basis of qualitative evidence on occasional Late Latin texts and inscriptions, accompanied by opportune examples of Romance developments (e.g. Herman 1987, Zamboni 2000b).[112] The scarcity and uneven temporal and areal distribution of the non-standard texts, together with the general problems related to conservative written texts used as evidence of linguistic change, make the analysis challenging. It is particularly difficult to approximate the chronology of the phenomenon although it would be highly important for evaluating the plausibility of semantic alignment theories.

Using traditional philological methods, Herman dates the stage of the three- or two-case system to the end of the 5th century AD and its subsequent disappearance in Italo-Romance to the interval between the 6th and 8th centuries.[113] Herman's view seems to be rather commonly accepted. Cennamo (2009) sketches a more detailed chronological panorama that will be analysed below. This panorama then appears in recent studies, such as Rovai 2012b. The aim of the present corpus-based study is to apply the approach more systematically and to a larger set of data than has been possible earlier. This is likely to make even the chronological estimations more reliable.

I next present preliminarily those linguistic constraints that previous research has considered to be the most crucial for the morphological realignment of Late Latin: a) control of the subject over the

---

[111] Pensado 1986, 274; Tekavčić 1972, 35; Väänänen 1981, 117.

[112] This method was dictated by necessity given that no parsed corpora of Late Latin existed before LLCT. In contrast, several important studies with traditional small example collections have been conducted on issues relative to the semantically-based alignment of Latin: Cennamo (1999b) shows how the pleonastic reflexives *se* and *sibi* reflect intransitivity split in Late Latin; Lazzeroni (2002) examines the semantically and syntactically motivated gender selection between thematic masculine/neuter doublets in the Indo-European framework; Rovai (2007) develops Lazzeroni's ideas as far as Latin is concerned; Rovai (2012a) shows that the reanalysis of feminine singulars as thematic neuter plurals in Latin is also semantically motivated; Fabrizio (2010) and Fabrizio (in press) discuss semantically-oriented syntax of subject infinitives in Latin; and Clary (2014) shows that even the case marking of cognate objects in Latin is semantically motivated.

[113] Herman 1998, 18, 21–22; Zamboni 1998b, 664–665; Zamboni 2000a, 93. For a chronology of Latin in Gaul, see Banniard 1992, 515–519.

verbal event, b) semantic type of verb, c) syntactic/pragmatic factors, such as syntactic construction and topicalisation. Because the realignment has been viewed mainly as semantically-motivated, the most important constraints seem to be semantic by nature. These are also the ones that I shall mainly concentrate on in the following because they are rather complicated to define and require, thus, a more detailed breakdown than the syntactic factors that will be discussed more fully in chapter 5.

Certain theoretical considerations, originally arisen in the field of language typology, have appeared to be helpful in explaining the variation and change related to Late Latin case use. Over the last twenty-five years, several typologically-oriented studies have suggested that the supposed binary (or ternary) case system of Late Latin was likely to be partly ergative/absolutive or active/inactive aligned.[114] Ergative/absolutive and active/inactive, as well as nominative/accusative, are different forms of morphosyntactic alignments. I shall explain below what the mentioned alignment types mean.

To put it simply, an alignment describes how the arguments of transitive verbs and those of intransitive verbs differ from each other. The distinction can be morphological or syntactic: syntactic alignment is based, for example, on word order while morphological alignment functions through verbal agreement and/or case marking, i.e. through morphological markers of grammatical relations.[115] Case marking is what will be examined in this study. According to Dixon (1994), there are three universal syntactic-semantic primitives, i.e. nuclear or core arguments, that apply to verbal clauses in all languages: A, S, and O. Transitive verbs have two nuclear arguments: A, which is covered by the syntactic function of the subject, and O, which is covered by the syntactic function of the object. Intransitive verbs have only one nuclear argument, S. What the alignments are fundamentally about is aligning S with either A or O, or both. These cross-linguistically attested nuclear arguments or syntactic-semantic primitives were originally coined by Dixon (1968).[116]

---

[114] Plank 1985; La Fauci 1988, 1991, 1997; Zamboni 1998a, 2000a; Cennamo 2001b, 2001c, 2009; Rovai 2005, 2010, 2012b; Ledgeway 2011, 462. For a typological comparison, see especially Rovai 2012b, 92–93, 111. For Cennamo, I will cite in this study mostly Cennamo 2009 (Argument structure and alignment variations and changes in Late Latin), a thorough overview that summarises the results of her previous articles. Cennamo 2001b is an early overview on the phenomena related to the extended accusative, while Cennamo 1998, Cennamo 1999a, Cennamo 1999b, Cennamo 2000, Cennamo 2001a, Cennamo 2001c, and Cennamo 2008 concentrate on the reorganisation of the voice system concomitant to the case realignment. In Cennamo 2011, the author discusses the role of the impersonal constructions in the alignment change. She has also examined the alignment change from the viewpoint of Italian dialectology in many articles, such as Cennamo 2001d, 2006, 2010, and 2014.

[115] Comrie 1989, 111, 125; Donohue 2008, 25–28; Rovai 2012b, 12. Donohue (2008, 27) introduces the term 'primary morphosyntactic device' to cover the use of head-marking (agreement), dependent-marking (case marking), and position (word order) as tools of determining alignment.

[116] Dixon 1968; Dixon 1994, 6–8; Rovai 2010, 317–318; Rovai 2012b, 17; Siewierska 1988, 49. Label P is sometimes used in place of O (e.g. Comrie 1989), while Lazard (1991, 3–58) utilises X, Y, and Z. La Fauci (1997, 9–10) defines

The ultimate advantage of the nuclear arguments is that with them it is possible to formalise the ways in which semantic roles and syntactic functions are mapped together in a language. Nuclear argument A corresponds broadly speaking to the semantic macrorole actor, while O matches the semantic macrorole undergoer; S is located somewhere between them on the continuum that I shall call the 'continuum of control' (Figure 3.1.). The macroroles are generalisations of specific semantic roles: the actor is an umbrella term for agent-like semantic roles and the undergoer lumps together the patient-like semantic roles.[117] The semantic features that contrast the actor with the undergoer (and determines whether a subject aligns with the O end or with the A end of the continuum) can be defined in several ways. I have chosen 'control' as the first semantic gradient in Figure 3.1. because the findings of Rovai and Cennamo prove that control plays an important role in the case marking of Late Latin.[118] Degree of control is an inherent feature of the subject and measures the control exercised by the subject over the verbal event. It reflects the degree of the primary responsibility of the subject over a given verbal process and can be thought to subsume comprehensively other important semantic features of subjects, such as animacy, referentiality, and individuation. Due to its centrality, control will be discussed on many occasions in this work: for example, in section 4.3., where it is tightly related to the central theme of that section, i.e. transitivity.

Figure 3.1. Continuum of control and semantic (macro)roles.

```
[– Control] <-------------------------------------------------------------------------------> [+ Control]
---- Undergoer ------------------------------------> <--------------------------------------- Actor ----
-------- O ------------------------------------------ S --------------------------------------- A --------
   Patient        Theme        Location      Instrument    Experiencer    Efficient      Agent
```

---
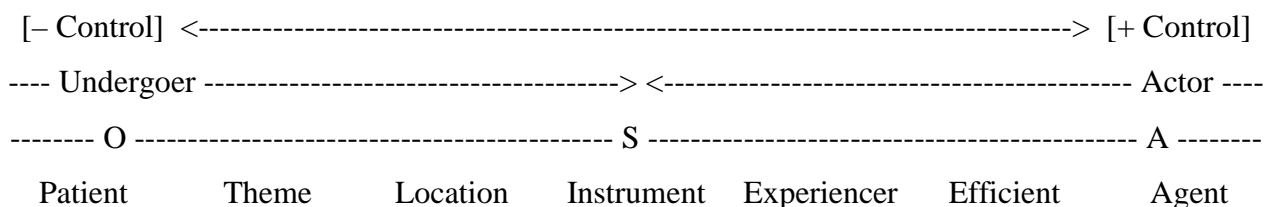
the nominal elements of the verbal nucleus by way of structural traits [±unary] and [±binary] but ends up with an essentially similar outcome as Dixon. In La Fauci, S refers simply to subject and O to object. The O can occur alone with trait [+unary] or together with S with trait [+binary] (and [-unary]). S can also occur alone; then it is [-binary]. The nuclear structures resulting from the combinations of these traits are called unaccusative, transitive, and unergative, respectively. Thus, S does not have the same meaning in La Fauci's terminology as in Dixon. The predicating lexical items, i.e. verbs, are then linked to this frame according to their valency.

[117] Van Valin 1990, 226; Van Valin 2005, 61–63; Rovai 2010, 317–318; Rovai 2012b, 23–25; Foley & Van Valin 1984, 29–36. On semantic roles in general, see e.g. Van Valin 2005, 53–58.

[118] Comrie 1989, 59–62; Cennamo 2009, 327–328; Rovai 2012b, 31, 52, 93, 106; Lakoff 1977, 244, 248–253; Lehmann 1991, 211–217. Lehmann (1991) considers control the decisive feature of distinguishing between the 'active' and 'inactive' participants of an intransitivity split. Consequently, he regards the 'active' and 'inactive' as grammatical relations that neutralise some important semantic distinctions just like any grammatical relation does. For other contrasting semantic features, see Mithun 1991; Donohue 2008; Sorace 2000. For the idea of gradiency in split intransitivity, see Sorace 2000, 861–863.

The continuum of control is completely based on generalisations of various types, i.e. on prototypicality. It suggests that prototypical subjects are actors and prototypical objects are undergoers. Prototypical subjects have considerably higher control over the verbal event than prototypical objects. Indeed, according to Van Valin (2005), the generalised semantic roles, actor and undergoer, can be assimilated to "the two primary arguments of a transitive predication, either one of which may be the single argument of an intransitive verb". This is how the linkage between the semantic layer and the syntactic functions, subject and object, takes place. However, as there is no one-to-one correspondence between the two macroroles and the three nuclear arguments, their connection is necessarily asymmetrical. Alignment is the strategy by which the non-existent correspondence between S and a specific macrorole is forced into existence.[119]

Cross-linguistically, languages tend to treat two of the three nuclear arguments (A, S, and O) in one way and the third one in another way. This is reflected as different alignments. In Figure 3.2., I show what this means for case marking. The nuclear arguments that are enclosed in an oval share the same case marking. In the nominative/accusative alignment, prevalent in Classical Latin, the subjects of transitive and intransitive verbs (A and S) are opposed to the object of the transitive verb (O). A and S are coded by the nominative case and O by the accusative case, hence the name of the alignment. In ergative/absolutive alignment, the subjects of transitive verbs (A) are opposed to all the other nuclear arguments (S and O): A is encoded with a case that is conventionally called ergative and the other nuclear arguments with a case called absolutive. The criterion of case marking in the ergative/absolutive alignment is syntactic, as it is also with the nominative/accusative alignment. The nominative/accusative alignment dominates European languages but languages with ergative/absolutive features are not cross-linguistically infrequent. They are most abundantly encountered in the Caucasus and in Oceania.[120]

Figure 3.2. The major alignment types.

Intransitive clause						S				S_A	S_O				S	
Transitive clause				A			O		A			O		A		O
				Nominative/accusative					Active/inactive					Ergative/absolutive		
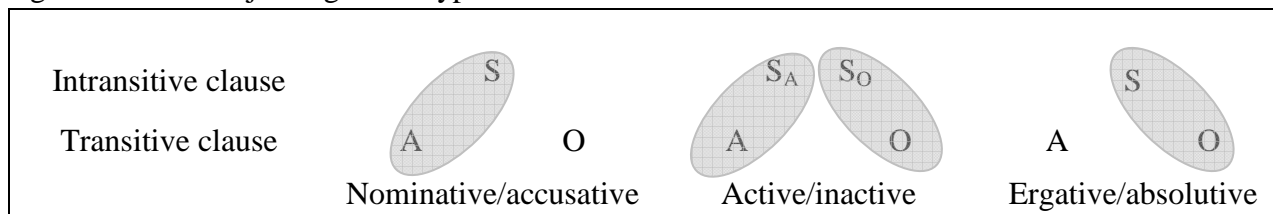
[119] Van Valin 1990, 226–227; Rovai 2010, 318; Van Valin 2005, 60–61; Sorace 2000, 874–877.
[120] Comrie 1989, 125–126; Dixon 1994, 8–11; Rovai 2012b, 35–50; Rovai 2010, 319. Even tripartite systems exist but they are marginal in comparison with other alignments (Rovai 2012b, 39–40).

On the other hand, the boundary between different alignments is not clear-cut. Several languages have some nominative/accusative and some ergative/absolutive characteristics or, what is of utmost importance for this study, they may split the nuclear argument S into two semantically-based arguments, S_A and S_O. The former represents the semantically active intransitive actor argument and formally aligns with A, while the latter is the semantically inactive undergoer argument and formally aligns with O, hence the name active/inactive alignment. This kind of 'split intransitivity' differs radically from the already discussed organisations in that the distinction between nuclear arguments is motivated on semantic, not on syntactic, grounds. Languages with a clear semantically-based argument marking are 'exotic' from the European point of view: they include, among others, Lakota, Mohawk, Guaraní, and Laz. On the other hand, the modern Romance languages show several phenomena motivated by split intransitivity, such as the auxiliary selection, *ne*-cliticisation, and absolute participle use.[121] The following schematic clauses exemplify how split intransitivity functions in Lakota, a Siouan language spoken in North and South Dakota.[122]

(1) transitive clause with A prefix *wa-*: *waktékte* "I will kill him"

(2) transitive clause with O prefix *ma-*: *maktékte* "he will kill me"

(3) intransitive clause with A prefix *wa-*, i.e. S_A: *wahí* "I came"

(4) intransitive clause with O prefix *ma-*, i.e. S_O : *maxwá* "I am sleepy"

Clause (3) shows that an agentive S_A subject of intransitive clause is marked in the same way as that of transitive clause (1). Clause (4) reveals that the undergoer-role subject (S_O) of the intransitive clause receives the same marking as the direct object of the transitive clause (2).

Wichmann and Rovai argue plausibly that a system with split intransitivity should be called 'semantic alignment' or 'semantically-based alignment' instead of 'active/inactive alignment' (or 'active/stative alignment', which is also used). The latter names suggest that it is the aktionsart of the verb that makes the distinction between the arguments but, in fact, crosslinguistic studies have shown that languages may condition the intransitivity split either by semantic properties of the verb,

---

[121] Van Valin 1990, 231–240; Dixon 1994, 70–110; Rovai 2012b, 47–52; Ledgeway 2012, 312–314, 321–327, 337–340; Zamboni 2000a, 103; Cennamo 1999b, 105–113; Sorace 2000.
[122] Mithun 1991, 514; see also Lehmann 1991, 212; Rovai 2012b, 57–58.

by semantic properties of the subject, or by both.[123] In this study, I shall use mainly the term 'semantically-based alignment'. It is, however, not clear to what extent the semantically-based alignment that has been postulated for Late Latin in earlier research is affected by the verbal semantics (mainly aktionsart). Rovai argues strongly for the view that the semantic features of the subject NP, such as animacy, individuation, and referentiality, which together constitute the above-mentioned 'control', are decisive in Latin. Cennamo is open to construction-motivated interpretations as well.[124] This study aims to clarify this aspect by discussing both the subject-inherent semantic properties (by way of animacy/referentiality distinction) and the verbal-event-inherent semantic properties (by way of a fourfold construction type distinction A, S_A, S_O, and S_O of the passive).

To summarise, active/inactive alignment or S split or semantically-based alignment is a semantically-based system of grammatical relations where the unagentive, intransitive undergoer subjects (S_O) are aligned, in terms of their grammatical properties, such as case marking, with the objects (O) of transitive constructions, whereas the agentive, intransitive actor subjects (S_A) align with the subjects of transitive verbs (A). The opposition is, thus, between the semantic distinction of the macroroles actor and undergoer. To avoid confusion, the following terminological conventions should be noted: the verbs with S_O subjects are often called unaccusative (5) and the verbs with S_A subjects unergative verbs (6). The terms 'unaccusative' and 'unergative' originate from the theoretical frameworks of generative and relational grammars, whereas the notions S_O and S_A are usually preferred by linguists working on functional theories.[125] Note that the terms 'unaccusative' and 'unergative' do not refer to the above-discussed alignment types, but describe the behaviour (transitivity, aktionsart) of the verb in a given clause, instead. In this study, these terms are utilised only when the focus is explicitly on the verb. In order to avoid duplicate terminology, I shall usually quote the subjects of unergative constructions as 'S_A subjects', the subjects of unaccusative constructions as 'S_O subjects', and the subjects of transitive constructions (7) as 'A subjects'.

---

[123] Mithun 1991, 510; Wichmann 2008, 4; Rovai 2012b, 47–48; Dixon 1994, 70–73, 83–85. Dixon's term, i.e. 'S split', is inadequate in its emphasis on the category S, which is relevant only to syntactic alignments. Rovai (2012b, 59–66) discusses languages where it is the aktionsart that causes the intransitivity split.

[124] Rovai 2012b, 47, 106; Dixon 1994, 71; Cennamo 1999b, 108; Cennamo 2009, 307; Cennamo 2011, 184–185; Pieroni 1999; Van Valin 1990; Valentini 2012, 91–93. As for the role of verbal semantics, Cennamo (1999) shows that the semantic properties of the verb affect the *se*/*sibi* alternation of the Late Latin pleonastic reflexives.

[125] Perlmutter 1978; Van Valin 1990, 222; Sorace 2000, 879–880; Cennamo 1999b, 107–108. In the terminology of Harris & Campbell, the unaccusative and unergative verbs are 'inactive' and 'active', respectively (Harris & Campbell 1995, 241). Cennamo 1999b contains a thorough discussion on the theoretical roots of the research on split intransitivity.

(5) unaccusative verb: CDL 273 (AD 772) *casa ubi ipsi genitor tuus* (S$_O$) *antea residet*

"the house where your father (S$_O$) once resided"

(6) unergative verb: CDL 248 (AD 770) *ut sacerdos* (S$_A$) [--] *a Domino deprecare deueas* [= *debeat*]

"that the priest (S$_A$) [--] should pray to God"

(7) transitive verb: CDL 90 (AD 747) *medietate* (O) *de ipsa terrola possedeat ipsa sancta Dei uertute* (A)

"[that] the holy church (A) of God possesses half (O) of that plot"

Though the morphosyntactic alignments were first discussed by scholars as early as over a hundred years ago, split phenomena received attention only later.[126] Restricted sub-systems of semantically-based alignment may appear within a dominant nominative/accusative alignment, where the split is usually determined by the agentivity, referentiality, control, and/or animacy of the subject.[127] As regards Late Latin, Rovai suggests that the semantic split shows between low-animacy and high-animacy nouns. According to Rovai, low-animacy subjects allow the extension of the accusative into unaccusative and passive constructions. They arguably preserve the nominative marking only with transitive and unergative constructions. High-animacy subjects still pattern with the nominative in all the subject types, just as in Classical Latin (see Table 3.1. and Table 3.2.). Animate nouns, and especially the human ones, involve a high control over the verbal process and, therefore, tend to occur as A or S$_A$ subjects. Inanimate nouns, instead, occur typically as direct objects (O) or, when subjectised, as S$_O$ subjects. The latter is the context in which the semantic realignment is supposed to become 'visible' when the inanimate S$_O$ subjects adopt the accusative case, originally reserved only for O.[128]

---

[126] Uhlenbeck (1901–1902) did not yet use word 'ergativity' in his pioneering treatise *Agens und Patiens im Kasussystem der indogermanischen Sprachen* on the nominative/accusative and ergative/absolutive systems. The term seems to surface in a description of the Caucasian Rutul language by Dirr in 1912. Sapir is probably the first to delineate systematically the basic alignment types including the active/inactive system in 1917. Plank (1979) and Lehmann (1985) introduced the study of ergativity into Latin linguistics and it seems to have been only in the late 1980's that La Fauci really began to analyse certain phenomena of Latin morphosyntax in terms of split intransitivity. For the history of alignment studies, see Wichmann 2008, 5–7; Itkonen 1974, 379; Rovai 2010, 320; La Fauci 1997, 12; Dixon 1994, 2–3.

[127] Donohue 2008; Rovai 2012b, 92–93. Although languages usually have one primary alignment, they very often contain differently aligned subsystems. For a cross-linguistic overview on the parameters conditioning the semantically-based splits, see Mithun 1991, 510–512; Donohue 2008, 39–73; Van Valin 1990, 251–252.

[128] Rovai 2012b, 110–112; Siewierska 1988, 49; Zamboni 2000a, 93; Zamboni 1998a, 137; Pensado 1986, 273; Silverstein 1976, 175, 198. For the Table 3.1. and Table 3.2., see Faraoni 2014b, 2.

Table 3.1. Behaviour of low-animacy nouns in Late Latin (= semantically-based alignment).

Construction	Transitive	Unergative	Unaccusative	Passive	Transitive
Nuclear argument	subject (A)	subject (S_A)	subject (S_O)	subject (S_O)	object (O)
Morphological marking	nominative		accusative		

Table 3.2. Behaviour of high-animacy nouns in Late Latin (= nominative/accusative alignment of all nouns in Classical Latin).

Construction	Transitive	Unergative	Unaccusative	Passive	Transitive
Nuclear argument	subject (A)	subject (S_A)	subject (S_O)	subject (S_O)	object (O)
Morphological marking	nominative				accusative

In language typology, it is helpful to present animacy and referentiality of an NP as a hierarchy, such as the one in (8). The higher in the hierarchy a noun is, the higher its animacy/referentiality degree and, consequently, the higher its control over the verbal process in which it is involved. This kind of noun is more likely to occur as the subject of a transitive construction and, thus, to assume the nominative case typical of the subject function.[129]

(8)   first/second person pronoun < third person pronoun < proper name < human common noun < non-human animate common noun < inanimate common noun

In this study, a simplified version of animacy/referentiality hierarchy is used. It will be defined in section 4.1.1. where animacy is further discussed from the viewpoint of the LLCT data.

It is intuitive that animacy/referentiality plays a central role in the diachronic evolution of alignments. The intermediate stage of the transition from nominative/accusative to ergative/absolutive alignment (or vice versa) is typically characterised by a semantically-based encoding of the arguments, just as it seems to have been in the case of Late Latin. Of course, in the case of Latin, the realignment that was taking place in the Late Latin argument structure probably ended up neutralising the case contrast before reaching a full ergative/absolutive alignment. Nevertheless, this evolution path can be reconstructed for some ergative/absolutive languages.[130]

In the transition from the mainly nominative/accusative alignment towards the ergative/absolutive alignment (see Table 3.3.), the marking of O (accusative case) does not extend immediately to all

---

[129] Croft 2003, 129–130; Siewierska 1988, 30, 49; Rovai 2012b, 90–93, 101; Dixon 1994, 83–85.
[130] Dixon 1994, 187–188.

the S subjects of intransitive constructions, but first to those that share most characteristics of the macrorole undergoer with O, i.e. the S_O subjects. The nouns that correspond to this definition usually occupy the lowest positions in the animacy hierarchies. The animate and, especially, human-referent nouns, instead, share inherently several characteristics of the macrorole actor and tend to occur as A/S_A arguments.[131] Thus, the animacy does not determine the case of the subject *per se*, but is one of the intertwined factors that together define the amount of control of the subject over the verbal event.

Table 3.3. Ergative/absolutive alignment.

construction	transitive	unergative	unaccusative	passive	transitive
nuclear argument	subject (A)	subject (S_A)	subject (S_O)	subject (S_O)	object (O)
morphological marking	nominative	accusative			

As already said, my intention is to observe the alignment in LLCT from two different semantic viewpoints: the inherent semantic properties of the subject and the semantic properties of the verbal event. These two viewpoints are connected by way of the subject selection constraints of the verbs. As was just stated, low-transitivity construction types usually involve subject arguments that are non-active and inert by nature and, thus, of low control. The low-control subjects of this kind are usually inanimate while the high-control subjects that favour high-transitivity constructions are usually animate or personal. Consequently, the intransitive constructions seem to attract the extended accusative.[132] This said, it is not necessarily quintessential in semantically-based alignment whether the subject belongs to a given construction type. What matters is whether the subject is an actor or an undergoer. Nonetheless, the above-mentioned fourfold construction type classification is a sound technical tool for assessing the actor/undergoer status of the subject, as will be shown in section 4.1.2.

Thus, construction types can be used as a metrics for assessing, from the standpoint of the verb, the same phenomenon that the animacy degree measures, i.e. the control of the subject over a given verbal process. In this study, I refer by construction types to the fourfold classification of transitive constructions (with A subjects), unergative constructions (with S_A subjects), unaccusative constructions (with S_O subjects), and passive constructions (with S_O subjects) (see section 4.1.2.). This classification introduces into the study the concept of transitivity that will be discussed in detail in sec-

---

[131] Rovai 2012b, 92–93; Plank 1985, 289–293.
[132] Cennamo 2009, 312, 314.

tion 4.3. The classification and its ideal connection to the case marking patterns of the different alignment types can be seen in the above three tables. It is obvious that the prototypically inactive, low-control S_O subjects of unaccusative verbs tend to adopt the case form of O (accusative) instead of the case form of A (nominative), which is reserved for transitive and unergative clauses in a semantically-based alignment.[133]

On this occasion, something must be said about the accusative as the functionally unmarked case in (Late) Latin because this assumption serves as the premise of all the syntactic analysis conducted in chapter 5. The discussion will be deepened in section 6.2. where it can be revised in the light of the statistical evidence of LLCT. In Latin as well as in many Indo-European languages, both the nominative and the accusative case are morphologically marked, although cross-linguistically the nominative is expected to have no marker in a nominative/accusative alignment.[134] Yet, the dominant default or minimum-effort form can be defined as unmarked, whereas the other, secondary one, is marked.

The nominative was quite obviously the unmarked default case of Classical Latin. Some scholars have suggested plausibly that the accusative underwent a markedness shift and became the unmarked pole of the nominative/accusative contrast sometime in Late Latin. This corresponds to the picture that arises from LLCT in chapter 4 and chapter 5. The idea of markedness-driven change also matches well the evidence seen in section 3.2.: the accusative, which is turning into the unmarked case, *cas régime* as Pensado calls it, is bound to expand at the expense of the marked case (nominative, *cas sujet*). It extends first to the subjects of contexts where the (low) animacy degree of the subject does not distinguish the subject from the object (e.g. non-prototypical subject *claudicatio* 'limp' in (20) in section 3.2.).[135]

So far only semantic factors have been discussed. However, certain syntactic and/or pragmatic constraints ought to be taken into account as well, and they will be discussed in detail in chapter 5. In chapter 6, I shall examine whether and to what extent these syntactic factors interact with the above-discussed semantic factors in the data of LLCT.

---

[133] Rovai 2012b, 47, 112; Cennamo 2009, 314.
[134] Pensado 1986, 271; Lehmann 1985, 247.
[135] Cennamo 2009, 308–309, 327; Collinge 1978, 623–624; Lehmann 1985, 246–247; Vincent 1982, 88–89; Rovai 2012b, 107, 110; Smith 2011, 277–278; La Fauci 2001, 21–22; La Fauci 1988, 55; Pieroni 1999, 120; Galdi 2013, 77; for critical views, see Adams 2013, 254–256; Benucci 2004.

In contrast to Rovai and Plank (1985), Cennamo (2009) does not emphasise the re-semanticisation of the nominative/accusative opposition, but describes the change, instead, as an active/inactive realignment of certain coding properties, such as case marking and agreement. However, Cennamo does consider animacy and control crucial concepts because most of the attested accusative subjects are confined initially to inactive arguments. According to Cennamo, these semantic parameters conflate gradually with the syntactic and pragmatic features that involve the type of clause, the degree of syntactic cohesion, and the integration of an extra-syntactic accusative topic into the predicate nucleus of the clause, like *herbam* in (9). The same is supposed to apply also to examples (15) and (16) in the next section.[136]

(9) <u>*herbam*</u>, *quae Gallice dicitur blutthagio*, *nascitur locis umidis*, *eam teres* (*Marcell*. 9.132)
'the <u>herb</u> that is called blutthagio in Gaulish grows in humid places, grind it'

Here the inanimate noun *herbam* 'herb' is the sentence-initial topic. It is also the (S_O) subject of the verb *nascitur* 'grows' and its semantic macrorole is clearly an undergoer. The final *-m* of the 1st declension is, of course, not very probative, but here the accusative form is well motivated not only semantically, but also pragmatically.[137] Once the topic is related to the sentence-initial position, the correlation that is found between topicalised subjects and their case form also implies a correlation between syntax, i.e. word order, and subject case selection. To conclude, I must also mention the possible extra-linguistic constraints that are likely to influence the subject case selection of LLCT. They will be discussed where relevant in most of the chapters of this study.

3.2. Extended accusative and the Late Latin data

This and the following section concentrate on the evidence and on how it has been interpreted in previous research. Section 3.2. discusses Latin data from the late Empire to the early Middle Ages. The analysis will be carried out by reviewing the main results of Cennamo's (2009) study. Section 3.3. introduces the other type of data, i.e. those that can be retrieved from both modern and historical Romance languages.

---

[136] Cennamo 2009, 326–328; Cennamo 1999b; Cennamo 2011, 172, 179–180; Herman 1987, 103; Rovai 2012b, 116–118.
[137] Cennamo 2009, 329.

Ledgeway (2012) suggests that the spread of the semantically-based alignment to the nominal declension was instigated by the active/inactive orientation of certain areas of the Latin verb system. Indeed, the *perfectum* paradigms of the non-deponent verbs of Classical Latin were active/inactive oriented with A and S_A marked by a synthetic paradigm (e.g. *feci* "I made", *cucurri* "I ran") and S_O alone by a periphrastic paradigm (e.g. *omnia per ipsum facta sunt* "all things are made by him"). In Late Latin, the active/inactive orientation expanded with the genesis of the perfective auxiliary construction (*habere* vs. *esse*) and the concomitant patterns of participle agreement. This orientation is then supposed to have expanded even to the nominal system.[138] Other possible points of departure or contributory developments for the semantically-based alignment include the temporary loss of grammatical dimension of voice in Late Latin and the reanalysis of impersonal constructions.[139] These will not be discussed in this study. The following passages will investigate how the actual attestations of extended accusatives in Late Latin texts match the above assumptions.

There are attestations of accusative subjects in place of nominative subjects at least from the 2nd century AD onwards. These so-called 'extended accusatives' are likely to lead the way to the generalisation of the accusative form as the principal form underlying Romance nouns. The concept 'extended accusative' was first introduced by Moravcsik (1978) in her study on ergative-based subsystems in languages that represent mainly nominative/accusative alignment.[140] Later, La Fauci and Cennamo introduced the concept into the study of grammatical relations in Late Latin. Rovai (2005) was the first to gather a systematic corpus of 134 cases of extended accusative from eight non-standard texts from between 4th and 9th centuries: 85–90% of the accusatives turned out to be S_O subjects of intransitive or passive predicates.[141] In the following, I present typical contexts of extended accusatives according to a chronology proposed by Cennamo.[142] The example phrases are taken mainly from Rovai (2005) and Cennamo (2009).

Cennamo presents the relative chronology of the semantic alignment as a sequence of attestations of the extended accusative with different verb types. According to Cennamo, the point of departure is nominal clauses, such as commands, exclamations (10), lists (11), various types of topicalisations

---

[138] Ledgeway 2012, 314–318; Cennamo 2008, 116–121.
[139] Cennamo 2009, 307, 312, 334–335; Cennamo 2011, 177–179, 180–185.
[140] Moravcsik 1978, 241.
[141] Rovai 2005, 77–87. Rovai excludes from his corpus the attracted subjects of relative clauses, the so-called recipe accusatives, and the accusatives occurring in impersonal constructions. See also Valentini 2012 for some important additions to Rovai's corpus.
[142] Rovai sketches a similar chronology in Rovai 2005, 70–71. Earlier attempts towards an absolute chronology of the extension of the accusative are found in Herman 1987 and Herman 1997.

((11) to (13)), presentative constructions (13), and impersonal constructions with an expressed argument (14).[143] These are attested in Latin from the earliest texts onwards. The common factor is the loose or lacking linkage to any verb.

(10) *me infelicem et scelestam* [--] (Plaut. *cist*. 685; 3rd/2nd c. BC)

"oh me unhappy and cursed [--]"

(11) *Puteolos, Antium, Tegeano, Pompeios, hae sunt verae coloniae* (*CIL* IV 3525; 1st c. AD)[144]

"Puteoli, Antium, Tegeanum, Pompei, these are real colonies"

(12) *portionem ad eos qui sanguinem meient* (*Mul. Chir*. 822; 4th c. AD)

"a drink for those who pass blood"

(13) *sed eccum Amphitruonem, advenit* (Plaut. *Amph*. 1005; 3rd/2nd c. BC)[145]

"but here comes Amphitruo"

(14) *vitam vivitur* (Enn. *trag*. 214; 3rd/2nd c. BC)

"one lives life"

The first known occurrences of alleged accusative subjects that seem to be more closely attached to verbs are found in Northern African execration tablets from the 2nd to 3rd century AD. They occur with intransitive verbs that denote change of state or location. Unexpectedly, all the accusative arguments are animate, many of them even personal names. On the other hand, a comprehensive scrutiny of all the execrations reveals that the connection between the accusative argument and the verb is not particularly close. The structures can be regarded as exclamations and, consequently, the arguments in the accusative as topics.[146] They can all be interpreted as S_O arguments.

---

[143] Cennamo 2001b, 6–12; Cennamo 2009, 315–329; Cennamo 2011, 177–179, 184–185; Pieroni 1999, 119–121, 126–127; Gerola 1949–1950, 209–219; cf. Benucci 2004, 12.

[144] *Tegeano* is likely to be *Tegeano‹s›*.

[145] Note that *eccum*, which marks the topic, is crystallised into the accusative form. See TLL V:2, 23–25; Väänänen 1981, 123.

[146] Audollent 1967, 387, 392; Cennamo 2009, 316; Herman 1987, 103–105; Sornicola 2008, 234; Sornicola 2011, 33–35; Herman 1997, 25. The breakdowns of agreement in number, as in (17), also suggest that the accusative arguments are not closely integrated into the sentence (Adams 2013, 227–228, 250). On the other hand, it has even been questioned whether the forms in *-u* are really to be interpreted as accusatives in these rather early texts. They could also be nominatives without *-s* (or with a mobile *-s*, as Herman puts it; see Herman 1987, 103–104; Herman 1997, 24–25). For the final *-s*, see section 2.3.3.

(15) *Epafu, Victore cadant, Lydeu cadat* (*def. tab.* 278A.3)

"let <u>Epafus</u>, <u>Victor</u> fall, let <u>Lydeus</u> fall"

(16) *Supestite russei servu Reguli cadat, vertat* (*def. tab.* 283A.4)

 "let <u>Superstes</u>, [member] of the red faction, Regulus' <u>servant</u>, fall, turn"

(17) *Delusore cadant* (*def. tab.* 278A.9)

"let <u>Delusor</u> fall"

From the 4th century onwards, extended accusatives can be found in texts that come from several other areas of the Empire. Now, the accusatives represent a larger variety of intransitive patterns ranging from fientives and anticausatives to passives. The subjects or their predicative complements, as in (19), mainly inanimate, are still exclusively S_O arguments.[147] Many – but not all – of the earliest occurrences of S_O subjects are restricted to abstract deverbal nouns, e.g. (18) and (20), which are clearly at the bottom end of the animacy hierarchy – the place where the extension of the accusative assumedly first began. Another major group is inanimate common nouns, as in (21), (22), (23), and (24). The following texts date back to the 4th to 6th centuries. This is the stage that can be viewed as reflecting most genuinely an active/inactive realignment of the case marking system in the spoken language: the S_O arguments are supposed to have aligned with O while the S_A arguments still aligned with A.[148]

(18) equative: *ustionem necessaria res est* (*Mul. Chir.* 153; 4th c. AD)

"<u>cauterisation</u> is necessary"

(19) equative: *ut crudastros sint* (*Anth.* 11.21; early 6th c. AD)

"so that they are <u>rawish</u>"

(20) change of state: *nascitur ei genuorum contractionem et claudicationem* (*Mul. Chir.* 516; 4th c. AD)

"its knees are developing a <u>contraction</u> and <u>limp</u>"

(21) change of location: *ut sanguinem exeat copiosum* (*Mul. Chir.* 618; 4th c. AD)

"so that the underline blood may run out abundantly"

(22) anticausative: *multos languores sanantur in ipsis locis* (*Ant. Plac.* 9; late 6th c. AD)

"several illnesses heal in these places"

(23) impersonal: *cum factum fuerit missam* (*Per. Eg.* 32.2; late 4th c. AD)

"when the mass is over"

(24) passive: *omnes cibos comedantur* (*Anth.* 1; early 6th c. AD)

"all kinds of food should be eaten"

The first reliable attestations of extended accusative with agentive/dynamic motion activities come only from texts from the 8th century. The subject is S_A and animate by definition: the inanimate subjects of motion verbs are S_O, as in (21) above. The S_A subjects no longer imply a pure active/inactive state of case alignment, but rather indicate an incipient transition to an ergative/absolutive stage of the coding system or, perhaps more plausibly, the beginnings of the final neutralisation of case opposition.[149] Indeed, the neutralisation is likely to have already begun by that time because the accusative has reached some of the A subjects of transitive events as well (see below). Passage (25) seems to be the earliest S_A example; in LLCT they total 87. I quote here one controversial case from AD 757 (26) and another, unambiguous, case from AD 812 (27).

(25) *si sequenter ipsum currit* (*Lex Alamannorum* 94.3, cod. A; c. AD 720)[150]

"if he runs away subsequently"

(26) *numquam me, heredis meis contra hanc offerta seo dotalia mea ire quandoque presumat* (CDL 125, AD 757)[151]

"let neither me nor my heirs dare to go against this my donation or endowment"

---

[149] Cennamo 2009, 319; Ledgeway 2012, 331; Zamboni 2000a, 114–115; Cennamo 2001b, 8. Ledgeway quotes the passage *ipsos filios sedeant inienuos* (Spain, Cartulario S. Vicente 42.16; AD 781–1200) as an example of S_A subject. However, *sedere* is here likely to be equivalent to modern Spanish *estar*, the meaning thus being "the children were hungry" and not "the(se) children were seated", as Ledgeway translates it. The subject is then S_O.

[150] *Lex Alamannorum* (manuscript A) is dated nowadays to the 8th c. (c. 730), not to the 6th/7th c. (Schmidt-Wiegand 2001, 201–202).

[151] *Me, heredis meis* may be a contamination of two structures.

(27) *quam filiis qua <u>mulierem</u> tua vadam* [= *vadant*] *ubi volueris* [= *voluerint*] (CDT 89, AD 812)

"both your sons and your <u>wife</u> may go where they wish"

According to the theory, the last stage of the realignment before the collapse of the case opposition was the gradual extension of the accusative to some transitive patterns. The earliest alleged transitive attestations date, however, back to Egeria (late 4th century AD) (28) and to 5th-century inscriptions of Africa (29).[152] Especially the early examples are scarce and some of them rather dubious, e.g. (29) where *filios et nepotes* is a recurring crystallised phrase that is likely to have been induced here erroneously, or (30), in which the juxtaposition of *haec* and the 1st declension *causam* seems to imply just a mere spelling error in *causam* (the same may be true for (9) and (23) as well).[153] Undeniable occurrences, such as (31), become more common only in the 8th century.

Many of the sentences are also rather low in transitivity, i.e. semantically closer to intransitive constructions than to the prototypical examples of high transitivity. In (28), *fontem* 'fountain' is actually low in transitivity because the predicate *colorem mutare* 'to change colour' can be interpreted as a single (intransitive) lexical item.[154] Moreover, *fontem* is a non-prototypical affected agent that undergoes an action. In a similar vein, the general transitivity degree of the transitive constructions of LLCT (especially of those with an accusative subject) is particularly low, as will be demonstrated in section 4.4.1. On the other hand, these observations corroborate the theory: the extension of the accusative to transitive patterns is likely to have begun with the constructions with the lowest possible degree of transitivity and probably never reached the highest domains of transitivity before the collapse of the case contrast. Indeed, many of the most agentive nouns were transferred to the Romance in their nominative form (e.g. It. *uomo* < *homo* 'man'), as will be seen in the following section.

(28) *<u>fontem</u> vero ubi testa saniam radebat quater in anno colorem mutat* (*Egeria, excerpta* 13.1; 4th c. AD)

---

[152] Cennamo 2009, 324–325. Adams considers the evidence on the extension of accusative subjects to transitive verbs to be very weak (Adams 2013, 249). Rovai (2005, 81) also quotes six other rather early examples of accusative with transitive verbs from Mulomedicina and Apicius. I do not enclose them here because, to my mind, they are not completely reliable. For the moment, I also leave aside the feminine plural subjects in *-as* (Rovai 2005, 86–87).

[153] Several inscriptions with *filios et nepotes* are known from Africa (see Cennamo 2001, 9). As *faciat* in (30) approaches the sense of *fiat*, *causam* would be better interpreted as an S_O subject. See also Bonnet 1890, 522. For the critique of the 1st declension *-m* being used as evidence, see Adams 2013, 248.

[154] Rovai 2005, 63.

"indeed, the <u>fountain</u>, where he scraped the pus with a crock, changes its colour four times a year"

(29) <u>*filios*</u> *et* <u>*nepotes*</u> [--] *memoria posuerunt* (*CIL* VIII 7467; 5th c. AD)

"his <u>children</u> and <u>grandchildren</u> [--] set the tombstone"

(30) *ne faciat scandalum haec* <u>*causam*</u> (*historia Francorum* 5.18; 6th c. AD)

"so that this <u>business</u> would not become a scandal"

(31) *nec hoc quod* <u>*eos*</u> *quesierunt habere debent* (*Lex Curiensis* 2.9; 8th c. AD)

"and <u>they</u> should not get what they have asked"

By the extension of the accusative to transitive structures of this kind, the accusative occupied the nuclear arguments O and S as well as the lowest fringes of A. The remaining, rather marginal, high-animacy pocket of nominative marking dealt with such a small fraction of nouns that it was not sufficient to maintain the binary case opposition. The case system, whose functional load had already been gradually taken over by other grammatical means, such as word order and increasing preposition use, ceased to exist.[155]

3.3. Traces of case system in Romance

This chapter discusses the Romance traces of the bicasual system in some detail. The semantic cases genitive, dative, and ablative will not be analysed. As stated above, the backwards extrapolation from the lexicon of the Romance languages is the other main approach to the Late Latin case system. The Romance data offer a different viewpoint on the history of Latin case system, but this also involves different problems. Nowadays, even those researchers who are not convinced of the case realignment accept rather widely that most Romance nouns derive from Latin accusatives.[156] This seems to be particularly evident, for example, with Spanish plurals in *-as*, *-os*, *-es* or with the 3rd declension singular imparisyllabic forms, such as It. *parte* (< Lat. *partem*).[157]

---

[155] On the role of prepositions and word order, see Pensado 1986, 274; La Fauci 2001, 23; Schøsler 1984, 113–161; Tekavčić 1972, 35–49. On article and object clitics, see Vincent 1997, 149–150, 163; Tekavčić 1972, 125.

[156] For different views, see footnote 235 in section 4.2.3.

[157] It is also known that the Sardinian 2nd declension nouns in *-u* (*ca(ba)ddu* 'horse') derive from the accusative-based form (*cavallu(m)*) and not from the nominative-based form (*cavallus*). Sardinian retains the word-final /s/. (Zamboni 2000a, 107).

The previous chapter ended in the conclusion that the Latin case system in Italy ceased to exist probably at the point where the accusative had occupied all the nuclear arguments O and S, as well as the least agentive A arguments. There are hints of a similar process even in the better documented Old French and Old Occitan.[158] The neutralisation of the nominative/accusative contrast meant that the case system was degrammatised, i.e. it lost its function but retained (partly) the old formal markers. The initially semantically-motivated distribution of accusative forms and the few extant nominative forms with the most agentive nouns were transferred to Romance, but without any syntactic function. Thus, almost all the low-animacy nouns continued the accusative-based form as the only descendant (e.g. Lat. acc. *parte*(*m*) > It. *parte*) while several high-animacy nouns established the nominative-based form as their Romance continuator (e.g. Lat. nom. *homo*, *mulier*, *rex* > It. *uomo*, *moglie*, *re*).[159]

It is probable that, in the centuries immediately following the neutralisation, the distribution of the Romance nouns observed the semantically-motivated pattern that had been functional immediately before the neutralisation. In other words, the distribution followed more strictly the semantic agentivity criterion than it does nowadays after more than a thousand years of diastratic, diaphasic, and diatopic variation that has confused the original semantically-motivated distribution. Although this cannot be proved, there are lexicalised hints of levelling or differentiation having taken place in later times. For example, some high-animacy nouns retained both the nominative-based and the accusative-based forms that were consequently subject to varying dialectal distributions (e.g. inside Italy *sarto – sartore*, both 'tailor') or differential lexicalisation processes (e.g. *Jacques* 'James' – *ja*(*c*)*que* 'jerkin').[160]

Smith (2011) wonders why only a small subset of animate nouns should have survived in the nominative. Once the distribution is not systematic, the reason may be precisely a later levelling and/or differentiation. Some of the nominative-based forms may have been 'normalised' and subsumed by their accusative-based counterparts as exceptions or anomalies. It is, however, to be noticed that in Italy the nominative is not continued in the personal name vocabulary as conspicuously as in

[158] Smith 2011, 282; Pensado 1986, 278–293.
[159] Lausberg 1962, 56; Seidl 1995, 96–97; La Fauci 1997, 58; Zamboni 1998a, 133–135; Zamboni 2000a, 106–107; Zamboni 2002, 12; Rovai 2012b, 110; Faraoni 2014b, 1–2; Tekavčić 1972, 45–46; Löfstedt 1961, 215–217; Bastardas Parera 1953, 24; van Reenen & Schøsler 2000, 330–331.
[160] Smith (2011, 285–287) calls this the 'refunctionalisation' of the old syntactic opposition. See also Adams 2013, 203; Rovai 2005, 64.

French, where e.g. *Charles*, *Jacques*, and *Georges* are of nominative origin. The 3rd declension *Guido* is one of the surviving continuators of the nominative in modern Italian.[161]

The above-mentioned cases are not, however, the only Romance vestiges of the nominative/accusative distinction in Italian nouns. It is assumed that the Gallo-Romance varieties of Northern Italy retained case opposition longer than the central and southern varieties. Especially in Friuli, the *s*-plural and place names reveal vestiges of ancient binary contrast. However, there is no area without nominative residues in Italy. Even in southern Italian varieties, there are lexical remains of the Latin nominative form.[162] In northern Italo-Romance, the examples abound: e.g. ancient Venetian *nievo* (< *nepos*), *pastre/pastro* (< *pastor*), *-adro* (< *-ator*), and Padovan *San Salvàro* (< *Salvàtor*).[163] Here, animacy/agentivity seems to be the common denominator. Recent studies have shown that some agentive nouns of ancient Tuscan, such as *nepòte* or *segnòre*, where the pre-tonic vowel has not risen to *-i*, reflect the originally stressed position of the vowel (*nèpos*, *sènior*).[164] It is also known that the dental imparisyllabic nouns in *-tas* continue the (inanimate!) nominative form in ancient Tuscan: e.g. *Cìvita* (< *civitas*), *potèsta* (< *potèstas*). Even the southern Italian varieties display continuators of the nominative form: e.g. *omo*, *prèvete* (< *prèsbyter*), *curàto-lo* (< *curàtor*).[165]

It is interesting to notice that some studies on Italo-Romance are based as much on the Late Latin evidence as on the Romance data. I shall review some findings of this method briefly because it opens interesting insights into the inherent traces of the Latin case system in the (Italo-)Romance lexicon. Indeed, even the much debated Italian plural seems to be mainly of accusative origin, as the 1st declension is likely to have undergone a uniform phonological process: /as/ > /aj/ > /e/ (see (32)). This is probable because, had the point of departure been the nominative, the ending /e/ would have palatalised the preceding occlusive (*amike* > **amitʃe*). In the 2nd declension, the result of /os/ > /oj/ > /o/ led to ambiguity concerning the number, hence the borrowing of the (agentive) ending *-i* for both active and inactive subjects.[166]

---

[161] Smith 2011, 283. See, however, Formentin 2012, 56–57, for 14th-century Romanesco nominative/accusative doublets of personal names: the speakers utilised pairs of nominative- and accusative-based forms, such as *Cecho* and *Ceccone*, in free variation, but were still aware of their being of the same origin.

[162] Politzer & Politzer 1953; Sabatini 1965; Zamboni 2000a, 104–114; La Fauci 1997, 58; Faraoni 2014a, 103. For Friulian, see Finco 2013, 343–350.

[163] Zamboni 2000a, 108–109; Faraoni 2014b, 2.

[164] Formentin 2002, 300–301.

[165] Zamboni 2000a, 108–109.

[166] Faraoni 2014a, 113–114; Maiden 1996, 173; Maiden 2000, 177. Indeed, the agentive Italian plural nouns with /k g/ stem have palatalised the occlusive in /tʃ ʒ/ (continuants of the nominative ending *-i*, such as *amico – amici* 'friends'), whereas the non-agentive /k g/ stem plural nouns display no palatalisation (continuants of the accusative ending *-os*,

(32) /*amik-as* > **amik-aj* > *amik-e*/

In general, studies on the origin of Italian plural forms have proven to be especially helpful for research on the Late Latin case system. Sabatini's (1965) pioneering study of Late Latin charter texts suggests an early date of neutralisation (6th century) for Southern and Central Italy (Table 3.4.), but Faraoni's (2014) systematic observations on charter Latin show a discrepancy between the singular and the plural (Table 3.5.). Faraoni postpones the neutralisation of case opposition for the plural to the end of the 8th century in Tuscany and Central Italy and to the 9th century in Northern Italy. Many singular forms are likely to have lost the bicasual opposition earlier than the plural: the phonetic difference between the nominative and the accusative forms vanished early in the 1st declension (-*a* vs. -*am*) while the 2nd declension probably kept the endings distinct for a longer time (-*us* vs. -*u*(*m*)/-*o*). Only the 3rd declension imparisyllabic nouns remained clearly distinctive (e.g. *pars* vs. *partem*).

Table 3.4. Late Latin plural forms: Sabatini's (1965) model.

	Southern and Central Italy				Northern Italy			
	1st decl.	2nd decl.		3rd decl.	1st decl.	2nd decl.		3rd decl.
sg.	-*a*	-*o*		-*e*	-*a*	-*o*		-*e*
pl.	-*e*	-*i*	-*a* / -*as* (neut.)	-*i*	-*as*	-*i* (SBJ) -*os* (OBL)	-*a* (neut.)	-*es*

(Sabatini 1965, 979, 982, 987; Faraoni 2014a, 103)

Table 3.5. Late Latin plural forms: Faraoni's (2014a) model.

Century		Southern and Central Italy, Tuscia included			Northern Italy		
		1st decl.	2nd decl.	3rd decl.	1st decl.	2nd decl.	3rd decl.
7th c. onwards	sg.	-*a*	-*o*	-*e*	-*a*	-*o*	-*e*
7th c.	pl.	-*as*	-*i* / -*os*	-*es* /-*is*	-*as*	-*i* / -*os*	-*es* /-*is*
8th c.	pl.	-*as* / -*e*	-*i* / -*os*	-*es* /-*is* / -*i*	-*as*	-*i* / -*os*	-*es* /-*is*
9th c.	pl.	-*e*	-*i*	-*i*	-*as* / -*e*	-*i* / -*os*	-*es* /-*is* / -*i*
10th c.	pl.	-*e*	-*i*	-*i*	-*e*	-*i*	-*i*

(Faraoni 2014a, 113)

---

such as *fuoco – fuochi* 'fires') (Maiden 2000, 176–177). For the vast literature on the origin of the Italian plural forms, see the summary in Faraoni 2010, 19–20. For further proof of the existence of a bicasual system behind the 2nd declension plural forms (in Venetian), see Formentin 2004, 110–116.

It is necessary to discuss here the implications of the proposed chronologies. First, it has to be remembered that the above analyses do not only include subjects, but all plural nouns *together* (objects, subjects, and complements of prepositions). This is understandable from the viewpoint of Romance languages that do not contrast case forms any more, but scarcely appropriate in a study of Late Latin evidence. Moreover, the data of the above analyses come from the nominal lists of the free parts of charters. These lists are obviously more or less asyntactic: they contain only a few subjects and the syntactic status of the listed items often remains ambiguous, as they are only loosely connected to the governing verb.[167] Therefore, it is not surprising that Sabatini's method fails to recognise a semantically-motivated subject case opposition in the plural, as pointed out by Faraoni.

To do justice to these important studies on the plural, I emphasise that these studies do not aim to examine the bicasual opposition. Indeed, the singular goes more or less unheeded in these studies that focus on the "eternal problem" of the origin of Italian plural forms. As the studies are solely interested in the ending, they even treat together all the 3rd declension singular imparisyllabic nouns. Their morphology still distinguishes between nominative and accusative although the ending is *-e* (e.g. *parte*), just as it is also with the parisyllabic nouns (e.g. *teste*). To sum up, although Faraoni's account explains plausibly the origin of the Italian plurals, it cannot be utilised as evidence when studying morphosyntactic alignment because it does not tell the difference between subjects, objects, and other syntactic functions. Indeed, even if the case contrast had been completely neutralised in the 7th century, the nominal list data would not be able to show that reliably.

Nevertheless, Faraoni's phonological explanation for the plural seems to support the theory of prolonged nominative/accusative contrast in the 2nd declension singular: if there was still a contrast between plural forms *-e* and *-as* (e.g. /e/ vs. /aj/) and *-i* and *-os* (e.g. /i/ vs. /oj/) in Central Italy of the 8th century, the same treatment of final /s/ is likely to have applied to the 2nd declension singular nominative ending *-us* (perhaps /uj oj/). As a consequence, phonology still allowed a morphosyntactic opposition between the nominative- and accusative-based singular forms. These forms may have surfaced with different degrees of classicising normalisation, e.g. *amicus* vs. *amicum*, *amicus* vs. *amicu*/*amico*, or even *amico* vs. *amico*.

At a certain point in (very) Late Latin, the pronunciation of what had been /s/ had become something like /j/ in word-final position and was not necessarily any more perceived as an allophone of

---

/*s*/. Consequently, the scribes more and more often failed to write <*s*> in final position. The chronology of the final -*s* lenition and definitive loss is imperfectly known.[168] It is, however, possible that the letter <*o*> (*amic-o*) represents a nominative-based singular form with /*oj*/ (*/amikoj/ < /amikus/*) at the time of LLCT. This kind of paradigm is, of course, not very practical, as it cannot ascertain the number contrast. In Old French, however, an exactly similar contrast was able to survive for centuries because it was supported by the article and by conventionalised word order.[169] That was, obviously, not the case in Italy and, in the end, a complete neutralisation of the case contrast spread across the whole nominal declension. I shall argue in section 4.2.2. that the phonologically resistant nominative/accusative forms, such as the 3rd declension imparisyllabic nouns, may have been responsible for maintaining the contrast even in such categories as the 2nd declension singular, where the phonological difference had long been weakened. The residues of the semantically-based alignment that will be pointed out in this work support this conclusion.

3.4. Problems concerning the interpretation of the data

The next paragraphs discuss some of the problems involved in the interpretation of the data presented in the above examples. The phrases (5) to (31) seem to settle quite conveniently on a chronological continuum that begins from low-transitivity domains and ends in high-transitivity domains. There are, however, some wild cards with the interpretation of the data that diminish the reliability of the outlined progression. This is particularly true for the areal chronology. Due to the nature of the available non-literary material, the areal representativeness of the data is poor and, additionally, some of the texts cannot be attributed with certainty to any geographical location. For example, *Itinerarium Egeriae* cannot really be used as evidence of the extended accusative in Spain because the place of writing of *Itinerarium* is identified with Spain only with difficulty.[170] In general, the pre-LLCT occurrences are so sporadic and of such a diverse origin that any areal distribution based on them is tentative at best.[171]

---

[168] Proskauer 1910, 187–188; Adams 2013, 132–135; Zamboni 1998b, 660; Väänänen 1966, 81. The Politzer have studied, in the Lombard charters, instances of *s*-omission that are likely to be purely phonologically conditioned (e.g. *quesierimu* pro *quesierimus*). They show that in the Lombard charters of the 7th and 8th centuries the omission of final -*s* is perfectly graded from Central to North Italy, with the most omissions in Central Italy (Siena). The Politzer conclude that the awareness of the final -*s* (and, to some extent, of final -*t*) was less clear in Central Italy. This is thought to reflect a better retention of final -*s* in the Western Romance type of language in the north of Italy (Politzer & Politzer 1953, 14). The similar gradation is also observed with the plurals of the 1st, 2nd, and 3rd declensions, albeit with greater ambiguity (Politzer & Politzer 1953, 25, 27, 29–30).

[169] E.g. van Reenen & Schøsler 2000.

[170] Cf. Cennamo 2009, 326. On the origin of the *Itinerarium*, see Väänänen 1987, 8, 154–157; Adams 2007, 709–710.

[171] See also Adams 2013, 250–251.

Second, many of the above occurrences come from rather late manuscripts. Nothing guarantees that their text represents the original linguistic state of their date of redaction. The textual variants may easily derive from intermediate transmissional stages, as has been shown elsewhere for the Rule of St. Benedict.[172] The oldest surviving manuscript of *Mulomedicina* dates back to the 15th century, and no one knows how many manuscript generations separate it from the original 4th-century text. The same also holds good for *Historia Francorum* (7th c.), Antoninus Placentinus (8th/9th c.), Anthimus (9th c.), and *Itinerarium Egeriae* (11th c.), although the gap between the original text and the first manuscript witness is narrower.[173] Consequently, it is difficult to tell whether the attestations of extended accusative in, say, Antoninus Placentinus, ought to be attributed to the late 6th century when his *itinerarium* was compiled or to the late 8th century, i.e. the date of the oldest surviving manuscript. Only the material surviving as original is free of these restrictions.

Third, it is always problematic to use written code as evidence of the changes taking place in the spoken language. Any written text, however non-literary it may be, still obeys some of the conventions of the written code. This may conceal the current linguistic circumstances almost completely, as in Late Latin where the conservative written code was very far from the reality of the spoken language.[174] The emergence of linguistic innovations in written texts is supposed to be register-specific. Rovai discusses the role of different registers concerning the extended accusative. The extended accusative seems to manifest itself exclusively in low and middle registers, such as those of the *Berufsliteratur* and inscriptions, and later also in the laws of the Germanic nations and more ambitious literature (*Historia Francorum*).[175] This is also a counter-argument to the claim that the alignment change cannot be taken earnestly because there are no texts that follow systematically the active/inactive alignment or where the nominative really shows signs of regression.[176] There cannot be texts of this kind because, until the emergence of vernacular written texts and even later, all the texts written in Latin followed the conventions of written registers, albeit to a varying degree.

---

[172] Coleman 1987, 50; Coleman 1999, 352–356. See also Galdi 2013, 12–13, for the *Historia Romana* of Jordanes.
[173] Ledgeway mentions the textual corruption in passing (Ledgeway 2012, 329), while Adams refers to it regarding *Mulomedicina* (Adams 2013, 247, 253). For the dating of *Historia Francorum*, see Bonnet 1890, 16–17; for *Itinerarium Antonini Placentini*, Rovai 2005, 58; Milani 1977, 32; for *De observatione ciborum* of Anthimus, Grant 1996, 43–44.
[174] Ellis & Beattie 1986, 201–202; Halla-aho 2009, 28–29; Rovai 2012b, 174–175.
[175] Rovai 2012b, 108–109, 171–174.
[176] Cf. Adams 2013, 247: "it would be misleading to imply that subject accusatives are common in Late Latin or that there is an attested stage of the language in which they can be seen to be established, if still outnumbered". When speaking of Late Latin, Adams often seems to mean earlier centuries than Cennamo and Rovai. LLCT clearly represents an attested stage in which the extended accusative is as established as it can be in written texts.

It is also worth recognising that the first attestation of any phenomenon in a text does not correspond to its first attestation in the language. Texts are highly problematic sources as far as the dating of linguistic change is concerned. Even though the occurrences presented in this chapter seem to follow a pattern predicted by the theory of semantic alignment change, the chronology proposed on the basis of these attestations is merely the chronology of when a certain phenomenon happens to occur for the first time in the corpus of those texts that survive up to our own days. There is also little evidence of the relative frequencies of accusative subjects. Such frequencies, proportioned to all the possible environments, i.e. all the subjects, would be of vital importance in examining linguistic change, as stated below.

This is, of course, the eternal problem of Latin linguistics: one has to cope with insufficient evidence, which leads to results that are less reliable than those that are obtained with modern languages. Nevertheless, a result can be viewed as convincing if it is representative of the available evidence and in concord with a relevant theoretical framework. The case of semantically-based alignment in Late Latin seems to fulfil these conditions relatively well. Most of the data, including those yielded by this work, seem to support the theory. Given the nature of the available texts, it is, however, likely that there will never be more convincing evidence concerning the early centuries of the development (c. 4th to 7th century).[177] This fact should not prevent using the by now under-used block of 8th/9th-century Italian charters to shed light on the final phases of the realignment – not perhaps any longer the stage of the active/inactive alignment, but the later ergative/absolutive phase, or the phase when the accusative had already entered the low-transitivity domains of the transitive patterns.

Next, I discuss briefly some problems of a clearly linguistic nature. It has to be understood that some of the extended accusatives may have nothing to do with a hypothetical alignment change. Adams (2013) proposes individual explanations to nearly all the extended accusatives presented here in section 3.2., especially to the earliest ones. It is certainly true that everything cannot be reduced to the alignment change. On the other hand, many seemingly unrelated occurrences of accusative subjects that can be explained with *ad hoc* explanations, such as anacoluthon, may nonetheless be symptoms of a large-scale linguistic change that led finally to the accusative becoming the default case (a notion vehemently opposed by Adams): when the writers slipped, they more and

---

[177] The only thinkable and thus far unexploited major sources of the extended accusative seem to be the Vandal charters from the *Tablettes Albertini*, the Ravenna papyri, and the Merovingian charters. For the usability of written texts for Latin linguistics, see Herman 2002, especially 32–33, 44. For a good example of how to use written documentary texts to study the spoken Latin, see Fiorentino 1994.

more often unconsciously ended up producing the accusative form. In earlier times, they had usual-
ly ended up producing the nominative, as Adams duly emphasises.[178] Generally speaking, Adams's
critique seems justified concerning the blind application of the realignment explanation to early
data.[179]

From the linguistic viewpoint, the main problem of the chronology proposed in section 3.2. with
respect to the underlying theory seems to be the fact that the first (rather ambiguous) occurrences of
the extended accusative with transitive constructions (sentences (28) and (29)) make their first ap-
pearance almost in tandem with the emergence of the extended accusative with intransitive (S_O)
constructions (i.e. in the 4th/5th centuries). According to the theory, the extension of the accusative
to intransitive constructions should have been a clearly precedent evolutionary stage. Similarly, the
first occurrences of the extended accusative with S_A subjects are rather late and post-date the first
occurrences of the A subject occurrences. Thus, the chronology emerging from the data is not as
neat as has been claimed. This is, however, not necessarily disastrous for the theory because the
LLCT data may actually explain the mentioned 'problems'. I refer here to chapter 5 of this thesis,
where it will be shown that, at least in LLCT, the S_A subjects are unexpectedly high in transitivity,
while the A subjects are unexpectedly low in transitivity. Therefore, the A subjects are likely to slip
into using the accusative more easily than would be expected offhand.

The point of departure for all linguistic change is variation. A linguistic change often appears for a
long time as a bias towards certain variants, and this can be perceived only through the relative fre-
quencies of different variants. These distributions can be observed most reliably in systematic cor-
pus studies, such as the one conducted here on LLCT. This intimate relation of variation and change
also applies to the realignment of the Late Latin case system: the different stages certainly existed
synchronically with certain subject types occurring more often in the accusative than the others. As
time went by, this distribution changed but, at any point of time, several variants with different fre-
quencies were to be observed. This is why accusative A subjects can be found in rather early texts
simultaneously with S_O (and S_A) subjects: in the 4th/6th-century texts, for example, there are a few
(early) attestations of accusative A subjects while the only really considerable evidence is for the
extension of accusative to S_O subjects. In other words, accusative A subjects have begun to appear

---

[178] Adams 2013, 234–256. Benucci (2004) advocates a similar explain-it-away view on the extended accusatives of
Italic languages (and Latin) although my impression is that the Italic data presented by Benucci might well reflect a
semantically-based alignment.
[179] The possible manifestations of the semantic tendency in Early Latin and Indo-European lexicon are beyond the
scope of this study. For the evidence concerning the (thematic) neuter, see Rovai 2007; Rovai 2012a; Lazzeroni 2002.

little by little (probably first with the least transitive constructions), although the nominative/accusative contrast (probably for the most animate/agentive A subjects) has still been in effect for the most part.[180] There is nothing to prevent us from assuming that, as far as the written code of Late Latin reflects any traits of the spoken language, it reflects more or less the alignment patterns described above in section 3.2.

On the strength of what has been said here, it is only natural that there are no texts with complete active/inactive encoding nor even too many texts in which the accusative subjects "can be seen to be established, if still outnumbered".[181]

4. Semantic constraints of subject case selection in LLCT

The following sections 4.1.–4.4. examine semantic factors underlying case marking alignment in LLCT: the animacy degree of subject, the transitivity degree of construction, and their interaction in the selection of the subject case. Section 4.1. first defines the concepts animacy and construction type and then discusses some problems concerning the classification of certain construction types that are particularly frequent in LLCT. The next two sections 4.2. and 4.3. apply the numerical and statistical method to the LLCT data. Chapter 4.2. is aimed at clarifying the role of animacy and the construction type in subject case selection while section 4.3. consists of a case study that evaluates the transitivity degree of the different construction types basing on a sample of 471 LLCT clauses. Section 4.4. addresses a further theme which arose in the preceding chapters: the peculiar verbal-semantic panorama of the charter Latin genre. Chapter 5 examines syntactic factors that also seem to affect the subject case selection in LLCT.

4.1. Animacy and construction type
4.1.1. Defining animacy

Chapter 4.1. seeks to test the theory, presented in section 3.1., according to which animacy plays a crucial role in alignment change. If semantically-based subsystems still prevailed in the (spoken) Latin of 8th- and 9th-century Tuscany and if these subsystems were somehow visible in the written code, one should expect that in LLCT the case distribution of subjects would reflect, at least rough-

---

[180] For the nature of linguistic change, see Halla-aho 2009, 27: "often what is called a change can only be discerned from some temporal distance, as usually a change is in fact a slowly emerging tendency, visible for a long time only as a statistical preference for one variable".
[181] Adams 2013, 247.

ly, the animacy of those subjects: inanimate common names should be the first and the foremost to occur as accusative subjects and animate proper names the last and the fewest.[182] The actual operationalisation of this rather abstract idea will be discussed in section 4.2. Animacy/referentiality is utilised as an indicator of alignment change because it is an easily measurable property of the subject and is likely to reflect closely certain higher-level semantic properties, the most important of which is the control of subject over verbal event.[183] Although control describes the actor/undergoer status of the subject argument better than animacy, technically measuring the degree of control would require a systematic analysis of a set of semantic traits, including animacy and referentiality.

Indeed, the theoretical models of Latin case alignment change draw on a general notion of animacy (and related properties) as a continuum.[184] As was shown in section 3.1., Croft (2003) has induced from typological data a chain of universals that can be presented as the 'extended animacy hierarchy':

(1)   first/second person pronoun < third person pronoun < proper name < human common noun < non-human animate common noun < inanimate common noun[185]

Similar hierarchies are usually referred to simply as animacy hierarchies although they involve three distinct but related functional dimensions of which only one is about animacy proper.

(2)   Person: first/second person < third person
        Referentiality: pronoun < proper name < common noun
        Animacy: human < animate < inanimate[186]

In practice, it is often not necessary or possible to distinguish all the degrees of the extended animacy hierarchy of (1). This usually leads to fragmentation of the data. On the other hand, a mere hierarchy of animacy is not sufficient, as animacy *per se* is not cross-linguistically reported to be decisive for alignment-split phenomena.[187] Additionally, the personal names of LLCT seem to form a peculiar group of their own and cannot always be examined together with other human-referent nouns. For the current study, I have chosen only those degrees of 'extended animacy' that are rele-

---

[182] Rovai 2012b, 111–112.
[183] Siewierska 1988, 49; Comrie 1989, 59–62, 185–199.
[184] A good example of these is Rovai 2012b, 111–112.
[185] Croft 2003, 129–130; Comrie 1989, 185–188; Dixon 1994, 83–85; Silverstein 1976, 175–176; Siewierska 1988, 30, 49.
[186] Croft 2003, 130.
[187] Cf. Haig & Schnell (in preparation).

vant to the analysis of semantic alignment. The inherent semantic features of subjects will be divided into three classes as follows:

(3)   personal name < animate common noun < inanimate common noun

Henceforth, hierarchy (3), not hierarchy (1) proposed by Croft, will be referred to when the term 'animacy' is used, unless otherwise specified. The hierarchy of (3) is a modified extended animacy hierarchy and consists of the right-tail end of hierarchy (1), except that no difference is made between human common nouns and non-human animate common nouns. I am aware that the hierarchy of (3) is not purely about 'animacy', but for the sake of brevity the title 'animacy hierarchy' is chosen.[188] The personal pronouns are ignored for the reasons explained in section 2.5.1. Personal pronouns are, of course, highly definite and referential and, therefore, even more resistant to the supposed alignment change than the personal names. This is seen, indeed, in LLCT as an extremely high rate of the nominative form of personal pronoun subjects *ego* (section 4.2.1.) and *tu*, as well as in their retention of functional contrast in Romance languages, e.g. It. *io* vs. *me*.

It is noteworthy that, in (3), the animate common nouns can be either human or non-human. This implies that, in LLCT, the class includes, in addition to the titles and kinship terms of the commissioners and addressees of the legal act, even the two animals, *boves* 'oxen' and *porco* 'pig'. They occur, unsurprisingly, as the subjects of passive constructions. This could be pernicious to the analysis, were the number of animals higher. Treating animals separately would be undesirable though, as the subgroup would be too small for reasonable analysis and even the subgroup of human animate nouns would diminish. There is also an additional class that contains other proper names, such as toponyms, but this will not be discussed here.[189] For the sake of brevity, I shall call the three categories illustrated in (3) 'personal names', 'animate', and 'inanimate'.

4.1.2. Defining construction types

This chapter discusses the other semantic variable that has been utilised to detect alignment change: the construction type of the clause in which the subject occurs. In section 3.1., the basics of the the-

---

[188] It is, actually, rather usual to refer to this kind of compound hierarchies by the name of only one of their components (see Croft 2003, 130).

[189] Toponyms are referential and individuated but, in spite of that, usually non-agentive. In addition, toponyms involve certain peculiarities of their own, such as an even greater spelling variation than personal names and crystallisation in locative/directive case forms. I propose to discuss the toponymy of LLCT in a future study.

ory and terminology concerning morphosyntactic alignment were presented. The cross-linguistically attested nuclear arguments A, S, and O, as well as the split intransitivity "arguments" S_A and S_O were introduced. It was also stated that, cross-linguistically, the intransitivity split may be motivated either by the inherent properties of the subject or of the verb. Therefore, in this study, the technical measures, animacy and construction type, are examined on a par to explicate to what extent they are relevant for the alignment of charter Latin. Animacy departs from the semantic features of the subject while construction type has the semantics of the verb as its starting point. Although theoretically separated measures, in practice they can be seen to intersect with each other in the control of the subject over the verbal event, a parameter that was discussed in section 3.1. This is because low-control subjects are usually inanimate and favour low-transitivity constructions while high-control subjects are usually animate or personal and favour high-transitivity constructions.

In section 3.1., I also introduced the following verb type classification: active, unergative, and unaccusative. Here, one must add the passive, which is usually considered to be unaccusative.[190] As it is not guaranteed that the passive behaves on a par with the other unaccusative constructions in LLCT, it will be examined separately. The verb types of this classification correspond to the respective nuclear argument relations A, S_A, and S_O (and S_O passive), respectively. For example, an unergative verb always has an S_A subject, an unaccusative verb an S_O subject, and so on.

As said in section 3.1., I shall utilise mainly the nuclear argument relations in order to avoid terminological overlap. The literature usually refers to the terms 'transitive', 'unergative', 'unaccusative', and 'passive' as verb types or clause types. Although I am mainly interested in the case form of the subject, I want to observe the verbal event as a whole. Therefore, I deem it more justified to use term 'construction type' to refer to constructions involving A, S_A, S_O, or S_O passive subjects plus the respective verb. Table 4.1. lists the fourfold construction type classification that will be utilised in this study. It also maps together the verb types and the construction types and locates them on the control/transitivity continuum. As the study concentrates on the subjects, nuclear relation O is omitted from the table.

---

[190] Rovai 2005, 61; Rovai 2012b, 105.

Table 4.1. Construction type classification applied to the subjects of LLCT.

Verb type	Construction type	Control of subject/ Transitivity of construction	
transitive constructions	A	⬆ ⬇	high
(intransitive) unergative constructions	S_A		
(intransitive) unaccusative constructions	S_O		low
passive constructions[191]			

Even though constructions are located continuously over the continuum, they must be grouped in manageable categories for linguistic analysis. The above-defined fourfold construction type classification is based on various criteria of diverse motivation: the passive is distinguished from active clauses on syntactic grounds, while the split between unaccusative and unergative intransitive clauses is a semantic one. The primary opposition is, however, between the transitive and intransitive clauses (passive included or not included). The borderline between them can be defined in several ways that will be discussed in the following.

This classification is rough and obviously not trouble-free, but its undeniable advantage is its simplicity and accessibility to everyone trained in traditional linguistic theory.[192] Every study of real corpus material requires simplifications that make the analysis possible by providing it with a limited set of easily definable parameters. The defects of the chosen classification are discussed in section 4.3.3. and even earlier when they become particularly apparent.

At this stage, the concept of transitivity has to be discussed in more detail because it is necessary to make a distinction between S_A/S_O subjects and certain non-prototypical A subjects. In other words, it is not always clear whether a verbal event is intransitive or transitive. Later in the case study of section 4.3., I shall define the transitivity degree for a sample of 471 LLCT sentences. There, transitivity will be treated as a technical tool that consists of a set of binary components. For the time being, it is enough to reflect on what actually makes a verbal event transitive or intransitive.

In grammar books, transitivity is about the opposition between transitive and intransitive verbs, i.e. whether the verb has a direct object or not. The traditional semantic definition of transitivity underlines the effective transfer of action from a subject to an object or, better, from an agent to a patient. As its simplest, transitivity is considered a semantic property of a verb that describes the relation of

---

[191] Note that there is no *a priori* reason to locate the passive at the bottom of the scale, as the S_O subjects of the unaccusative and passive constructions are usually thought to be equivalent.
[192] This division is adopted by Rovai 2005, and it is also the underlying assumption in many articles of Cennamo.

that verb with its subject and possible direct object. An equally traditional syntactic definition describes transitivity by way of the number of arguments the verb takes: the verbs taking one argument are intransitive while the verbs taking two arguments are transitive.[193]

Nowadays, transitivity is usually considered a combination of semantic and syntactic plus pragmatic factors (e.g. Hopper & Thompson 1980). Definitions differ also in regard to whether transitivity is viewed as a property of verbs or clauses.[194] In this study, transitivity is seen fundamentally as a property of the clause or verbal event involving the verb and its arguments, not as an inalienable property of the verb itself. For the purpose of any language-specific data-driven study, it is useful to consider transitivity to be a semantically conditioned continuum rather than a binary category. I also consider transitivity to be a continuum where the verbs that take direct object are located at the higher end and the intransitive, objectless verbs at the lower end of the continuum. This is what Hopper and Thompson did in their seminal paper in 1980: they proposed a scalar notion of transitivity that is characterised by the interplay of a number of unhierarchised parameters. Hopper and Thompson's model is discussed in detail in section 4.3.

Within this kind of multifactorial framework, even verbs with objects can have differing degrees of transitivity, which turns out to be important for this study. Some recent corpus-based studies that are inspired by functional linguistics focus more on the actual language use, such as spontaneous conversation, and emphasise the genre-dependent nature of transitivity, at the same time renouncing most of the conceptual apparatus traditionally constructed around it.[195] Some of their results can be exploited when analysing LLCT (see section 4.4.).

It is usually rather straightforward to decide whether an intransitive verb is unaccusative or unergative (not to mention passive), as only a few transitivity components pertain to the decision. As stated, the problems arise only when one is asked to draw a line between intransitive and (non-prototypical) transitive constructions. The role of the O participant is traditionally considered to be essential to the definition of transitive and intransitive clauses.

Nevertheless, the LLCT corpus material, as well as conversational data, as pointed out by Thompson 2002, show that several frequent object categories are indeterminate or, at least, non-

---

[193] Kittilä 2002, 20–25.
[194] Kittilä 2002, 25–26.
[195] E.g. Thompson & Hopper 2001.

prototypical. This is to say that it is necessary to use violence in forcing some of the subject/verb combinations, i.e. clauses, of LLCT into one of the four categories presented in Table 4.1. (transitive, unaccusative, unergative, passive). It is a matter that certainly does not do justice to the scalar nature of transitivity but is required by the operability of the analysis. The principles of this forcing are explained briefly in the following two sections, section 4.1.3. and section 4.1.4. Some of the choices are not self-evident but must be justified or, at least, clearly written out. A few issues that will be discussed are borderline cases that probably would not be worth discussing with other corpora but are frequently attested in LLCT.

4.1.3. Object deletion and transitivity

In this and the following chapter, I shall explain how I deal with the problems involved in the classification of semantically transitive clauses with no overt object participant (section 4.1.3.) and the status of speech verb clauses (section 4.1.4.).

In many languages, certain semantically transitive verbs can be employed without any overt direct-object argument. This phenomenon, which Naess (2007) calls indefinite object deletion, can be viewed as an instance of a more general valency reduction. Hopper and Thompson (1980) do not define whether the semantically required participants really need to be present in the surface structure or whether they can remain unrealised. It is clear, however, that they consider the prototypical situation to involve overt participants.[196]

Naess divides the object deletion into context-dependent and context-independent deletion. Most languages allow omission of objects whose reference can be retrieved from context. Such context-dependent object deletion takes place where the object has already been mentioned in the preceding discourse or where the general context provides sufficient clues to the identification of the object.[197] In charter Latin, the absence of the logical direct object of a semantically transitive verb can often be described as a technical (context-dependent) ellipsis favoured by the text type, e.g. avoidance of repetition inside a formulaic clause. This is especially common in comparative clauses, such as (4). The logical object of (5) is also found in the previous discourse. The context-deducible arguments are added in curly brackets.

---

[196] As is known, Latin favours pro-drop, i.e. omission of the pronominal subject. Thus, any supposition about two overt participants must be revised for Latin.
[197] Naess 2007, 124–125.

(4) MED 505 (AD 830) *angaria vero vobis facere debeamus per singulos annos*, *sicut illi alii vestri massarii vobis* {*angaria*} *faciunt*

"we have to do the corvée labour for you in the same manner as your other tenants do {the corvée}"

(5) CDL 261 (AD 772) *et post meo decesso*, *filii mei similiter* {*uno tremisse* (OBJ), *tibi* (IOBJ)} *dare deueas* [= *debeant*], *sicut ego supra premisi*

"and after my death, my sons will have to give {you (IOBJ), one *tremissis* (OBJ)} in the same way as I promised above"

I classify as transitive those semantically transitive verbs whose semantic valency includes a patient/theme[198] relation as a direct object, notwithstanding their syntactic valency, i.e. whether they have an overt object in their surface structure or not. Once, along with Hopper and Thompson, transitivity is seen as a multifactorial phenomenon, the mere number of realised participants is not decisive *per se* because reducing or adding participants does not necessarily change radically other transitivity components. The semantic structure of the verb *dare* 'to give' includes semantic relations <agent, theme, recipient>, of which only the agent is realised in (5), while the theme (*uno tremisse auro aut oleo*, *cira* "one *tremissis* of gold or oil [or] wax") as well as the recipient (*ad ipsa sancta ecclesia*) can be deduced from the preceding discourse. Sentence (6) immediately precedes sentence (5).

(6) CDL 261 (AD 772) *per singulus anus dare deueas* [= *debeam*] *ad ipsa sancta ecclesia uno tremisse auro aut oleo*, *cira*, *quem habuero*

"I must give every year to that church one *tremissis* of gold or oil or wax, whichever of those I happen to have"

In this chapter, I refer to certain transitivity components that feature in Hopper and Thompson's (1980) transitivity scale, which will be presented in section 4.3.1. For the present, it is necessary to understand that Hopper and Thompson's scale consists of ten transitivity components, each of which can have a binary value 0 or 1. The scale includes parameters that are both semantic, syntactic, and pragmatic.

---

[198] Theme is a semantic or thematic role that is in a state, changes its state, or undergoes an action but is not particularly affected. E.g. "I set the charter on the table" or "I gave him a charter".

In (5), the following high-transitivity components of Hopper and Thompson remain unchanged: action, telicity, punctuality, volitionality, affirmation, potency of A (agent/subject). Only the components related to O (object) become impossible to define because O is left out. Actually, the theme argument of the verb *dare* would not be very affected in any case (it is theme, not patient).[199] This is also the case with *uno tremisse auro aut oleo*, *cira*, which are rather non-individuated in the current sentence. They are all scarcely referential in the prevailing context because their referents are not present. As O is omitted, it seems plausible to think that, as a consequence, both its affectedness and individuation are reduced to a minimum. To put it another way round, at least in charter Latin, O is more likely to be deleted if it is low in affectedness and individuation (hence Naess's 'indefinite' object deletion). The normal interpretation of a transitive verb used without an object seems to be that the general action is of more interest than the given unspecified object.[200]

Hopper and Thompson's list (section 4.3.1.) is mainly intended to be used for cross-linguistic classification of transitivity and for illustrating different morphosyntactic manifestations of transitivity alterations, such as a change in case marking. It quickly becomes obvious that certain components must be further divided into more specific parameters in order to make case-specific analyses. Affectedness and individuation of O are cover terms for a bundle of parameters that together determine the prototypicality/non-prototypicality of O: animacy, definiteness, referentiality. Animacy and referentiality can be opened up in the form of an animacy hierarchy (human < animate < inanimate) and a referentiality hierarchy (pronoun < proper name < common noun), respectively (see section 4.1.1.).[201] These parameters help in understanding the special status of objectless transitive verbs. The deleted objects seem to be always inanimate and lowly referential in LLCT (cf. (4) and (5)) and their relevance for the current verbal act is so scarce that they can be left out.

Theoretically, it can be discussed which one is more important to transitivity, an agentive A or an affected/individuated O. In this study, where the case selection of the subject is in focus, it is the prototypical A, i.e. the universal syntactic-semantic primitive of the transitive-clause subject, that is considered more essential for transitivity than the prototypical O. I find it important to emphasise

---

[199] The subject of the verb *dare* 'to give' can be interpreted as an affected agent because it is deprived of what is given, whence the reduced potency of A. For the affectedness of A in object deletion verbs, see Naess 2007, 126–127.

[200] Munro 1982, 305.

[201] Malchukov 2006, 330–350; Croft 2003, 130. Malchukov (2006, 333) seeks to defeat the inconsistencies of Hopper and Thompson's transitivity scale by proposing a new scale that stretches from subject-related parameters via verb-related to object-related properties. The scale is intended to explicate mutual semantic affinities between individual transitivity parameters.

the fact that the semantic valency/subcategorisation frame of the verb remains the same in spite of the presence or absence of the object. The conclusion drawn from this in this study is that context-dependent object deletion does not diminish the transitivity of the verbal event to the extent that the event ought not to be viewed as transitive any longer. As a consequence, the 'transitive' clause category, as defined in this study, includes verbal events of vastly differing degrees of transitivity. This is also why it is necessary to observe the transitivity degree of different construction types separately in the case study of section 4.3.

4.1.4. Speech verbs and effected objects

Another special case of O are the speech verbs or *verba dicendi* (*et declarandi*). They usually lack a prototypical (referential and definite) pronominal or nominal object. The verbs denoting speaking and speech acts are considered transitive because they can have a clausal object complement (7), an accusative and infinitive complement (8), a clausal complement in direct speech, i.e. a syntactically unmarked quotation (9), or a pronominal or nominal object (10).[202] They can occur, however, without any object (11) like the above-discussed object deletion verbs. In the latter case, the omission of the object can be interpreted as being context-independent (see above). Clausal complements and direct quotations are by far the most common strategies of speech verb complementation in LLCT.

(7) MED 309 (AD 803) *prenominati sacerdotes* [--] *iurando* <u>*dixerunt*</u> *ut veritatem fuissent*
"the above-mentioned priests [--] <u>said</u> under oath that it was the truth"

(8) MED 298 (AD 801) <u>*promitto*</u> *me esse componiturus tibi ipsas prefatas res*
"I <u>promise</u> that I shall compensate you the mentioned property"

(9) MED 335 (AD 807) <u>*respondebat*</u> *ipse Alprandus clericus*: *veritas est quia abeo res ipsa quam dicis*
"Alprandus, the clerk, <u>answered</u>: 'the truth is that I possess that property that you speak of'"

(10) CDL App. (AD 786) *ut* [--] *omnia ei certam* <u>*diceret*</u> *ueritatem*
"that [--] should <u>tell</u> him the whole and exact truth"

---

[202] Munro 1982, 304–306.

(11) MED 539 (AD 838) *isti omnes per sagramenti testum equaliter <u>dixerunt</u>*

"they all <u>spoke</u> in the same way swearing by the holy scripture"

Cross-linguistically, speech verbs are often imperfectly transitive, as the quotations and object clauses they introduce are usually very far from being prototypical objects. Munro (1982) discusses mainly unmarked direct quotations of type (9), but her observations go for (marked) clausal structures as well. In Munro's terminology, direct quotations are viewed as unmarked while indirect, syntactically bound quotations are marked. Cross-linguistically, there are several structures that align speech verbs with intransitive rather than with transitive verbs. In this context, it is especially important that the speech verbs pattern syntactically with intransitive verbs. They also display absolute subject marking in certain ergative/absolutive languages.[203]

The special nature of speech verbs is due to the fact that their undergoer arguments are effected objects. An effected object is one that does not exist before the verbal act begins but only comes about as a result of that act. Speech verbs are typical representatives of verbs involving effected objects, and some of these verbs can occur with object deletion, as was stated above (see (11), cf. *dicere* 'to say', but also 'to speak'). Effected objects are non-referential by nature and this non-referentiality of the object is inherent to the semantics of the speech verb itself. In general, effected objects show a low degree of distinguishability from the agent and distinctiveness from the general background. This low degree of distinctness makes them highly susceptible to omission.[204] In LLCT, the omission of effected objects often seems to be context-dependent because an adverb, such as *similiter*, is usually present. What is important concerning the transitivity component parameters of Hopper and Thompson (1980) is that effected objects cannot be affected by verbal action because they come to exist only through the very same action. As effected objects do not undergo a change of state caused by the verb, they cannot be described as prototypical patients.[205]

In sum, it seems quite clear that the described speech verbs can be classified as transitive. Even though the direct quotations or the clausal/infinitival complements are effected objects that are non-referential, low in distinctness, and hardly affected by the act of saying, there is still a volitional agent.[206] It is true that the lack of object affectedness and individuation is cross-linguistically found

---

[203] Munro 1982, 302–312; on verbs of saying in Classical Latin prose, see Spevak 2010, 157–164.
[204] Naess 2007, 103–104, 127–128
[205] Naess 2007, 106.
[206] Naess 2007, 127; Munro 1982, 313, 316.

to be associated with several morphosyntactic intransitivity phenomena,[207] but the relative weight given to the affectedness of the patient (in respect to having a volitional agent) depends on the objective of each individual study. As with object deletion verbs, I continue to consider speech verbs transitive enough for the purpose of this study on case alignment. This choice is in conflict, though, with the one preferred by Thompson and Hopper (2001) in their study of modern conversational English.

Thompson (2002) sketches a radically different picture on object complement clauses. She not only denies that clausal object complements are arguments and that they are subordinate to the main clause, but she also rejects the whole conceptual division between the main clause and the complement clause. These functionally-based observations cannot, of course, be directly extrapolated to charter Latin, but the idea is that, as far as discourse dynamics is concerned, clausal complements are often more important than the main clauses. All this is based on the observation that certain highly frequent main clauses, such as *I think* or *I guess*, seem to be limited to expressing epistemic/evidential/evaluative value, i.e. reporting the speaker's stance towards the content of the complement clause.[208]

In the same way, several main clauses of the main clause plus clausal/infinitival complement combinations of LLCT can be interpreted as not having a proper independent value. It cannot be denied that in the phrases (7) to (9) the most important information is carried by the clausal/infinitival complements while the reporting clause predicates *dixerunt*, *promitto*, and *respondebat* can be viewed as conveying the evidential background for the fact expressed in the complement. They tell who is the responsible author of these legally binding statements, which is an important piece of information in the context of legal documents (phrases (7) and (9) come from a trial record, (8) from the *sanctio* clause of a concession charter). The reporting clauses are essential in the charters in so far as they address the pertinent persons who pledge themselves to the agreed contract, but the real content of their commitments is recorded in the clausal/infinitival complement. The secondary role of the reporting clauses is even more accentuated in phrases (12) and (13).

(12) MED 385 (AD 813) *qui* <u>*dixit*</u>: "*missa postea non cantavit* [= *cantavi*]"; *tunc ibidem nobiscum aderant idonei homines qui* <u>*dixerunt*</u>: "*et quare negas ut missas non cantasses quia nos te vidimus missas cantare*?"

---

[207] Hopper & Thompson 1980, 256–259.
[208] Thompson 2002, 125–139.

"he <u>said</u>: 'I did not sing the mass thereafter'; then there were with us some trustworthy men who <u>said</u>: 'and why do you deny that you sang the mass because we saw you sing the mass?'"

(13) MED 335 (AD 807) *respondebat Teusprandus cl*(*ericus*) <u>*dicens*</u>: "*tu res ipsa nec eccl*(*esiam*) *avere non debes*"

"Teusprandus, the cleric, <u>answered</u>: 'you shall not have the property nor the church'"

In (12), the two predicates *dixit* and *dixerunt* merely state the fact that what follows is a direct quotation. They also serve to emphasise the endeavour of recording verbatim the exact (fraudulent) words of the accused as well as the aggravating reply of the testimonies.[209] Present participle *dicens* in (13) illustrates well the versatile functions of speech verbs in charter Latin. *Dicens* seems to be redundant at first sight, as it is preceded by a finite-form speech-verb predicate *respondebat*. Yet, *dicens* carries an important discourse feature: it specifies the opening of the direct quotation that could otherwise remain obscure because the connection with the finite verb is broken by the (rather unusual) postposition of the subject, *Teusprandus*. The function of *dicens* in late and medieval Latin can be likened to the colon that is the modern textual means to indicate the beginning of direct quotation. Since medieval writing did not use punctuation in the modern sense, it was practical to separate quotations from other texts using customary marker words.[210] The same seems to concern the word *et* as well: *et*, which usually appears as the iconised logogram &, is used to mark the sentence or phrase border, often substituting the modern full stop plus capital letter combination. Thus, *dicens* and *et* form a kind of proto-punctuation although their use is not consistent. What is said of *dicens* applies, *mutatis mutandis*, to the finite forms of *dicere* and other speech verbs, as was seen in (12).[211]

The Classical Latin standard complementation method of *verba sentiendi et dicendi*, i.e. the accusative and infinitive construction, has lost ground to clausal complementation by the time of LLCT.

---

[209] The *notitia iudicati* of MED 385 is a reconfirmation of the sentence passed ten years earlier (AD 803) on Alpulus, the priest. Alpulus had not recognised the sentence, hence the bishop-judge's zeal to render the new protocol as non-appealable and final as possible.

[210] *Dicens* may have been originally adopted from Biblical Latin where it may have been a calque of the Greek source text. The frequency of the phenomenon in Greek Biblical language can be an effect of the similar Hebrew usage, mediated by the *Septuaginta*. E.g. *Gen.* 1.15–16: *locutus est autem Deus ad Noe* <u>*dicens*</u>: "*egredere de arca*" (Vulgata), καὶ εἶπεν κύριος ὁ θεὸς τῷ Νωε <u>λέγων</u>: "ἔξελθε ἐκ τῆς κιβωτοῦ" (*Septuaginta*), וַיְדַבֵּר אֱלֹהִים אֶל־נֹחַ לֵאמֹר׃ צֵא מִן־הַתֵּבָה ("and God spake unto Noah, saying, go forth of the ark"). On the other hand, the same framework conditions concerning punctuation prevailed in the whole ancient world and, therefore, it is not impossible that a marker word based on verb 'to say' was an independent innovation in all these languages.

[211] From the viewpoint of conversation analysis, (12) also indicates a turn shift in the legal protocol where lengthy addresses are reported between the opposed parties. For *et*, see Halla-aho 2009, 65–67; Beaman 1984, 59–61.

Various conjunctions, such as *ut* and *quod*, are used to introduce the complement clause, but this study cannot go into the discussion on the Latin complementation and its diachronic development.[212] The exact nature of the syntactic relation between the subordinated ACI infinitive plus its accusative subject and the subordinating matrix verb has been the subject of debate. Nevertheless, in the present study, the accusative and infinitive construction is considered an object complement of the matrix verb on a par with other complementation strategies.[213]

One peculiar feature of the legal jargon of charter Latin is valency reduction and an apparently abnormal realisation of arguments. The direct object has sometimes been replaced by a prepositional phrase, as in (14). The normal wording is *me confirmavit*, but here the argument structure of verb *confirmare* has been modified to fit a novel *sicut* clause. There are two possibilities: either the originally transitive verb has been intransitivised or the verb remains transitive with a deleted object (e.g. *rem* 'the property') and an extra beneficiary argument (*in me*). The interpretation obviously affects the classification of the clause. In this study, I have classified structures of this kind as transitive with a deleted object.

(14) MED 414 (AD 818) *sicut ipse quondam Richipertus in me per textum cartule confirmavit*
"as the late Richipertus confirmed [e.g. the property] to me through the charter text"

Lastly, I run through a few other principles that are followed in the construction type classification of this study: modal verbs and the *videri* 'to seem' construction. The modal verbs do modify the transitivity degree of the main verb, but I still classify modal verb + main verb combinations according to the main verb properties. The modifying power of the modal verbs will be taken into account in the case study of section 4.3. where the transitivity degree is diagnosed by way of ten transitivity components defined by Hopper and Thompson (1980). For example, *debere* 'to have to' in clause (15) reduces the volitionality and potency of the subject (see Table 4.10. in section 4.3.). Similarly, *posse* '(to) can' in clause (16) signifies that the verbal event is not described as real but potential. This obviously affects the degree of control the subject has over the action conveyed by the verbs *molestare* and *resubtragere*.

---

[212] I refer here to the comprehensive and detailed studies of Greco (2012, 23–50) and Adams (2005).

[213] Bolkestein (1976, 269–272) divides the accusative and infinitive constructions into two classes: in *dico* type constructions, the infinitive + NP is claimed to be the direct object of the matrix verb while, in *admoneo* type constructions, it is only the NP that is the direct object of the matrix verb, the infinitive being the third argument of the verb. Rivas (2000, 94–115) objects to this view by showing how the borders between the two alleged classes are not impermeable.

(15) MED 301 (AD 802) *angaria vero per singulas septimana vobis facere debeamus*

"we have to perform you the corvée every week"

(16) MED 172 (AD 778) *neque ego qui supra offertor* [--] *possit molestare aut resubtragere mea offerta*

"nor can I, the above-mentioned donator, [--] contest and reverse my donation"

In this study, *videri* is not interpreted as a passive or medial but as an evidential marker specifying the source of information when occurring in *videor/videtur* + main verb combinations.[214] *Videri* occurs exclusively in formulaic phrases, such as (17) and (18). The synthetic passive was a moribund category by the time of LLCT. Cases, such as (18), show that the forms *videor* and *videtur* were crystallised remnants of Classical Latin and were not always utilised according to classical grammar.

(17) MED 427 (AD 819) *res illa* [--] *ubi ipsi genitor meus avitare visus fuit*

"the house [--] where my father dwelled [= was known to have dwelled]"

(18) MED 195 (AD 784) *quantum habere videor* [= *videtur*] *rectore et guvernatore*

"as much as its rector and governor appears to have"

Note that clause (18) is classified as active and transitive with the direct object *quantum*. *Videri* is viewed as an evidential modal verb that reduces the reality degree of the verbal event (see Table 4.10.). Some related issues will be further developed in section 4.3. This and the previous chapters have prepared the ground for the numerical analysis of the LLCT data that will be performed in section 4.2.1. and section 4.2.2.

---

[214] Thompson 2002; Cuzzolin 2010, 255. On various semantic values of *videri*, see Orlandini 1996, 415–427.

4.2. Animacy, construction type, and subject case in LLCT

4.2.1. Numerical analysis of the whole corpus

In section 4.2., I shall examine whether and to what extent the two interrelated variables defined in section 4.1.1. and section 4.1.2., i.e. animacy and construction type, influence subject case selection. In this and the following chapter, the following two hypotheses, H_1^1 and H_1^2, will be tested.

H_1^1: The lower the animacy class of the subject is, the more often is the subject expected to occur in the accusative.

H_1^2: The less transitive the construction type of the subject is, the more often is the subject expected to occur in the accusative.

These hypotheses are formulated on the basis of the theoretical framework that I sketched in section 3.1. drawing heavily on the work of Rovai (2012b) and Cennamo (2009). This framework is what I mean whenever I refer to "theories of alignment change" in this and the following chapters. The framework observes the subject case marking from two different angles: the inherent semantic properties of the subject NP and the semantic properties of the verbal event. As for the subject properties, it is assumed that the extended accusative manifests itself first and foremost in the lowest animacy domains. Together with the animacy parameter, construction types approximate the control or primary responsibility of the subject over a given verbal process.

Hypotheses H_1^1 and H_1^2 are tested with two statistical methods: chi-square-based contingency tables and decision trees. Contingency tables are simple matrix-format tables that display the frequency distributions of the variables. Contingency tables are often called crosstabs because they are based on cross-tabulation of the variables on the *x*- and *y*-axes of the matrix.[215] Decision trees are a computer-assisted method of modelling the connections between the independent variables. I shall present the decision tree method in more detail below. Both cross-tabulation and decision trees exploit the chi-square test.

Chi-square test (χ^2) is an umbrella term of statistical hypothesis tests for nominal, i.e. categorical, variables. Categorical variables are variables that have two or more categories with no built-in or-

---

[215] E.g. Agresti 2007, 21.

dering of the categories. In the chi-square test, the sampling distribution observes a chi-squared distribution when the null hypothesis, i.e. the hypothesis that there is no relationship between two measured phenomena, is true. The null hypotheses for the above hypotheses H_1^1 and H_1^2 are H_0^1 and H_0^2:

H_0^1: The animacy class has no significant effect on the case form of the subject.

H_0^2: The construction type has no significant effect on the case form of the subject.

In sum, the chi-square test assesses how likely the observed differences between categorical data sets are due to chance. If the likelihood is low enough, it can be claimed that there is a significant dependence between the two variables. In this study, I shall utilise mainly Pearson's chi-squared test, which is the most common type of chi-square tests for large samples. Pearson's test is also the default statistic of the IBM SPSS CHAID application that will be exploited for the decision trees. My objective is to assess with the chi-square test whether the cross-tabulated observations of the two variables are independent of each other (so-called chi-square test for independence).[216]

Pearson's chi-squared test evaluates the significance of the dependence between the variables, but it does not tell how the individual observations contribute to this dependence. It is, however, often helpful to find out in which cells of the contingency table "something is going on". Therefore, I have decided to analyse the dependence between the variables in cross-tabulations by way of adjusted standardised residuals. The residuals are counted by subtracting the real frequency that is observed in the data from the expected frequency, i.e. the frequency that would occur when the relationship between two measured phenomena is due to the differing sample size only (null hypothesis). The residuals are then standardised in order for fractions of different size to be comparable. Conventionally, an adjusted standardised residual higher than 2.0 indicates that the number of cases is significantly larger than would be expected if the null hypothesis were true. In contrast, an adjusted residual that is less than -2.0 indicates that the number of cases is significantly smaller than would be expected if the null hypothesis were true. In other words, the adjusted residuals point to where the real associations are in a cross-tabulation with a significant chi-square.[217]

---

[216] E.g. Agresti 2007, 34–35.
[217] Agresti 2007, 38–39. See also 'Interpreting adjusted residuals in Crosstabs cell statistics' (https://www-304.ibm.com/support/docview.wss?uid=swg21479605 (accessed 2 April 2015)).

As said, I shall utilise the decision tree analysis (CHAID) to have a look at the dependence between the dependent variable, i.e. the case form of the subject, and the independent variables, which are animacy and construction type as well as the number of the subject (Table 4.2.). No hypothesis is formulated for number, as the theoretical framework does not predict anything concerning it. Thus, number will be examined heuristically as a byproduct in the animacy and construction type decision trees. Finally, I shall test the independent variables separately on the dependent variable in order to observe the individual behaviour of each variable.

Table 4.2. Dependent and independent variables with their categorical levels (section 4.2.).

Dependent variable		Independent variables	
Variable name	Categorical levels	Variable name	Categorical levels
case form of subject	nominative or accusative	animacy class	personal, animate, inanimate
		construction type	$A/S_A/S_O/S_O$ passive subject
		number of subject	singular, plural

By combining these results with theoretical reflection, I seek to assess whether one of the variables can be deemed the fundamental one and, in case it cannot be clarified, which one seems to have the most effect on the subject case selection in LLCT. In the next section, i.e. section 4.3., the construction type analysis will be further developed by taking under examination a new independent variable, i.e. transitivity, as defined by Hopper & Thompson (1980). The variables and their categorical levels discussed in this section are summarised in Table 4.2. All the variables are categorical, i.e. they can assume only a limited number of values.

The mentioned statistical tool that I shall call the 'decision tree' is the Exhaustive Chi-squared Automatic Interaction Detection tree (CHAID). CHAID is a multivariable analysis method for categorical variables and it is provided by IBM SPSS (version 22). The tool is based on a decision tree algorithm that calculates, by way of recursive classification, a statistic model of the sample from the given variables.[218]

[T]he essence of decision tree analyses is that they operate by partitioning the data into sets with the same values of a variable. They first find the independent variable that accounts for the largest majority of the variation in the dependent variable and partition the data into sets, called

---

[218] IBM 2011, 1.

branches, based on that variable. Each of the resulting branches is further subdivided based on the values of other independent variables until all of the data is accounted for. The result is a tree structure that indicates what independent variable or combination of variables is associated with particular values of the dependent variable.[219]

Decision trees illustrate which levels of the variables resemble each other and what kind of connections there are between the variables. The advantages of decision trees are that their output is highly visual and rather easy to interpret. Contrary to regression-based models, decision trees do not assess the significance levels of their rules. Decision trees just tell how often the model correctly applies to the information in the database, hence their liability to overgeneralise. Nevertheless, they are very helpful in pointing out interactions that need to be further explored.[220] In this study, I shall utilise decision trees specifically to see how the data split into statistically significant ($p < 0.05$) branches. These branches are then examined by analysing in more depth the case distribution on each categorical level of the independent variables. In the following decision trees, each branching node, i.e. parent node, has to have at least 20 occurrences of subjects and each branch, i.e. child node, at least 7 occurrences.[221] The chi-square and the level of significance (p-value) are calculated according to Pearson's test and the significance levels are adjusted using the Bonferroni correction.[222] In the following decision trees, the construction type is represented as levels A, S_A, S_O, and S_P, the latter standing for the S_O subjects of passive constructions.

---

[219] Eddington 2010, 267.
[220] Eddington 2010, 265–272; IBM 2011, 1–2; Magidson & Vermunt 2005, 1–3. Priiki 2014 and Priiki (in press) are good examples of how decision trees can help to detect successfully dependences in a relatively small linguistic dataset.
[221] IBM 2011, 9; Priiki (in press), 8.
[222] IBM 2011, 10.

Figure 4.1. Decision tree for subject case (dependent) and animacy, construction type, and number (independent).
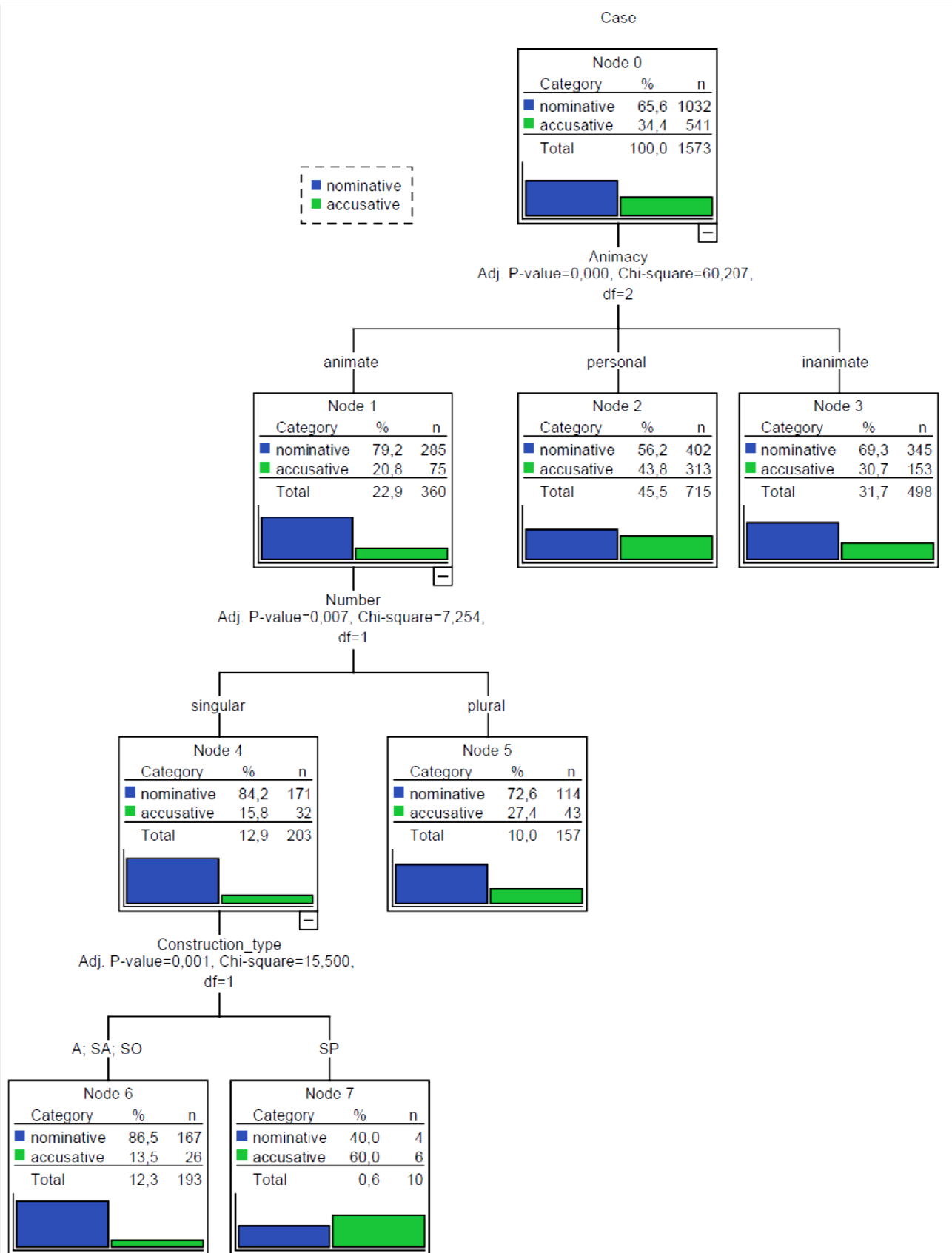
Figure 4.1. illustrates the connections between the case form of the subject and all the relevant subject-inherent variables, i.e. animacy, construction type, and number.[223] Node 0 presents the nominative/accusative distribution of the 1,573 basic query subset subjects of LLCT. A considerable number (34.4%) of the subjects are in the accusative. It is seen that animacy creates the first branch (nodes 1–3), and the chi-square of this first ramification is very high ($\chi^2 = 60.21$, $df = 2$, $p < 0.001$). Indeed, it is the animacy that is responsible for the most important split every time a decision tree is forced out of the entire basic query subset of LLCT.

Other ramifications turn up as well: the variable 'number' splits the class 'animate' further in a statistically significant way (nodes 4–5). However, its chi-square can be considered lower ($\chi^2 = 7.25$, $df = 1$, $p = 0.007$) than that of the animacy split. On the other hand, the comparison is not completely happy because the degrees of freedom (df) are not equal. Finally, even the singular animate subjects effect a statistically significant split by construction type (nodes 6–7). The relative significance order of the independent variables can be illustrated as follows:

(19) animacy < number < construction type

The class 'animate' is the only place where the variable 'number' partitions the data. No statistically significant split arises when the variable 'number' alone is applied to the whole corpus. This illustrates the central weakness of decision trees: the ramifications below the first branch do not give an overall picture of the data, but are restricted to the occurrences of the branching node. In the case of Figure 4.1., the variable 'number' applies only to the 360 animate subjects and nothing else, i.e. the dependency between the number and the subject case is not statistically significant in the other animacy classes. The splits of the lower branches, although they are statistically significant, are representative of only the branch they split. The plural animate subjects in Figure 4.1. seem to occur more often in the accusative than the singular subjects, albeit both the singular and plural rate lower than in the other animacy classes.

The generalisability of tree models can be estimated by a classification percentage that defines how large a portion of the occurrences of the data a given model can classify in correct classes. Thus, were the tree model generalised to the population that contains the observed sample, the value of

---

[223] The declension type of the subject was also taken into account as an independent variable. Declension did not, however, appear to be a statistically significant factor in any set of variables that have been tested, hence I have left it out of the analysis.

the dependent variable could be predicted to a certain extent.[224] The tree in Figure 4.1. could explain correctly 65.7% of the subject occurrences. The standard error (S) for this is 0.012. The objective of this study is not to predict, but to illustrate the data and the relations of the variables. Therefore, the classification percentages with standard errors will not be given for the decision trees of this and the following chapters.

As for the construction type, the split is based on only 203 animate singular subjects. The ramifications of this kind may sometimes reveal important dependencies that would remain otherwise unheeded, but usually they have little to give to the analysis of the whole data. The construction type split of the singular animate subjects is, indeed, rather unexpected, as the active A and S_A subjects align with the inactive S_O subjects. According to the theory, the S_O subjects of the unaccusative verbs and the S_P subjects of the passive verbs would be expected to pattern together. On the other hand, the opposition of the S_P subjects to all the other subject types can be thought to reflect a distinction between the active and passive voices.

In general, the pattern that is revealed by the first ramification in Figure 4.1. seems to contradict the theories of alignment change: the accusative percentage of the personal names is much higher than would be expected (43.8%). It is also difficult to tell why it is only variable 'animate' that splits into statistically significant categories in Figure 4.1. It may be symptomatic of the fact that at least the class 'personal name' involves a particularly high number of distracting factors (mainly the 2nd declension personal names that will be discussed below). Unlike number, the construction type effects a split in the whole data if animacy is excluded. This can be seen in Figure 4.2.

---

[224] IBM 2011, 27; Priiki (in press), 8. See also Eddington 2010, 269; Lewis 2000; Magidson & Vermunt 2005.

Figure 4.2. Decision tree for subject case (dependent) and construction type (independent).
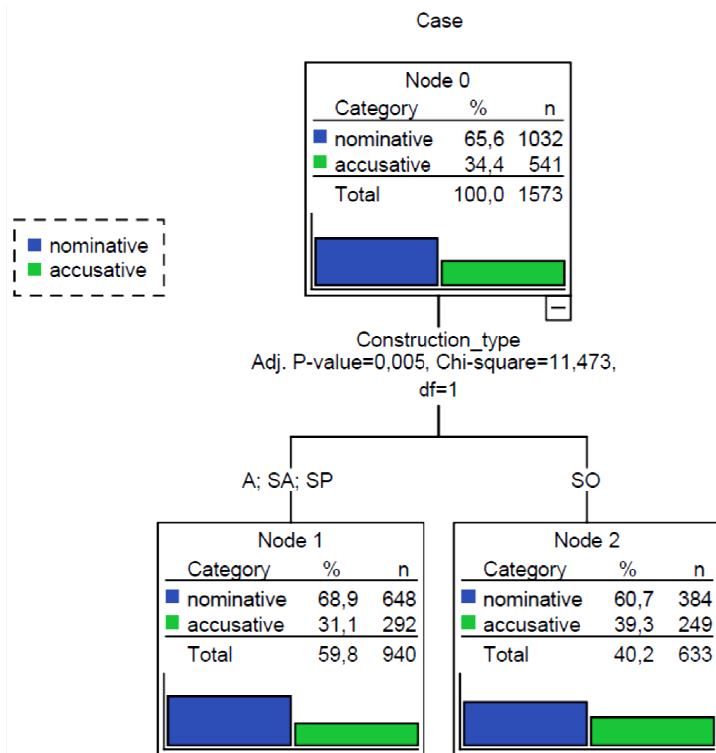
Figure 4.2. shows that the construction type splits the whole data in a manner different from that of the animate subjects above: now the branches contrast S_O against A, S_A, and S_P (*sic*). The joint accusative percentage of the A, S_A, and S_P subjects (31.1%) is lower than that of the S_O subjects (39.3%). This is surprising because, as stated above, the theory predicts that the S_O subjects of the unaccusative verbs and the S_P subjects of the passive verbs align with each other and contrast with the A and S_A subjects. All this suggests that a more detailed examination of the passive and, on the other hand, of the personal names is desired. It can be assumed that it is the 2nd declension (singular) *o*-stem personal names that represent the major distracting factor in the above decision trees. In the analysis that is carried out below in section 4.2.2., the *o*-stem personal names and other controversial categories will be excluded from the analysis. The following pages will show which kinds of problems the presence of the *o*-stem personal names causes for the distributional patterns. As a prelude to this, I consider briefly what kind of accusative percentage distributions are likely to be expected when a corpus is thought to reflect a certain stage of alignment change.

According to Rovai (see section 3.1.), animacy or, better, inanimacy, is the privileged locus where the extension of the accusative can be most easily observed within a predominantly nominative/accusative aligned language. The reason for this is that low-animacy subjects have a weak control over the verbal event, i.e. they are typically S_O subjects. Consequently, they tend to adopt the

case form of O (accusative) instead of the case form of A (nominative). The high-animacy subjects, instead, are typically S_A or A subjects and adopt the case form of A (nominative) both in the nominative/accusative and semantically-based alignment, so they cannot be used as indicators of the realignment.[225] Moreover, there is recent typological evidence that animacy *per se* favours certain selection phenomena that have been commonly explained by other factors, such as discourse drive.[226]

In this study, both these indicators, animacy and construction type, are discussed as separate but interacting variables. So, if the case distribution of LLCT is semantically-motivated at least to some extent, it should reflect, however imperfectly, the animacy distribution of its subjects as well as their relative position on the transitivity continuum as represented by the construction types (cf. Figure 3.1. in section 3.2.2.).

Previous research has studied the alignment change of Latin by classifying and setting in chronological order sample sentences that have been collected from various sources across a long time span. These are then observed in the light of theoretical considerations that for their part derive from typological alignment studies. The only study that displays quantitative data is Rovai 2005. In general, the pattern extracted from the traditional Latin sample sentences seems to follow rather well the one that Rovai and Cennamo have predicted on theoretical grounds. Inanimate common names seem to be the first and the foremost to occur as accusative subjects and animate personal names the last and the fewest, animate common nouns appearing somewhere in between. Basing on this, the extension of the accusative is assumed to have departed from the low-transitivity domains from where it may have penetrated gradually into the higher-transitivity domains.

LLCT is the first and thus far the only material where the morphosyntactic alignment of (Late) Latin can be studied with corpus-linguistic and, hence, statistical methods. Because corpora of this kind have not been available, the following question has never before arisen: what is the distribution of the accusative subject expected to be like in different animacy or construction type categories at an arbitrary time point? The answer is, of course, related to how variation is understood. In this study, I define the chronological priority of a linguistic phenomenon as follows: if an initial chronological

---

[225] Rovai 2012b, 111–112.
[226] Haig & Schnell (in preparation) and Everett (2009, 5, 11–15) suggest that whether a full NP or pronoun/zero is selected as the subject depends heavily on the animacy degree of that subject: highly animate subjects are preferably non-lexical. See also Dahl 2000, 42–48. This view contests the theory of Du Bois (1987, 829–839), who proposed a discourse-based interpretation of ergativity, implying that as the A arguments tend to be discursively connected to given information, they are often realised as pronouns or zero. In contrast, the S and O arguments often introduce new referents and, thus, tend to be full NPs. This discourse explanation of ergativity has turned out to be poorly generalisable cross-linguistically and seems to be replaceable by an animacy-based explanation.

stage of a linguistic evolutionary process can be attested as feature *x* and if any later chronological stage of the same process appears as feature *y*, then at any time point during the process, the percentage of feature *x* must be higher than the percentage of feature *y*. In other words, a given phenomenon is likely to consolidate its position in domains where it has first appeared. Thus, its presence is relatively higher in that domain than in the domains where it appears only later. Consequently, the synchronic variational profile is by default different at the beginning of the process than at the end of that process.

Drawing on what was said, I find it justified to assume that if a corpus of an originally nominative/accusative aligned language is thought to reflect an ongoing semantic realignment, the following should apply at a certain time point: inanimate/S_O subjects display the highest accusative share while animate/S_A subjects display the second highest accusative share and the personal name/A subjects display the lowest accusative share. This model does not predict the absolute percentages that would be expected at a certain time point, but it predicts the order of those percentages. However, the LLCT accusative percentages of the below tables are all more or less on the same level (about 20 to 40%), which seems unexpected: the differences between the percentages are not particularly striking considering that, theoretically, it could even be possible to find a distribution of type A 0% – S_A 50% – S_O 100%.

It is worth discussing what kind of accusative percentages should be reasonably expected. It is obvious that real-world situations do not elicit such percentage distributions as the one presented above. It is, instead, necessary to take into account the variational nature of language change and the mechanics of how language change is projected in written medium, the latter being likely to reflect several mutually conflicting linguistic and stylistic tendencies that date back to different diachronic stages. Moreover, the classificatory concepts, such as animacy degree or construction type, do not model the real world in an exact manner, but are mere tools conceptualised by the researcher.

To summarise, what is essential in interpreting percentage distributions in a material that is supposed to reflect alignment change are the systematic differences in size between the levels of that variable. I will proceed now to presenting the contingency tables for the chi-square tests that have been conducted on the variables defined in the above hypotheses H_1^1 and H_1^2. Table 4.3. presents the dependence between subject case and animacy in LLCT.

Table 4.3. Dependence between the case and animacy of subject in LLCT.

Case		Animacy			Total
		inanimate	animate	personal	
nominative	N	345	285	402	1,032
	%	69.3%	79.2%	56.2%	65.6%
	residual	2.1	6.2	-7.2	
accusative	N	153	75	313	541
	%	**30.7%**	**21%**	**43.8%**	34.4%
	residual	**-2.1**	**-6.2**	**7.2**	
Total	N	498	360	715	1,573
Chi-square		$\chi^2 = 60.21$, $df = 2$, $p < 0.001$			

The accusative percentages of the inanimate and animate subjects of Table 4.3. seem to comply with the postulated semantically-based case alignment: the percentage of the inanimate subjects (30.7%) is higher than that of the animate subjects (21%). It is, however, the high accusative percentage of the personal names (43.8%) that breaks the pattern. The personal names would have been expected to feature the lowest accusative percentage of all. All the adjusted standardised residuals surpass the cut-off of ±2.0, i.e. all the levels of the independent variable are important, hence the tripartite division of animacy in Figure 4.1. (cf. below Figure 4.4.). The higher residuals of the animate and personal name subjects (-6.2 and 7.2. for the accusative) suggest, however, that these two categories are responsible for the high dependence between the dependent and independent variables.

In Table 4.3., singular and plural have been treated together because the classification tree in Figure 4.1. showed that number is not a statistically significant principle of distribution for the entire corpus. However, number did effect a split with animate subjects in Figure 4.1. Appendix 4.1. demonstrates that when the subjects of Table 4.3. are studied separately in the singular and the plural, only the singular shows a statistically significant dependence between case and animacy ($p < 0.001$) while the plural shows no such dependence ($p = 0.18$). Indeed, the dependence observed in Table 4.3. is exclusively due to the singular subjects. The perceived pattern, however, is practically the same whether singular and plural are treated together or not: the personal names show the highest accusative percentage against expectations.

Before turning to the analysis of the distracting factors, I shall have a look at the second semantic variable, i.e. construction type. Table 4.4. presents the dependence between the case of the subject and the construction type of the clause in LLCT.

Table 4.4. Dependence between the case of the subject and the construction type in LLCT.

Case		Construction type				Total
		A	S$_A$	S$_O$	S$_O$ pass.	
nominative	N	501	65	384	82	1,032
	%	67.9%	75%	60.7%	71%	65.6%
	residual	2.2	1.9	-3.2	1.3	
accusative	N	237	22	249	33	541
	%	**32.1%**	**25%**	**39.3%**	**29%**	34.4%
	residual	**-2.2**	-1.9	**3.2**	-1.3	
Total	N	738	87	633	115	1,573
Chi-square		$\chi^2 = 13.41$, $df = 3$, $p = 0.004$				

The S$_O$ subjects of unaccusative constructions display the highest accusative percentage (39.3%). This matches well the assumption that the extension of the accusative began from the low-transitivity domains where the subjects are [-Agentive] and their relation with the verb characterised by [-Control]. On the other hand, the accusative percentage of the A subjects (32.1%) is higher than those of the S$_A$ subjects of the unergative constructions (25%) (and the S$_O$ subjects of the passive constructions (29%)), which is contrary to what was expected. The theory suggests that the extended accusative penetrated into the A subjects of transitive clauses only very late when the semantically-based active/inactive and even the ergative/absolutive stage of the realignment had passed and the case opposition was being neutralised for good. There is also a difference in size (about 10 percentage points) between the S$_O$ active and S$_O$ passive subject percentages (39.3% and 29%, respectively). This is contrary to what I suggested in section 4.1.2.: the S$_O$ subjects of the passive and the S$_O$ subjects of the active unaccusative constructions can be expected to pattern on a par.[227]

The high accusative share of the A subjects can again be explained by the behaviour of the personal names, which are understandably frequent as the subjects of transitive clauses. I shall return to this in section 5.3. The dependence between case and construction type in Table 4.4. is statistically significant ($p = 0.004$), but treating the occurrences separately for singular and plural in Appendix 4.2. reveals that the statistical significance is again limited to singular subjects only ($p = 0.031$). In plural, there is no significant dependence ($p = 0.109$), and the situation resembles in this respect the one observed for subject case and animacy. As for the singular, the pattern of Appendix 4.2. is almost identical to that of Table 4.4. The adjusted standardised residuals -2.2. and 3.2 reveal that it is

---

[227] This kind of pattern is, indeed, observed in section 4.2.2. (Table 4.6.).

the A and the S_O subjects that are really responsible for the bipartite division of the construction type in Figure 4.2.

In this chapter, it has often been stated that animacy and construction type are undoubtedly connected with each other. Figure 4.3. proves this explicitly. There is an extremely high dependency ($\chi^2 = 454.03$, $df = 6$, $p < 0.001$) between construction type and animacy in LLCT. It is, however, not only LLCT or Latin where low-animacy subjects occur particularly in intransitive constructions, but this has been observed cross-linguistically.[228]

Figure 4.3. Decision tree for construction type (dependent) and animacy (independent).

Subject_type

Node 0		
Category	%	n
A	46,9	738
SA	5,5	87
SO	40,2	633
SP	7,3	115
Total	100,0	1573

Animacy
Adj. P-value=0,000, Chi-square=454, 026, df=6

inanimate

Node 1		
Category	%	n
A	15,5	77
SA	1,6	8
SO	65,3	325
SP	17,7	88
Total	31,7	498

animate

Node 2		
Category	%	n
A	62,5	225
SA	14,2	51
SO	19,2	69
SP	4,2	15
Total	22,9	360

personal

Node 3		
Category	%	n
A	61,0	436
SA	3,9	28
SO	33,4	239
SP	1,7	12
Total	45,5	715
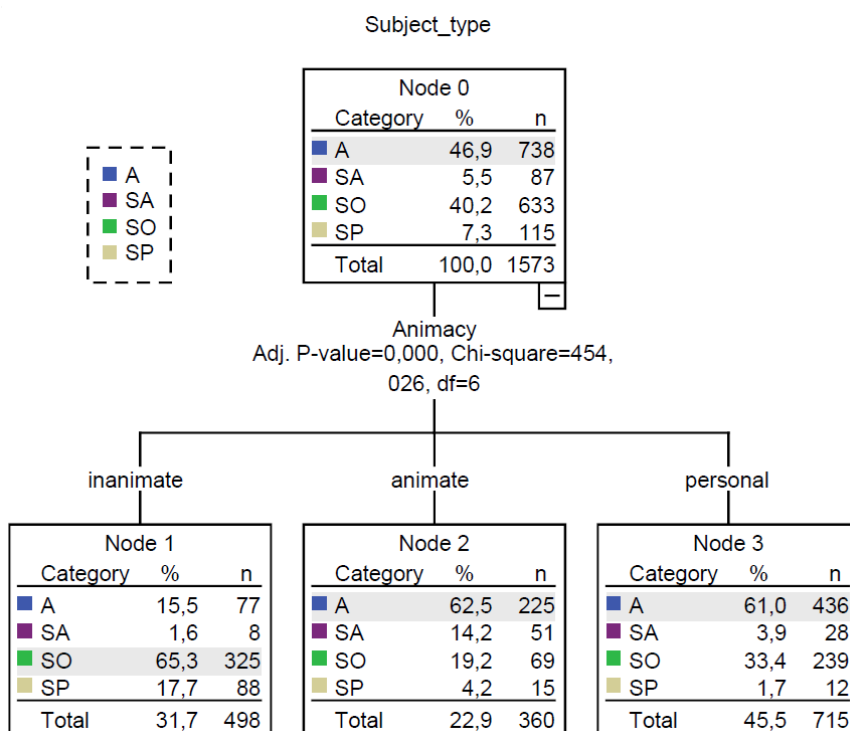
The construction types are distributed rather predictably within the animacy classes. The A subjects draw mainly on animate (62.5%) and personal name nouns (61.0%), while the S_A subjects appear to be particularly often animate nouns (14%). Unsurprisingly, the S_O subjects are often inanimate nouns (65.3%) but, on the other hand, one third of the personal names (33.4%) occur as S_O subjects. This may be due to the high frequency of copula structures with the verb *esse* (see section 4.4.1.). The passive S_O subjects are most often inanimate nouns (18%), which makes sense, as the subjects of passive constructions are demoted objects of active constructions. Animacy and construction type are, thus, tightly interrelated, and one can ask which of them is of primary importance for the

---

[228] Everett 2009, 1, 11–15; Haig & Schnell (in preparation).

case marking of the subject. Or should they both be considered manifestations of some more general common variable, such as control (section 3.1.)? Of course, this cannot be decided on the basis of statistical analysis; the data must be considered in the light of theory, instead. I shall return to this in section 6.1.

To summarise, all the above observations made in this chapter point to an atypical behaviour of personal names, namely their accusative percentage is higher than expected. On the other hand, 91.6% of the personal names of LLCT are 2nd declension names. Therefore, it is necessary to reconsider the status of the ending *-o*, which has been interpreted as accusative-based in this study (section 2.3.3.). In the following chapter, the data of LLCT are examined without the *o*-endings and without any other possibly controversial endings, such as the *e*/*em*/*i*-endings of the 3rd declension singular parisyllabic nouns. As far as subject case is concerned, the only language-historically reliable forms seem to be the 2nd declension plural nouns (*-i* vs. *-os*) and the 3rd declension singular imparisyllabic nouns, which have different roots in the nominative and accusative (*par-s* vs. *part-em*). Since the former are quite rare, I have chosen to create a subcorpus from the latter.

4.2.2. Numerical analysis of the 3rd declension singular imparisyllabic subjects

This section applies the same statistical tests that have been made in the previous section to the subcorpus that consists exclusively of 3rd declension singular imparisyllabic subjects. The 3rd declension singular imparisyllabic subjects are 392 in LLCT. Here, even the parisyllabic nouns of type *pater* are included because their stem changes: *pater* vs. *patr-em*.[229] The potential levelling caused by phonological evolution does not affect these nouns and, therefore, they can be thought to indicate reliably either the nominative or the accusative form. Figure 4.4. and Figure 4.5. show respectively the dependence between the subject case and animacy and between the subject case and construction type in 3rd declension singular imparisyllabic subjects.

---

[229] The parisyllabic nouns, such as *curtis* vs. *curtem* 'courtyard' and *Iohannes* vs. *Iohannem*, have been excluded because it is possible that the evolutionary end-products of the endings *-is*, *-es*, *-em* have been merged occasionally due to the weakening of word-final sounds. Note, however, that everywhere else in this study I have considered even the endings of the parisyllabic 3rd declension nouns to be distinctive (see section 2.3.).

Figure 4.4. Decision tree for subject case (dependent), animacy, and construction type (independent) in the subcorpus.
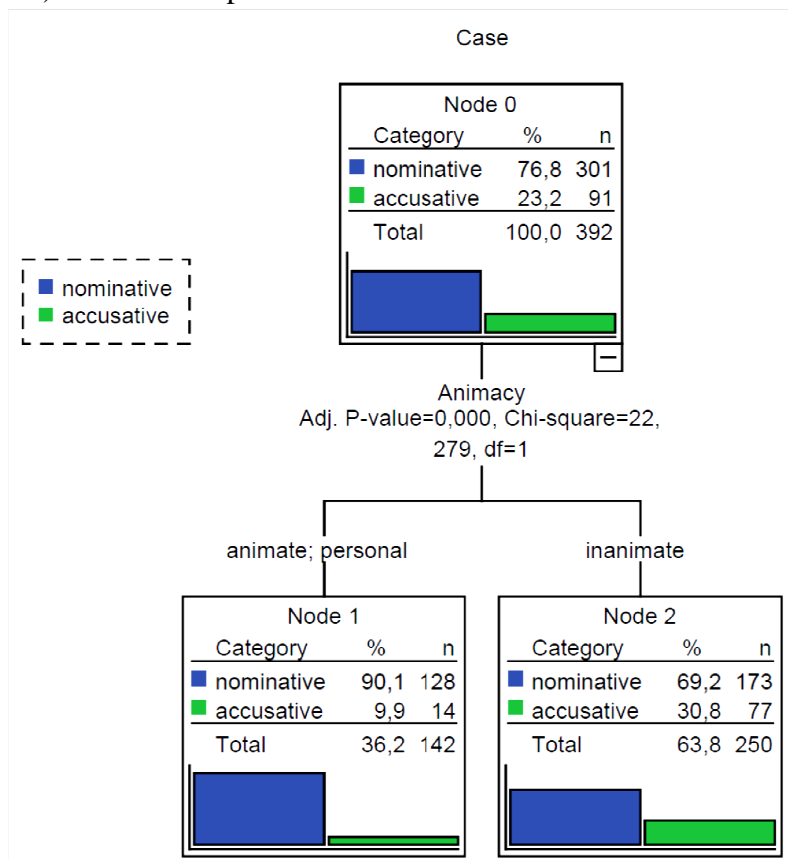
Figure 4.4. shows that 23% of the subjects are in the accusative in the subcorpus. The first split is again caused by animacy. This proves that animacy is the strongest distinguishing factor in the subcorpus, just as it is in the whole LLCT. Animacy provokes the first split in the decision tree whenever it is included. The chi-squares in Figure 4.4. ($\chi^2 = 22.28$, $df = 1$, $p < 0.001$) and in Figure 4.5. below ($\chi^2 = 15.16$, $df = 1$, $p = 0.001$) suggest that the construction type is secondary as the *explanans* of the subject case selection in LLCT: its chi-square is lower than that of the animacy split.[230] In contrast to the whole LLCT, in the subcorpus the animate subjects and personal name subjects align with each other and display a remarkably low joint accusative percentage (9.9%) while the inanimate subjects are opposed to them with an accusative percentage of 31%. The tight connection of the animate and personal name subjects is further discussed in section 4.3. and section 4.4.

---

[230] The respective chi-squares cannot be compared in Figure 4.1. and Figure 4.2. This is because in Figure 4.1. the construction type split is based exclusively on the animate singular subjects and not on the whole data. Second, the animacy split of Figure 4.1. ($\chi^2 = 60.21$, $df = 2$, $p < 0.001$) and the construction type split of Figure 4.2. ($\chi^2 = 11.5$, $df = 1$, $p = 0.005$) are not comparable because the degrees of freedom (df) are different. Both the chi-squares are still very high.

In Figure 4.4., the nodes that are split by animacy do not further branch by construction type. This is probably because animacy and construction type are so tightly interrelated (see Figure 4.3.) that a split between personal name/animate and inanimate is, in this subcorpus, practically the same thing as the split between A/S_A and S_O/S_O passive. After the animacy split, the effected nodes no longer contain enough variation for a statistically significant construction type split.

Figure 4.5. Decision tree for subject case (dependent) and construction type (independent) in the subcorpus.
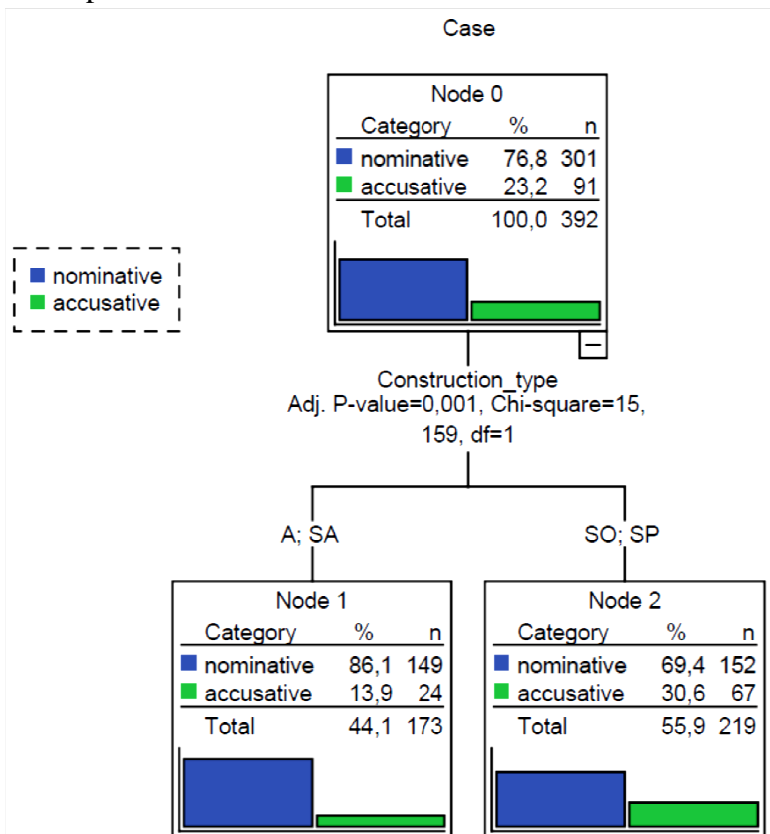
If the dependence is examined only between the subject case and the construction type, as in Figure 4.5., a completely theory-compatible split is obtained: the split aligns the A and S_A subjects as opposed to the S_O and passive S_O subjects. This time the S_O subjects of passive constructions behave as postulated: they align with the other S_O subjects and not with A and S_A, as they did in the decision tree in Figure 4.2. Now, a closer look at the data is needed in order to see the accusative percentages of each category.

As the decision trees group the categories into statistically significant nodes, it is not possible to see the accusative percentages of the categories separately. Therefore, I present here Table 4.5. and Ta-

ble 4.6., which give the accusative percentages for the subcorpus in the same manner as Table 4.3. and Table 4.4. gave them for the whole corpus in the preceding section.

Table 4.5. Dependence between subject case and animacy in the subcorpus.

Case		Animacy			Total
		inanimate	animate	personal	
nominative	N	173	118	10	301
	%	69.2%	90.1%	91%	76.8%
	residual	-4.7	4.4	1.1	
accusative	N	77	13	1	91
	%	**31%**	**10%**	**c.10%**	23%
	residual	**4.7**	**-4.4**	-1.1	
Total	N	250	131	11	392
Chi-square		$\chi^2 = 22.28, df = 2, p < 0.001$			

Table 4.6. Dependence between subject case and construction type in the subcorpus.

Case		Construction type				Total
		A	S_A	S_O	S_O pass.	
nominative	N	124	25	123	29	301
	%	85.5%	c.90%	68.7%	c.75%	76.8%
	residual	3.1	1.6	-3.5	-0.7	
accusative	N	21	3	56	11	91
	%	**15%**	**c.10%**	**31%**	**c.30%**	23%
	residual	**-3.1**	-1.6	**3.5**	0.7	
Total	N	145	28	179	40	392
Chi-square		$\chi^2 = 15.16, df = 3, p = 0.001$				

Table 4.5. shows that now the accusative percentage pattern complies with what is postulated by the theory: the personal names stick most firmly to the nominative (c.10% accusative) while the 31% of inanimate subjects is as high as it was in Table 4.3. for the entire LLCT data (30.7%). The 10% of the animate subjects is, however, of the same size as the c.10% of the personal names, although the animate subjects might be expected to show a higher accusative share than the personal names. Note, however, that the 3rd declension imparisyllabic personal name subjects are only 11. In Table 4.3., the percentages of animate and personal name subjects of the whole corpus were 21% and 43.8%, respectively. The adjusted standardised residuals 4.7 and -4.4 of Table 4.5. reveal the importance of inanimate and animate subjects in the bipartite animacy split attested in the decision tree of the subcorpus (Figure 4.4.). In the subcorpus, the personal names do not have a statistically significant role *per se*, as they had in Table 4.3. above (residual 7.2).

Table 4.6. indicates that the construction type also shows a nearly theory-compatible distribution, which is, in fact, an approximate replica of the animacy distribution. The S_O subjects of the unaccusative and passive constructions occur both clearly more often (15 to 20 percentage points) in the accusative than the A and S_A subjects. On this basis, it is justified to treat the S_O subjects of the unaccusative and passive constructions together, as Rovai, indeed, does in his studies. The theory postulates that the accusative percentage of the A subjects should be the lowest one, but here it is the S_A subjects with their c.10%. This will be explained in section 5.3.2., where the S_A subjects of LLCT turn out to be particularly high in transitivity. Of course, some of the frequencies are again rather low as well. The residuals -3.1 and 3.5 show that it is again the A subjects and the S_O subjects that are mainly responsible for the bipartite construction type split in Figure 4.5.

As an interim summary, the 3rd declension singular imparisyllabic subjects seem to follow the semantically-based case marking system to a certain extent. As was stated in the preceding section, an A 0% – S_A 50% – S_O 100% distribution is not realistic. Instead, it is realistic to expect a distribution that displays systematic, statistically significant, and theory-compatible differences between the accusative percentages of the categories of the independent variable. This is in effect what is seen both in Table 4.5. and Table 4.6. Contrary to what the statistics of the entire LLCT suggested in the preceding section, now even the S_O subjects of passive constructions fall into place.

As stated earlier, the 3rd declension imparisyllabic nouns are the locus where the alignment change can be most reliably detected. In theoretical terms, alignment splits can occur anywhere in a system, even between different declension classes. However, declension splits of this kind are not yet attested in any language, as far as I know. Therefore, until proved otherwise, I shall assume that once the semantically-based case marking or its residues are attested in the 3rd declension subcorpus of LLCT, this marking is likely to have been extant in the whole declension of LLCT as well. What impedes us from perceiving its extension properly is that late phonological changes make it impossible to discern the nominative forms from accusative forms in several other categories of the Latin declension that distinguished the mentioned case forms in Classical Latin.

Of course, categories in which the nominative and accusative endings have possibly merged into each other can be thought to reflect an already neutralised case opposition.[231] Likewise, the presence of the accusative in the S_A subjects seems to hint at an ergative/absolutive alignment. The first-mentioned neutralisation interpretation is problematic once it is proved that a systemic case contrast still existed somewhere in the Latin declension. If the speakers of Late Latin still understood that there are different case forms for different semantico-syntactic functions in some declensions, I find it possible that they perceived the same functional opposition even in those declensions where phonological levelling had abolished the morphological contrast between subject and object endings. All this boils down to the question of how many morphologically contrasting categories there must be in order to keep the functional opposition alive in the whole system. The evidence of Old French suggests that a surprisingly reduced presence of contrasting morphological marking was able to maintain a two case system for centuries.[232]

From this viewpoint, it remains plausible to ask to what extent, say, an individual *o*-ending was perceived as representing a nominative instead of an accusative. In my opinion, this section has shown rather clearly that one of the phonologically disabled categories must be the 2nd declension *o*-ending. This category has particularly important consequences in LLCT because charter Latin involves a considerable number of 2nd declension personal name subjects (655 in total). The relatively clear appearance of the semantically-based alignment in the 3rd declension singular imparisyllabic subjects suggests that the reason which prevents a similar pattern from arising from the entire LLCT is the disturbing effect of the 2nd declension personal names and of other morphologically ambiguous declensions.

In section 4.3. as well as in chapters 5 and 6, it is unfortunately seldom possible to treat the 3rd declension subcorpus separately because the required chi-square tests cannot be conducted on such a small sample ($N = 392$). A larger corpus (with possibly more advanced statistical tests) would be needed. My strategy will be as follows: in the remaining part of this work, I shall examine the morphological realignment mostly on the basis of the whole LLCT corpus, but whenever it is statistically possible, I shall utilise the 3rd declension subcorpus as a comparison material. It will turn out, for example in section 5.3.2. (Table 5.6.), that the entire LLCT and the 3rd declension subcorpus pattern

---

[231] Note, however, that it is not completely certain whether the morphophonological change that led to the merger is thoroughgoing in these categories. For example, the difference between the 2nd declension *-o*, *-um*, and *-us* may have been audible when pronounced carefully. This will be discussed later on.
[232] Sornicola 2011, 27–28; Schøsler 1984, 234, 240–242.

are essentially alike. To conclude section 4.2., I shall discuss briefly the 2nd declension personal names and the *o*-ending in the following excursus.

4.2.3. 2nd declension personal names

Personal names, as well as place names, reveal several interesting phonological and linguistic phenomena that are only scarcely visible in other categories of LLCT. This is because names do not belong to the basic vocabulary, but must be learned separately. Novel names did not become orthographically established in the Late Latin period because there was no unifying or standardising authority to establish them. This is especially true for the personal names of Germanic origin that were the most popular name types in 8th/9th-century Tuscany. The Germanic names differ from the Latin-based names in that they did not have a traditionally codified spelling, whereas many Latin-based names had occurred in the written code for centuries and, therefore, the scribes knew how to spell them. In this connection, it is possible to discuss the name category of LLCT only to the extent that it seems to be related to case marking.

This section will discuss in some length singular subjects that end in *-o/u/e/i* (2nd and 3rd declensions) and are tagged as accusatives in LLCT. It is necessary to say first a few words about how I see the status of these 'Romance-type' endings. I am aware that it is, sometimes justifiably, doubted whether forms that result from phonetic erosion of classical case forms can anymore be considered "accusative" or "nominative".[233] However, this attitude bans *a priori* any research on the topic. The Romance-type nouns in, say, *-o* or *-e* are often viewed as outsiders of the traditional Latin case system or as some kind of "neutral" forms. But are they really such in those *Latin* texts that are still proved to maintain a nominative/accusative case contrast? It is obvious that these forms become "neutral" as soon as the case contrast ceases to exist, i.e. in the Romance languages. In the meanwhile, I consider it important to examine what kind of functional domains the Romance-type forms can occupy in Latin *before* the neutralisation of the case contrast.

It was noticed in the preceding chapters that personal names, and most likely the personal names with the 2nd declension singular *o*-endings, apparently cause unexpected behaviour in the accusative subject distribution when the entire basic query subset of LLCT is examined. If all the subjects with

---

[233] See e.g. Sornicola (2011, 35) who refers to D'Ovidio and Schuchardt.

the 2nd declension singular *o/u/um*-endings[234] (748) as well as all other singular and plural subjects (433) whose endings are not completely reliable are excluded from the analysis, only the 392 3rd declension singular imparisyllabic subjects remain. It was stated in the previous section that, with the latter, the accusative distribution across the animacy classes and the construction types is more or less what could be expected of an ongoing semantically-motivated realignment (Table 4.5. and Table 4.6.). The open question is whether the distracting effect observed in LLCT is due to the 'nameness' of the 2nd declension singular personal names or to their *o/u/um*-endings. This can be examined by comparing the accusative percentages of the different construction types of the 2nd declension singular subjects of LLCT.

Table 4.7. The 2nd declension singular subjects of LLCT.

	Accusative		Nominative		N
Inanimate	11	c.50%	10	c.50%	21
Animate	19	26%	53	74%	72
Personal	291	44.4%	364	55.6%	655
$\sum$	321	42.9%	427	57.1%	748
Chi-square	$\chi^2 = 9.41$, $df = 2$, $p = 0.009$				

Table 4.7. shows that the inanimate and animate 2nd declension singular subjects present a some-what predictable accusative distribution: inanimate subjects have a bigger accusative share (c.50%) than the animate subjects (26%), although both the percentages are higher than the corresponding percentages in the 3rd declension subcorpus (31% and 10%, respectively; Table 4.5.) or even in the entire LLCT (30.7% and 21%, respectively; Table 4.3.). It is only the personal names that stick out with their 44.4%. This percentage is almost the same as the personal name percentage in the entire LLCT (43.8%, Table 4.3.), which is to be expected because the 2nd declension personal names amount to 91.6% of the personal names of LLCT.

I propose the following twofold interpretation of the above percentages: the *o/u/um*-ending and the 'nameness' are likely to be together responsible for the elevated accusative percentages of the 2nd declension singular subjects of LLCT. The fact that all the accusative percentages are higher than the respective percentages in the 3rd declension subcorpus or in LLCT on average is probably due to the nominative/accusative ambiguity of the *o/u/um*-ending: some of the *o/u/um*-endings are likely to derive from the accusative ending *-um* and some from the nominative ending *-us*. As is well known,

---

[234] The accusative ending distribution of the 2nd declension singular subjects is: *-o* 86%, *-u* 12%, *-um* 2% (see Table 4.8.).

some researchers consider the *o*-ending a phonologically motivated merger of the accusative and nominative endings (or a fusion of accusative, nominative, ablative, and dative endings).[235] My interpretation supports partly the nominative/accusative merger view, but emphasises that a part of the *o/u/um*-ending nouns were intended to be one of the two case forms, nominative or accusative, not an undifferentiated, Romance-type form that represented their merger. This is related to what was said above of the extant functional contrast between nominative and accusative cases.

According to what was just proposed, the boost caused by the ambiguity of the *o/u/um*-ending is likely to count for about 15 to 20 percentage points of the accusative percentages of the inanimate and animate subjects in Table 4.7. (compared to the 31% and 10% of Table 4.5.). The personal names, instead, seem to contain as many as 35 percentage points of extra accusatives (compared to the c.10% of Table 4.5.) due to the *o/u/um*-ending but also due to their being personal names. I seek to validate this suggestion in the following passages.

I will begin by asking why personal names would show a tendency against semantic alignment (as far as there is such). It has already been stated that personal and place names tend to reflect phonological evolution better than the basic lexicon that was supported by the written tradition and school teaching, whatever that teaching may have been. When a scribe was confronted with a new word, for example a name that he had never seen written down, he had to transcribe what he heard with the writing system of (Classical) Latin. It is, however, probable that this struggle with the unknown affected case endings only secondarily.

Names seem to behave against expectations even in several early non-literary Latin texts. Adams discusses the occasional non-inflection of names found in non-literary materials, such as papyri, inscriptions, and curse tablets. Most of his examples are about the nominative form being the non-inflected one, i.e. the base or default form, that tended to be realised in the 'name slots' of formulaic texts (see below) although the syntactic context would have required some semantic case. In regard to these occurrences that date back to well before the 8th and 9th centuries (mainly the 1st to 4th centuries AD), Adams concludes that "the nominative was felt to be the essence of the name". He underlines that this nominative does not reflect any collapse of the inflectional system in his material:

---

[235] For the nominative/accusative view, see Herman 1987, 104–105; Adams 2013, 136–147, 201; for the fusional view, see Vielliard 1927, 193; Pei 1932, 225–226; Pei 1937, 241–246, 264–267; Gaeng 1990, 116–117; Penny 1980, 501–509; Banniard 1992, 518; Penny 2002, 116. See also Sornicola 2011, 22, 33–35; Väänänen 1981, 116–117. The 2nd declension singular subjects with *o/u/um*-ending were included in the basic query subset in order to make it as comprehensive as possible (see section 2.3.3.).

it is only the personal names that are uninflected.[236] Adams's claim makes perfect sense in a linguistic context where the default case was still the nominative. Since the accusative replaced the nominative as the default case sometime in Late Latin, the accusative case and the preferred construction type of the personal name, i.e. the A subject (see Figure 4.3.), conflicted. At least in the restricted charter context, the 'nameness' of the name was likely to prefer the current default case, i.e. the accusative, although its agentive function favoured the nominative.

I shall next discuss some aspects that contribute to the discussion concerning the default case as the preferred case of personal names in non-literary texts, especially charters. I shall examine first in more detail the already-mentioned role of the formulaicity and, then, discuss two more phenomena that are connected with each other: the intimate nature of personal names and the exigency of clarity, i.e. the wish not to leave room for misunderstandings concerning the persons occurring in a document. This discussion is indebted to Adams 2013 in many respects.

As mentioned earlier, charter formulae have 'slots' that are more or less mechanically filled in with the names of the persons involved in the legal act. Adams cites examples of the *Tablettes Albertini*, Vandal North African documents from AD 493–496, that present several formulaic phrases with such slots for personal names. As the nominative was obviously felt to be the proper form of names at the time, the nominative was the case form that often ended in these slots regardless of the syntactic context. This, indeed, seems to be what has happened in the common subscription formula in (20). According to Adams, the incorrect nominative thus sometimes represents only a careless filling in of forms.[237] Filling in slots is, of course, what is also done in LLCT even though the formulae do not seem to be as rigid in LLCT as they were in the *Tablettes Albertini*. The difference is, however, that in LLCT the form that ends in the slot is no longer the nominative but the accusative (or the *o/u/um*-ending form in the 2nd declension singular). Example (21) cites a subscription from LLCT where it is exceptionally the nominative that fills in the slot while (22) presents the normal case where the fill-in form is a 2nd declension form in -*o*.

   (20) *ego Lucianus petitus a <u>Maxinus benditor</u>* (*Tablettes Albertini* IX.24)

   "I, Lucianus, [who was] asked by <u>Maxinus, the seller</u>"

---

[236] Adams 2013, 205–207, 211–215. About the vocative as the origin of certain fossilised names, see Wackernagel 1926, 297; Havers 1928, 104; Havers 1931, 94–95.

[237] Adams 2013, 213–215; Courtois & al. 1952.

(21) CDL 94 (AD 748) *signum + manus <u>Theopigntus arcidiaconus</u> testis*

"sign + [= Holy Cross] of the hand of <u>Theopigntus, the archdeacon</u> and witness"

(22) MED 503 (AD 829) *tali ordine ut ego <u>nominato Fusciano</u> vel meus heredes suprascripta vinea bene lavorare et gubernare seo ipsa terra vacua propaginare et in omnibus meliorare debeamus*

"with the regulation that I, the <u>mentioned Fusciano</u>, or my heirs must well cultivate and govern the above-mentioned vineyard and plant the vacant plot and ameliorate [it] in all respects"

On the other hand, it must be remembered that these slots were not slots in the modern sense: blanks in an otherwise prefilled paper or electronic form. In the context of LLCT, Adams's slots are to be understood as imaginary by nature, as the scribes produced the whole charter texts by memory. The mechanical aspect is likely to have been less pronounced in this kind of mental process. What is essential here is that the slots represent syntactic positions where something "out of the ordinary" had to be placed. This "out of the ordinary" had to be retrieved outside of the formula repertoire and, thus, required more mental processing than the normal reproduction of the repetitive formula. As a consequence, the syntactic connection of the fill-in name with the verbal nucleus of the phrase is obfuscated and the syntactically neutral default case results. Combined with Adams's observations on the older non-literary material, it would be easy to conclude that the nominative was felt to be the default form of nouns at least in the times of the *Tablettes Albertini* but, by the time of the LLCT charters, the accusative-based form (-*o/u/um* in the 2nd declension singular) was likely to have become the default form.[238]

Regarding the special status of personal names, Adams mentions that identifying the person in question was certainly very important in curse tablets in which the writer wanted to be sure of the identity of the victim of the curse. Adams quotes, indeed, some extra-syntactic nominative-form names that occur in curse tablets.[239] This observation can be generalised to the scribes' need to identify explicitly the persons involved in the redaction of a charter. I shall discuss this view in connection with the intimacy that is always present when quoting a living person's name. In charter Latin, one's personal name is a category of elevated importance. A charter is a legally binding document that registers a transaction in which certain economic interests are involved by definition. The validity of the charter as well as of the whole transaction requires that all the details are correct. This

---

[238] On the default case and related bibliography, see section 3.1. and section 6.2.
[239] Adams 2013, 212; Havers 1928, 96–97.

calls for the greatest possible identifiability of the transferred property and of the parties involved. Thus, not only the location and boundaries of each plot of land must be precisely indicated, but also the rogator, the purchaser, and the witnesses must be unambiguously identifiable.

With personal names, identifiability can be enhanced by epithets, such as *vir religiosus* and *clericus* or *filius quondam Dondoloni* in (23), as is the norm in LLCT. In this respect, it is possible that, at least in some cases, the meticulous phonetic spelling of the names may have been considered an important means of being precise.[240] It cannot be ruled out, however, that in certain other cases the opposite tendency of using classicising or etymologising spelling was favoured.

> (23) CDL 45 (AD 730) *consta me Candido u*(*ir*) *r*(*eligiosus*) *cliricus filius quondam Dondoloni*
> [--]
> "it is agreed upon that I, Candido, vir religious, son of the late Dondoloni [--]"

I find it plausible that the abundant use of the *o*-ending with the 2nd declension singular personal name subjects is partly related to this need of precision. Most scribes probably felt that it was their duty to indicate the identity of the participants of a legal act as clearly as possible: some of them choose to do that, perhaps, by providing a phonetic spelling of the names. This implies, of course, that the normal spoken form of these 2nd declension personal names ended in *-o*, just as it ends in Italian. This does not undermine the morphophonological conclusion that was made on the basis of Table 4.7., i.e. that part of the *o*-endings can be explained by the undeniable phonological merger of the outcomes of *-um* and *-us*, although the case contrast between the nominative and the accusative was still alive to some extent. As said, some of the *o*-endings may have been intended to be nominatives and some others accusatives. With *-o* it is usually impossible to tell.

What is crucial here is that names are always personal and intimate entities that are tightly related to the 'face' and identity of their carriers. It is socially unacceptable to question other's face by saying or writing his name wrongly, i.e. contrary to how the carrier himself presents (pronounces) his name. This normal human sensitivity of the scribes may have led them to transcribe even the pronounced ending /o/ as *-o* even when they knew that the subject function required the nominative in *-us*. Moreover, we do not know whether there was idiolectal (or sociolectal?) variation in using personal names: perhaps some persons preferred to be called and/or spelled in a given way. This can

---

[240] Adams 2013, 212; Lazard 2007, 516; cf. Pei 1932, 214; Taylor 1924, 100.

never be proved, but lists of names, such as (24), appear to be interesting environments for hypothesising the possible motivations behind the case assignment of personal names.[241]

(24) MED 202 (AD 785) *ubi nobiscum aderant sacerdotes et lociservatores seu haremannos*, *id est Iacobus diac(onus)*, *Rachiprandus presbitero*, *Rachifrid p*r(*es*)b(*iter*) *et Gausprandus l(oci)s(ervator)*, *Cunimundu*, *Bellerifunsu*, *Teuprand*, *Ghispert*, *Domnuccio*, *Filuartus*, *Vadilo*, *Lopo*, *Teoduald*

"where with us were present the priests and royal governors and *haremanni*, i.e. Iacobus, the deacon; Rachiprando, the priest; [--] Lopo, Teoduald"

The scribe of (24) is likely to have observed some principle of choosing the case forms of the participants of the list. It is, however, impossible to say if that principle reflects the real use. Certain persons whose name is of Germanic origin in (24), may have used the uninflected form (*Rachifrid*, *Teuprand*, *Ghispert*, *Teoduald*) while others may have used the inflected Latin form of the name (appearing as *Rachiprandus*, *Gausprandus*, *Cunimundu*, etc.). Nothing would have impeded the scribes from Latinising even *Rachifrid*, *Teuprand*, and others if they had wished. The fact that the names with Latin case ending, such as *Iacobus*, *Cunimundu*, and *Domnuccio*, have different endings may be a proof of the scribe's careful deliberation on how to spell the different names. On the other hand, it can be equally well a sign of carelessness: the scribe may not have been sure about how to spell the names and, therefore, chose a transcription that was phonetic or a sort of compromise between phonetic and etymologising spelling.

Table 4.8. provides a more detailed view on the distribution of the 2nd declension singular subject endings in LLCT. It is immediately noticed that personal name subjects do not favour *um*-endings but contain, instead, quite a few *u*-endings and many *o*-endings. Of course, *-um* is rather infrequent in each animacy class. However, the absence of *-um* in personal names may tell something about the scribes' attitudes towards personal names. The ending *-um* was probably known to be *the* accusative ending and, perhaps more importantly, a neuter ending,[242] both of which fit badly the personal name category that represents agentive, human actors. Writing *-um* in a personal name has probably looked awkward. If so, it can be seen as further proof that the highest transitivity domains were still considered proper to the nominative. At first sight, it seems paradoxical that writing *-o* or *-u* in

---

[241] On interpreting nominative/accusative lists, see Adams 2013, 226–232.
[242] In some Late Latin texts, *-um* is used hypercorrectly in almost any syntactic function with thematic neuters (Korkiakangas 2010, 139; Löfstedt 1961, 226–231).

127

a personal name subject was not obviously as awkward. According to my interpretation, the scribes may have chosen the *o*-ending, first, because they took pains to quote a person's name scrupulously and, second, because they may have thought that *-o* represented both the nominative and the accusative case. In any case, *-um* was a theoretical, learned relic that was also considered such.

Table 4.8. Endings of the 2nd declension singular subjects in LLCT.

Animacy	Ending	N	%
Inanimate	*-um*	3	c.15%
	-o	6	c.30%
	-u	2	c.10%
	-us	10	c.50%
	∑	21	
Animate	*-um*	5	7 %
	-o	14	19 %
	-us	45	63 %
	-r	8	11 %
	∑	72	
Personal	*-o*	255	39 %
	-u	36	5 %
	-us	364	56 %
	∑	655	

To sum up, (personal) names seem to have special properties that may contribute to the fact that the 2nd declension personal names so often occur with the ending *-o*. There appears to be an extra-linguistic or, better, sociolinguistic factor that works against semantic alignment: if semantically-based alignment were the only factor at play, 2nd declension personal names should display more *us*-endings than other animacy classes. Yet, because of its 'nameness' and the social factors involved therein, the form in *-o*, which was certainly common in the spoken language (whether it was functionally nominative or accusative), emerges more perceptibly in personal names than elsewhere.

The described hypothesis does not, of course, apply to the 3rd declension imparisyllabic personal names that were discussed in section 4.2.2. (Table 4.5.). In 3rd declension imparisyllabic personal names, no phonological merger takes place between the nominative and accusative forms. Assuming that the nominative/accusative contrast still existed, the possible personal spelling preferences are likely not to have crystallised as the only form either the accusative form, say, *Tasone* or the nominative *Taso*, because both continued to be functionally relevant. In the 2nd declension, this kind of crystallisation was more likely, as the function of *-o* had become ambiguous.

128

Next, I shall examine briefly how 3rd declension singular parisyllabic subjects behave. The 3rd declension singular parisyllabic nouns seem to follow the same pattern as the 2nd declension singular nouns in Table 4.7. The nominative and accusative forms of the 3rd declension parisyllabic nouns, which are, however, only 68, cannot be reliably kept apart from each other because of morphophonological developments that may have levelled the case endings *-is/es/em* at least in some contexts. This is because, contrary to the imparisyllabic nouns, the parisyllabic nouns do not have different roots in the nominative and the accusative (e.g. *fin-is* vs. *fin-em*).

Table 4.9. shows that the accusative percentage of the 3rd declension parisyllabic personal names (c.40%) is of the same magnitude as the 2nd declension personal name accusative percentage (44.4%) in Table 4.7. Similarly, the accusative percentage of the inanimate subjects is higher than that of the personal names with both the 3rd declension parisyllabic subjects (c.70%) and the 2nd declension subjects (c.50%). There is only one animate 3rd declension parisyllabic subject, so the animate accusative percentage cannot be compared to that of the 2nd declension subjects (26%). The 3rd declension personal names categorised as accusatives include, for example, *Leonaci*, *Iohanne*, and *Suave*.

Table 4.9. The 3rd declension singular parisyllabic subjects of LLCT.

	Accusative		Nominative		N
Inanimate	13	c.70%	5	c.30%	18
Animate	0	-	1	-	1
Personal	20	c.40%	29	c.60%	49
Σ	33	49%	35	51%	68
Chi-square	$\chi^2 = 6.01$, $df = 2$, $p = 0.028$ (Fisher's exact)[243]				

The percentages of Table 4.9. show that even the 3rd declension parisyllabic personal names are in the accusative more often than would be expected. As often stated, the theory of semantically-based alignment expects the personal name subjects to be the last category that was penetrated by the extended accusative. Based on what has been said in this chapter, I suggest that the relatively high accusative percentage of the personal name subjects (c.40%) results from the same influence of 'nameness' that was proposed for the 2nd declension personal names. The general high level of the accusative percentages (c.70% and c.40%) is probably due to the phonologically-motivated merger

---

[243] Fisher's exact test is a statistical significance test for contingency tables just like the chi-square test. Fisher's exact test can be used for samples that are not large enough for the chi-square test because it calculates the *p*-value exactly (see Agresti 2007, 45–46).

of the 3rd declension parisyllabic nominative and accusative endings, as it was in the 2nd declension data. In sum, what has been shown for the 2nd declension *o/u/um*-ending with personal names seems to apply, *mutatis mutandis*, also to the 3rd declension *e/em/i*-ending with personal names.

This chapter has revealed that it is problematic to assume that all the 2nd declension *o/u/um*-endings or all the 3rd declension *e/em/i*-endings would be either purely accusative-based or an inseparable outcome of a nominative/accusative merger. Instead, a certain portion of these forms, regardless of their identical appearance, may have been intended to be either nominative or accusative, not a form that represented their merger. This is likely to explain the unexpectedly high accusative percentages of the 2nd declension singular and 3rd declension singular parisyllabic subjects of personal names.

4.3. Case study: transitivity degree of the construction types in LLCT

The previous sections showed that the animacy variable is connected to the subject case more clearly than the construction type variable. It is, however, worth finding out whether the somewhat lower performance of the construction type variable is due to some skewing factors involved in its operationalisation. It was stated in section 4.1.2. that the technical fourfold construction type categorisation that is followed in this study and in certain other studies is a mere tool without a coherent theoretical basis: the contrast between A and S$_A$/S$_O$ subjects is motivated by syntactic transitivity, the contrast between S$_A$ and S$_O$ on exclusively semantic grounds, and the contrast between A/S$_A$/S$_O$ and S$_O$ passive subjects again on syntactic grounds. The question arises whether a more satisfying starting point can be found for the categorisation of the construction type variable. The concept of transitivity scale seems to offer one solution.

In this chapter, I shall conduct a case study in which the transitivity of a representative sample of LLCT clauses is decomposed into ten transitivity components according to the model proposed by Hopper and Thompson (1980). It is to be noticed that, owing to the multifactorial nature of transitivity, the model is also multifactorial and, thus, based on empirical observations and their subjective interpretation. Therefore, the model can be considered a more granular version of the fourfold construction type classification. Because of its ten semantic and syntactic components, the model is, however, supposed to yield transitivity degree values that can be proportioned with each other in a more meaningful way than the construction types.

Another motivation of this case study is that the research that has been made lately in the field of genre and text type studies has shown that different genres behave in a radically different manner as far as transitivity is concerned.[244] The present transitivity component analysis will provide the first results concerning the overall transitivity degree of Latin legal genre. As a byproduct, certain problems that arose in the numerical analysis of section 4.3., such as the almost equal share of the accusative percentage of the A and S_A subjects, can be explained. Although the question concerning the transitivity degree of A and S_A subjects goes well beyond the horizon of a study of charter Latin, it is necessary to treat it here, especially because LLCT, thus, provides its contribution to this cross-linguistically important discussion.

4.3.1. Hopper and Thompson's transitivity scale

Transitivity is the effectiveness with which an action takes place. Hopper and Thompson (1980) divide this effectiveness into ten components, some of which are related mainly to the nominal participants, others to the verb.[245] In this study, I view transitivity essentially as a technical continuum on which all the construction types and, thus, all the subject/verb combinations of LLCT can be located: syntactically intransitive constructions are low in transitivity, whereas syntactically transitive, i.e. two-participant, constructions are prototypically high in transitivity. (It is, however, not so clear what is the relative transitivity order of the S_O and S_A constructions.)

My intention is to find out whether Hopper and Thompson's transitivity component model predicts subject case selection in LLCT more accurately than animacy and construction type variables that have been discussed in the preceding chapters. Transitivity, understood as multifactorial according to Hopper and Thompson, measures essentially the same phenomena as the animacy degree and construction type. What is new is that the ten-component analysis easily yields comparable transitivity degree values that are likely to represent rather comprehensively the semantic parameters that affect the relation between the subject and the verb. Further, Hopper and Thompson's model includes many parameters that are thought to be constituents of what was called 'control' in section 3.1. Therefore, Hopper and Thompson's model can be viewed as a tool for measuring not just transitivity, but also the control of subject over verbal event.

---

[244] Thompson & Hopper 2001; Vázquez & García-Miguel 2009.
[245] Hopper & Thompson 1980, 251–252; Kittilä 2002, 27.

Table 4.10. presents Hopper and Thompson's (1980) transitivity scale. The left column gives the parameter name while the two other columns assign to it two categorical levels: one that implies high transitivity and another that implies low transitivity. A (agent) and O (object) of Table 4.10. refer to the two participants in a two-participant clause regardless of their syntactic functions.

Table 4.10. Hopper & Thompson's (1980) transitivity scale.[246]

Parameter	High transitivity	Low transitivity
Number of participants	2 or more participants (A and O)	1 participant
Kinesis	action	non-action
Aspect	telic	atelic
Punctuality	punctual	non-punctual
Volitionality	volitional	non-volitional
Affirmation	affirmative	negative
Mode	realis	irrealis
Agency	A high in potency	A low in potency
Affectedness of O	O totally affected	O not affected
Individuation of O	O highly individuated	O non-individuated

None of these factors is claimed to be decisive, but they are together responsible for the degree of transitivity according to which clauses can be ranked. Even though Hopper and Thompson's model may not be the best suited one for the current study, it is applied here because it is accessible and well known. Consequently, it enhances the possibility of comparing the results of this study with other studies, as far as this is ever possible. Moreover, the semantic emphasis of Hopper and Thompson's model feels sensible, as the current study discusses first and foremost the role of semantics in case marking.

I underline that Hopper and Thompson's scale is an abstract framework whose application to a real linguistic study is problematic for reasons that are both practical and theoretical. It is necessary to explain some central features of Hopper and Thompson's model in order to understand how the LLCT data are dealt with in what follows. Some of the transitivity component parameters are explicated below:[247]

- Kinesis is about lexical aspect or aktionsart (Vendler 1957): the distinction is made between state and the three other categories (action, achievement, accomplishment).
- Aspect denotes grammatical aspect. Here it is thought to present an action or event as being complete (telic) or incomplete (atelic) in some sense.

---

[246] Hopper & Thompson 1980, 252–253.
[247] The definitions come from Hopper & Thompson 1980, 252–253 unless otherwise indicated.

- Punctuality expresses momentary action that has no duration, e.g. 'to kick' vs. 'to carry'.
- Volitionality indicates whether the verbal event taking place is intended or instigated by the subject or not.
- Affirmation: certain transitivity features of affirmative and negative clauses are encoded differently in certain languages.[248]
- Mode: the realis mode presents a verbal event as 'real', and corresponds to indicative present, imperfect, or pluperfect in declarative sentence (not in interrogative clause or hypothetical subordinate clause).[249]
- Agency is an extremely complicated concept that involves the planned involvement of the subject in the verbal activity.[250]
- Affectedness of O describes to what degree an action is transferred to a patient.[251]
- Individuation of O refers both to the distinctness of the O from the A and to its distinctness from its own background.[252]

Hopper and Thompson proposed their model to be a tool for comparative typological study. They did not intend it to be a technical schema that would be calibrated on the basis of the minimum and maximum scores of transitivity points calculated for each component parameter. In other words, the components are not necessarily all equally significant, but depend on the language, structure and the purpose of each individual study. Non-hierarchisation is also the major weakness of the scale because it is impossible to tell which of the listed components are crucial in a certain, or in any, case. There are evident differences in the typological significance of the parameters as well. Hopper and Thompson have also been criticised for their selection of parameters some of which tend to co-occur and, thus, can be seen as a single parameter.[253]

These defects are also verified in the current study where the transitivity scale is used for quantitative analysis contrary to the original intention of Hopper and Thompson. The incommensurateness of the transitivity components turns out to be especially problematic: for example, the affirmation or mode are hardly as important in Latin as the other components. The central problem is that all the components have to be treated as binary categories in the analysis although, in reality, certain components (number of participants, affirmation, mode) are binary while others are scalar. Also, the overlapping of certain categories (especially agency) with others makes the analysis challenging. At

---

[248] Hopper & Thompson 1980, 276–277.
[249] Hopper & Thompson 1980, 277.
[250] The subject is not, however, necessarily animate or human (Hopper & Thompson 1980, 286; Ahearn 2010).
[251] Lehmann 1991, 217–221.
[252] The (complete) individuatedness of O is very challenging to define. Indeed, Hopper and Thompson consider it a scalar phenomenon (Hopper & Thompson 1980, 253). In my analysis, animacy and referentiality are the principal criteria. For a detailed analysis of individuation, see Timberlake 1975, 124–134; Timberlake 1977, 162.
[253] Kittilä 2002, 27.

this point, it is again necessary to underline the subjectivity that is always involved in semantic analysis. In some cases, my solutions concerning the transitivity gradients inevitably deviate from the common practice (if there is any), but at least I have been consistent in them.[254]

The difficulty of defining the semantic properties of certain subject/verb combinations of LLCT with accuracy is further increased by the fact that the semantic value of several expressions remains unclear in charter Latin. This is due to contamination of formulae, misinterpretations by the scribes, and, in general, the inaccessibility of certain expressions in the absence of native-speaker language knowledge. Good examples include the oscillating argument structures of verbs (*de*)*precare* 'to pray', (25) and (26), and *molestare* 'to disturb' (27).

(25) CDL 114 (AD 754) *ut sacerdotes qui in ipsas eclesias ordinati fuerent pro facinoribus meis Dominum deprecare debeant*

"that the priests who will be ordained in these churches must <u>pray</u> to the Lord for my sins"

(26) CDL 261 (AD 772) *ut sacerdos qui iuidem fuere mihi pro salute anime messarum precibus a Domino deprecare deueas* [= *debeat*]

"that the priest who will be ordained there must <u>pray</u> the requiem mass for the salvation of my soul from the Lord [?]"

(27) CDL 35 (AD 724) *et numquam nos uel posterus noster te de hanc dicto loco molestari praesumat*

"nor should we or our heirs ever dare to <u>expel</u> you out of this mentioned place"

It is not clear which construction is intended in (26). The prepositional phrase *a Domino* may be a dative periphrasis.[255] Case (27) may originate from a contamination of constructions involving *molestiam inferre* 'to bring disturbance' and *expellere* 'to expel', which both occur in charter Latin. Euphemistic expressions abound as well: both *migrare* 'to migrate' (28) and *recedere* 'to recede' are used to denote dying. There is no other way than to do the transitivity analysis on the basis of the real meaning of these figurative/metaphorical phrases, i.e. to parallel the verbs with the verb 'to die'.

---

[254] As mentioned above, Lakoff (1977, 244) also presents a transitivity scale. Lakoff describes a prototypical transitive clause (of English) that must fulfil fourteen transitivity criteria. The transitivity degree of any clause can be approximated according to them, at least in theory. Some of these criteria are again redundant and could be subsumed under some other criteria (Kittilä 2002, 28; Cennamo 2001c, 52; Comrie 1989, 61–62).

[255] It can hardly be interpreted as a prepositional object. See Ledgeway 2012, 350–351.

Cf., however, *mors occupauerit* and *mortis* [= *mors* or *mortem*] *obserueat* that are transitive but still denote dying (29). They cannot be interpreted as intransitive, of course.

(28) CDL 24 (AD 720) *si ipse abbas custus de hac luce <u>migraueret</u>*

"if the abbot in charge <u>leaves</u> this world"

(29) CDL 171 (AD 763) *et si me mors <u>occupauerit</u>*

"and if death <u>conquers</u> me"

4.3.2. Transitivity component analysis of sample clauses

I proceed now to the actual analysis of the LLCT case study. I have defined the transitivity degree of 471 sample clauses that are classified as evenly as possible into 36 number/case/construction type/animacy combinations, so that it is possible to test the transitivity degree on the number, case, construction type, and animacy class of the subject.[256] This classification can be seen in Appendix 4.3. that presents the whole sample data. The transitivity degree is assessed using Hopper and Thompson's transitivity scale by providing each of its components with a binary value (0 or 1).

In practice, this means that I have selected a random sample of 25 clauses for each number/case/construction type/animacy combination. The clauses were sampled systematically at regular intervals through the lists of sentences with the sentences ordered alphabetically by the predicate verb lemma. The selection interval was determined so that the sample size did not exceed 25. In case there were fewer than 25 clauses in a given number/case/construction type/animacy combination, I included all the available occurrences. Of the transitive sentences with inanimate singular subjects, for example, I have sampled all the 17 clauses with accusative-form subjects and 25 of the 49 clauses with nominative-form subjects.

When the predicate verbs of this sample material are compared to lists that are comprised of all the non-passive basic query subset clauses, it is seen that the sample corpus contains all the categories characteristic of the entire LLCT with roughly similar proportions. As a consequence, I consider the sample to be representative enough of the verb variation attested in the whole population of LLCT.

---

[256] The mentioned classification produces 2 * 2 * 3 * 3 = 36 analysis units: two numbers (singular and plural), two cases (nominative and accusative), three construction types (A, S_A, and S_O subjects), and three animacy classes (inanimate, animate, and personal name). The S_O subjects of passive constructions are not dealt with here, as it is not clear how Hopper and Thompson's transitivity component analysis should be applied to the passive.

In total, I have analysed 308 clauses of the 1,242 singular-subject non-passive clauses and 163 clauses of the 216 plural-subject non-passive clauses. All these clauses with their respective analyses can be viewed in Appendix 4.3.

Next, I shall formulate the hypotheses H_1^1 to H_1^3 that will be tested in this chapter. The corresponding null hypotheses H_0^1 to H_0^3 are given after each hypothesis. The dependent and independent variables are listed in Table 4.11. with their categorical levels. As regards the hypotheses, 'transitivity of the construction type' refers to the relative order in which the A, S_A, and S_O subjects occupy their positions on the transitivity continuum. Respectively, 'transitivity degree' refers to the value obtained on the basis of Hopper and Thompson's transitivity scale. The statistical analysis is again performed with contingency tables and decision trees.

H_1^1: The less transitive the construction type of the subject is, the lower the transitivity degree is expected to be.

H_0^1: The construction type has no significant effect on the transitivity degree.

H_1^2: The lower the animacy class of the subject is, the lower the transitivity degree is expected to be.

H_0^2: The animacy class has no significant effect on the transitivity degree.

H_1^3: The lower the transitivity degree of the verbal event is, the more often the subject is expected to occur in the accusative.

H_0^3: The transitivity degree has no significant effect on the case form of the subject.

Table 4.11. Dependent and independent variables with their categorical levels (section 5.3.2.).

Dependent variables		Independent variables	
Variable name	Categorical levels	Variable name	Categorical levels
transitivity degree	0, 1,…, 9[257]	construction type	A/S_A/S_O subject
		animacy class	personal, animate, inanimate
case form of subject	nominative or accusative	transitivity degree	0, 1,…, 9

---

[257] The transitivity degree is here interpreted as a categorical variable. See the discussion at the beginning of chapter 5.

Table 4.12. is an example of transitivity component analysis applied to sentence (30) that is a simple transitive clause (plus antecedent) with a rather prototypical transitive verb and an individuated, referential (effected) object. All the other transitivity components receive value 1, except for punctuality and affectedness of object. This is because the process of concluding a sales contract has a certain duration (i.e. the event is not punctual) and the object is not completely affected. The relative pronoun *quas* 'that', which refers to its antecedent *uinditione* 'sales contract', is an effected object and, hence, cannot be genuinely affected. Moreover, *uinditione* (or *quas*) is not a concrete object that is really made, but the making has to be understood figuratively.[258] Consequently, the clause under examination receives a transitivity value of 8 in 10 (i.e. 80%).

(30) CDL 253 (AD 771) *ipsa uinditione quas eorum* [= *eis*] *Brittulo fecit*
"the sales contract that Brittulo made to them"

Table 4.12. Example of transitivity component analysis of clause (30).

Subject	Verb	Object	2 participants	Action	Telicity	Punctuality	Volitionality	Affirmation	Realis mode	Agency	Affected O	Individuat. O	Σ
Brittulo	*fecit*	*quas*	1	1	1	0	1	1	1	1	0	1	8 (80%)

After evaluating the transitivity degree of each sample clause according to the ten transitivity parameters, I aggregate these values and obtain, thus, the transitivity degree values for each construction type. They can be viewed in Table 4.13. The numbers of the table are mean transitivity percentage values that are summed up from individual clauses for each transitivity parameter and for each construction type.

Table 4.13. Transitivity component values of each construction type.

Construction type	2 participants	Action	Telicity	Punctuality	Volitionality	Affirmation	Realis mode	Agency	Affected O	Individuat. O	Σ mean	N
A	79%	65%	57%	2.7%	64%	90%	46%	54%	7.0%	44%	**51%**	187
S$_A$	0%	100%	75%	0%	100%	88%	44%	78%	0%	0%	**49%**	77
S$_O$	0%	0%	5.3%	3.4%	1.0%	97%	41%	0%	0%	0%	**15%**	207

---

[258] For abstract noun + *facere* constructions, see section 4.4.1.

137

It is immediately noticed that there is a considerable difference between the transitivity degrees of different construction types. The gap is strikingly wide between the A/S_A and S_O subjects. Actually, the A and S_A subjects of LLCT can be said to align with each other against the S_O subjects: the distance is about 35 percentage points. In general, the difference in size of the mean transitivity percentages of the construction types accords roughly with the accusative percentages obtained in Table 4.6. of section 4.2.2.: the higher the transitivity degree, the lower the accusative percentage (A 15%, S_A c.10%, S_O 31%).[259]

In Table 4.13., the transitivity value decreases on the continuum $A - S_A - S_O$ (51% – 49% – 15%). This is a positive piece of evidence of the usability of Hopper and Thompson's transitivity scale as the basis of corpus analysis. Of course, some circular reasoning may be involved here, as the scale includes factors that relate to the inherent properties of the subject, such as agency. Therefore, it would be a desideratum to have a tailored version of the scale where the overlapping factors were removed as carefully as possible.

Differences are also found within individual transitivity component percentages. It is obvious that the parameters that are meant to distinguish between (syntactically) transitive and intransitive clauses (2 participants, affected O, individuated O) show 0% for the intransitive cases. What is, however, not obvious is that even the percentages of the A subjects are surprisingly low in categories 'affected O' and 'individuated O', just as they are with most of the parameters of Hopper and Thompson's scale. The mean transitivity percentage of A is, indeed, only 51% although a prototypical transitive clause could be expected to figure with 100% of transitivity. This will be discussed later in section 4.3.3. However, syntactically realised objects are not present in every (semantically) transitive clause of LLCT, as was stated in section 4.1.3. and section 4.1.4. Indeed, only 79% of the A constructions seem to have two overt participants, i.e. a direct object and a lexical subject, which is the default in the basic query subset that excludes pronominal and zero subjects.

I find particularly important the observation that the transitivity degree of the A and S_A subjects is almost the same. This is even more noteworthy, as three out of the ten transitivity parameters of Hopper and Thompson's are object-related and, thus, do not apply to the S_A subjects of unergative clauses at all while being operative in most of the transitive clauses. This relates undoubtedly to my

---

[259] Note that Table 4.6. is about the 3rd declension *imparisyllabic* subcorpus only.

decision to define transitivity chiefly on semantic criteria, not on the number of participants.[260] This (corpus-specific?) feature is certainly related to the almost equal accusative percentages of the A and S_A subjects that were observed in section 4.2.1. In LLCT, the S_A subjects seem to be highly agentive (78%) while the A subjects are rather low in agentivity (54%) and transitivity in general. Once the A and S_A subjects seem to align semantically with each other, they are also likely to behave in a similar manner when case marking is concerned. Indeed, Table 4.6. of section 4.2.2. shows that the accusative percentage of A is 15% and that of S_A is c.10% even though in Table 4.6. the question is about the 3rd declension subcorpus. On the level of the whole LLCT, the similarity of encoding is likely to be blurred because of the 2nd declension personal names.

Figure 4.7. Decision tree for transitivity degree (dependent), construction type and animacy (independent).

Transitivity_degree

Node 0

Category	%	n
0-1	26,5	125
2	23,1	109
3	7,6	36
4	7,9	37
5	9,3	44
6	13,2	62
7	5,7	27
8	5,3	25
9	1,3	6
Total	100,0	471

Construction_type
Adj. P-value=0,000, Chi-square=410,715, df=16

SO → **Node 1**

Category	%	n
0-1	55,1	114
2	40,1	83
3	4,8	10
4	0,0	0
5	0,0	0
6	0,0	0
7	0,0	0
8	0,0	0
9	0,0	0
Total	43,9	207

A → **Node 2**

Category	%	n
0-1	5,9	11
2	12,3	23
3	10,7	20
4	10,7	20
5	10,7	20
6	18,7	35
7	14,4	27
8	13,4	25
9	3,2	6
Total	39,7	187

SA → **Node 3**

Category	%	n
0-1	0,0	0
2	3,9	3
3	7,8	6
4	22,1	17
5	31,2	24
6	35,1	27
7	0,0	0
8	0,0	0
9	0,0	0
Total	16,3	77

Animacy (Node 1) Adj. P-value=0,000, Chi-square=51,397, df=4
Animacy (Node 2) Adj. P-value=0,028, Chi-square=20,233, df=8
Animacy (Node 3) Adj. P-value=0,000, Chi-square=45,859, df=4

animate → **Node 4**

Category	%	n
0-1	54,8	34
2	35,5	22
3	9,7	6
4	0,0	0
5	0,0	0
6	0,0	0
7	0,0	0
8	0,0	0
9	0,0	0
Total	13,2	62

inanimate → **Node 5**

Category	%	n
0-1	74,7	71
2	22,1	21
3	3,2	3
4	0,0	0
5	0,0	0
6	0,0	0
7	0,0	0
8	0,0	0
9	0,0	0
Total	20,2	95

personal → **Node 6**

Category	%	n
0-1	18,0	9
2	80,0	40
3	2,0	1
4	0,0	0
5	0,0	0
6	0,0	0
7	0,0	0
8	0,0	0
9	0,0	0
Total	10,6	50

animate; inanimate → **Node 7**

Category	%	n
0-1	8,0	11
2	16,1	22
3	10,9	15
4	10,2	14
5	13,1	18
6	16,1	22
7	12,4	17
8	10,2	14
9	2,9	4
Total	29,1	137

personal → **Node 8**

Category	%	n
0-1	0,0	0
2	2,0	1
3	10,0	5
4	12,0	6
5	4,0	2
6	26,0	13
7	20,0	10
8	22,0	11
9	4,0	2
Total	10,6	50

animate → **Node 9**

Category	%	n
0-1	0,0	0
2	6,8	3
3	6,8	3
4	36,4	16
5	45,5	20
6	4,5	2
7	0,0	0
8	0,0	0
9	0,0	0
Total	9,3	44

inanimate; personal → **Node 10**

Category	%	n
0-1	0,0	0
2	0,0	0
3	9,1	3
4	3,0	1
5	12,1	4
6	75,8	25
7	0,0	0
8	0,0	0
9	0,0	0
Total	7,0	33

---

[260] As was seen in section 4.1.4., I have also interpreted constructions with complement clauses as transitive constructions.

A tight statistical dependence between transitivity degree and construction type is also seen in the decision tree in Figure 4.7., where the construction type creates the first branch (nodes 1–3) of the sample with a high chi-square value ($\chi^2 = 410.72$, $df = 16$, $p < 0.001$). The decision tree is challenging to read because the categorical levels of the dependent variable are many. The same distribution that can be observed in the histograms of the construction type nodes 1–3 in the decision tree is perceived more easily in Figure 4.6. There, the construction types are located characteristically across the transitivity continuum: the S_O subjects have their locus in the low-transitivity domains, whereas the S_A subjects peak on the middle values. The A subjects are more evenly distributed, which is a very interesting observation indeed. This observation will be further analysed in section 4.3.3. As said, the transitivity degree is here interpreted as a categorical variable.

It is to be noticed that the *x*-axis in Figure 4.6. does not cover the whole ten-degree scale of the transitivity model. Transitivity value 10 is missing, as no clause receives this high score, and values 0 (5 occurrences) and 1 (120 occurrences) have been merged. The merger of 0 and 1 was imperative since otherwise the chi-square could not have been calculated because of too many small expected count cells. The cross-tabulation for transitivity degree and construction type with adjusted standardised residuals can be found in Appendix 4.4.

Figure 4.6. Construction type as a function of transitivity degree ($\chi^2 = 410.72$, $df = 16$, $p < 0.001$).
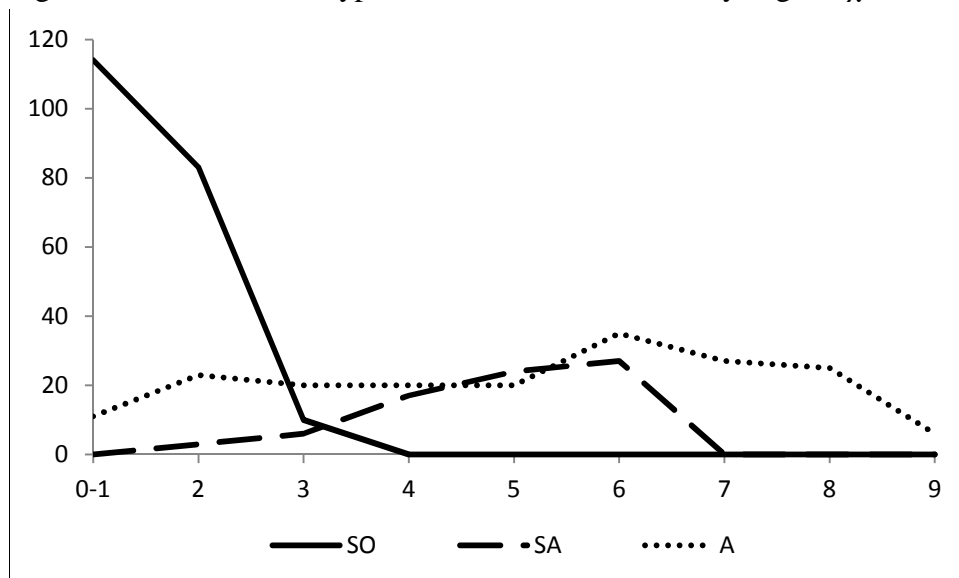
The decision tree in Figure 4.7. also reveals that a second split (nodes 4–6, 7–8, and 9–10) in the sample data is caused by animacy and it materialises under each construction type node. The transi-

tivity degree distributions are again interesting: they will be examined below with the help of Figure 4.8. As for the decision tree structure, the pattern is not completely as one might expect on general grounds: the inanimate and personal subjects align unexpectedly with each other in the S_A branch, although Figure 4.8. shows that, in general, inanimate and personal name subjects are not distributed in a similar way. It has to be remembered that the lower splits of decision trees do not represent the whole data but only the occurrences that are found within each first-split branch. Therefore, the alignment of the inanimate and personal subjects in the S_A branch reflects only the peculiar animacy distribution of the 77 S_A subjects. In the S_O branch, each animacy class creates a statistically significant node, which is also noteworthy, especially because the S_O personal name subjects appear to be very low in animacy.

Figure 4.8. Animacy as a function of transitivity degree ($\chi^2 = 145.50$, $df = 18$, $p < 0.001$).
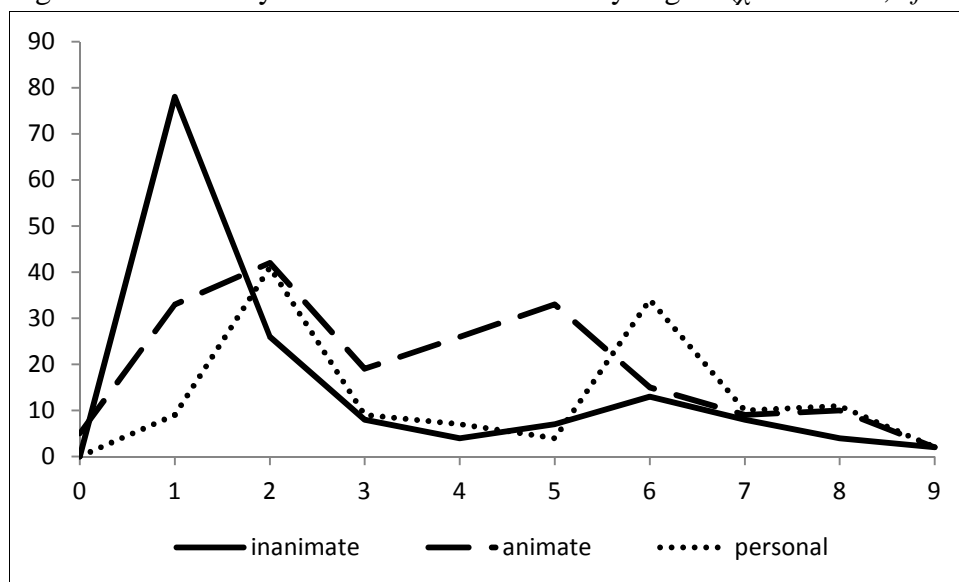
It has already been shown in section 4.2. that animacy and construction type are tightly interrelated. The frequency distribution of the occurrences of animacy classes across the transitivity continuum can be seen in Figure 4.8. This time, the x-axis covers the whole ten-degree scale of the transitivity model with the exception of value 10, which is lacking in the sample. The curves in Figure 4.8. show that inanimate subjects are predictably concentrated at the low-transitivity end, which holds good less strikingly also for animate subjects although they occur within a wider range of transitivity. This said, all the curves seem to display somewhat of a drop between values 2 and 6. It may be that the resolution power of Hopper and Thompson's model is not at its best round the middle of the scale. This may be partly due to certain causally connected parameters, such as action, volitionality, and agency, which usually all assume either 0 or 1 (see Table 4.13.), a phenomenon that causes the

transitivity value to jump by three units. This is possibly visible even in Figure 4.6. as a slight interim drop of the A subject graph and the sudden disappearance of the S_O subjects.

Back to Figure 4.8., the personal name curve seems to have two separate maximums at 2 and 6 because of the intermediate drop (see also nodes 10 and 6 in Figure 4.7.). The personal name subjects of LLCT are, indeed, surprisingly low in transitivity. This issue will be revised in section 4.3.3. Nevertheless, Figure 4.7. and Figure 4.8. suggest the following general trend: the higher the animacy class of the subject, the higher the transitivity degree of the clause. The cross-tabulation for transitivity degree and animacy with adjusted standardised residuals can be found in Appendix 4.5. Based on the chi-squares in Figure 4.6. ($\chi^2 = 410.72$, $df = 16$, $p < 0.001$) and Figure 4.8. ($\chi^2 = 145.50$, $df = 18$, $p < 0.001$), the dependence between construction type and transitivity turns out to be stronger than the dependence between animacy and transitivity (note, however, the different degrees of freedom). This is obviously related to the internal properties of Hopper and Thompson's model and does not undo the observations made in section 4.2. according to which animacy is likely to affect the subject case selection more significantly than the construction type.

To conclude, I shall have a look at the dependence of the subject case and the transitivity degree in order to see whether the ten-grade transitivity model provides a more potent tool for detecting the motivation of case marking than the fourfold construction type classification or the animacy classes. A decision tree for the case and the transitivity degree is found in Figure 4.9. The levels are again given as categories 0–1, …, 9, as with Figure 4.6. and Figure 4.7.

Figure 4.9. Decision tree for case (dependent) and transitivity degree (independent).
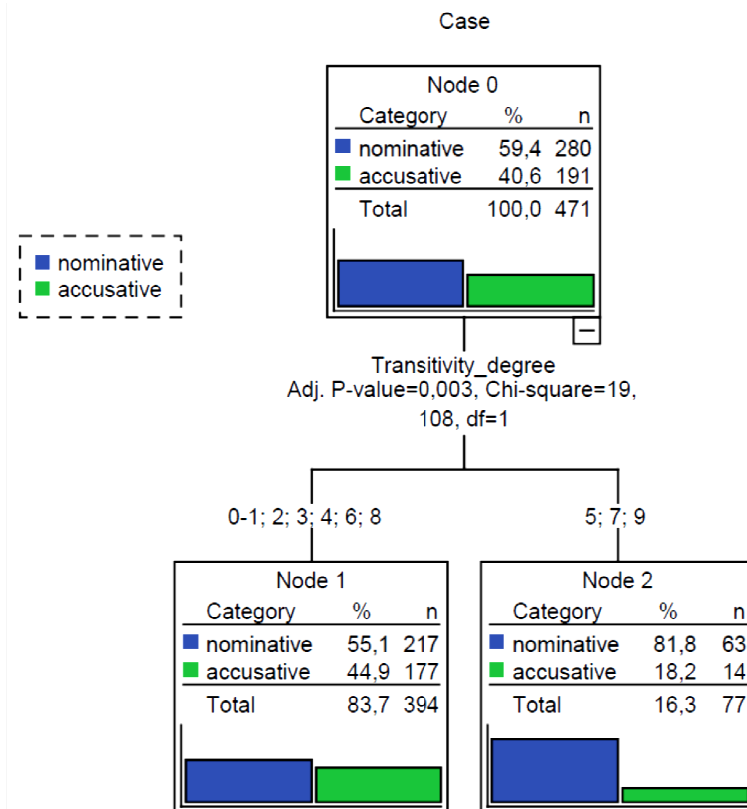
The decision tree in Figure 4.9. shows that the transitivity degree causes a statistically significant split, with a considerable amount of noise, however. Although the chi-square value is high ($\chi^2 =$ 19.11, $df = 1$, $p = 0.003$), node 2 with the lower accusative percentage (18%) displays discontinuous transitivity levels: values 5, 7, and 9 do not form a coherent unity because values 6 and 8 are aligned with the values of node 1. The transitivity degree, as defined by Hopper and Thompson, seems to be a rough indicator of subject case in a similar manner to the construction type or animacy degree. It cannot, however, be considered as having any particular advantage in comparison to those other variables.

Table 4.13. presents the dependence between the case of the subject and the transitivity degree of the construction where the subject occurs. The table shows that the adjusted standardised residuals are significant (> 2.0) precisely at values 5 and 7, where the accusative percentages are at their lowest (together with value 9). This suggests that these two transitivity degrees are responsible for the statistical significance of the cross-tabulation of Table 4.13. Residuals -2.9 and -2.8 indicate that the categorical levels 5 and 7 have significantly fewer occurrences of the accusative than would be expected if the null hypothesis were true, hence the low accusative percentage 18.2% in node 2 of the decision tree. Note that in Table 4.13., where the categorical levels are examined as categories 0–1,

…, 9 (*df* = 8), the chi-square is not comparable to the chi-square in Figure 4.9. where the decision tree algorithm groups the levels in two nodes (*df* = 1).

Table 4.13. Dependence between case of subject and transitivity degree of construction.

Case		Transitivity degree									Total
		0-1	2	3	4	5	6	7	8	9	
nominative	N	68	61	18	20	35	39	23	11	5	280
	%	54%	56%	c.50%	c.55%	c.80%	65%	c.85%	c.45%	c.85%	59.4%
	residual	-1.3	-0.8	-1.2	-0.7	**2.9**	0.6	**2.8**	-1.6	1.2	
accusative	N	57	48	18	17	9	23	4	14	1	191
	%	46%	44%	c.50%	c.45%	c.20%	35%	c.15%	c.55%	c.15%	40.6%
	residual	1.3	0.8	1.2	0.7	**-2.9**	-0.6	**-2.8**	1.6	-1.2	
Total	N	125	109	36	37	44	62	27	25	6	471
Chi-square		χ^2 = 22.64, *df* = 8, *p* = 0.004									

Figure 4.10. visualises the accusative percentages of Table 4.13. across the transitivity continuum. A relatively clear trend is observed: the higher the transitivity degree is, the higher the conservation of the nominative and the lower the accusative percentage. Nevertheless, the high-transitivity end is problematic: value 8 presents a surprisingly high accusative share and the frequencies behind the low accusative percentage at value 9 are only 1 in 6. Note that the general accusative share of the sample is elevated (40.6%), undoubtedly because of the 2nd declension personal names. Moreover, value 9 is not the only place where the frequencies are low.

Figure 4.10. Accusative subject distribution across the transitivity continuum.
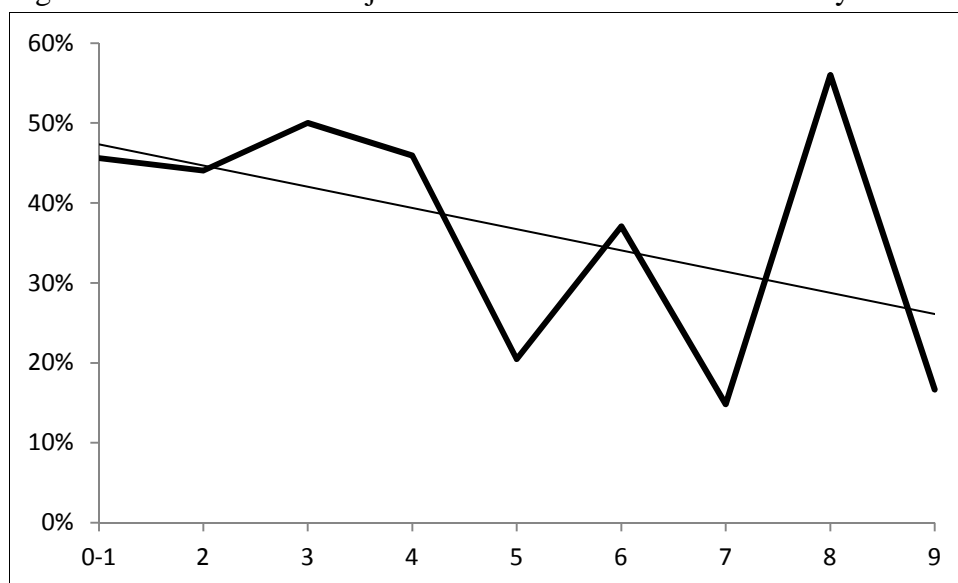
Table 4.14. offers a different viewpoint on what has been seen in the above tables and decision trees. Table 4.14. presents the distribution of the individual transitivity parameters classified by case and construction type (A, S_A, S_O). It can be seen that practically the only group whose transitivity degree really influences the subject case selection is construction type A, i.e. the subjects of verbs traditionally classified as transitive. In other words, the difference of 7 percentage points in favour of the nominative A subjects implies that those transitive clauses that select the accusative subject tend to be lower in transitivity (47%) than those that select the nominative subject (54%). At any rate, both the percentages are very low considering that they do belong to A subjects of transitive clauses. The S_A and S_O subjects, instead, have almost equal transitivity degrees, whether they are in the nominative or the accusative (S_A 48% vs. 49%, S_O 15% vs. 14%).

Table 4.14. Transitivity component values of each construction type for nominative and accusative subjects.

Case	2 participants	Action	Telicity	Punctuality	Volitionality	Affirmation	Realis mode	Agency	Affected O	Individuat. O	Σ mean	N
Nominative												
A	75%	70%	62%	3.8%	72%	90%	50%	64%	8.6%	43%	**54%**	105
S_A	0%	100%	74%	0%	100%	84%	43%	83%	0%	0%	**48%**	58
S_O	0%	0%	6.8%	5.1%	1.7%	95%	41%	0%	0%	0%	**15%**	117
Accusative												
A	84%	57%	51%	1.2%	52%	89%	41%	41%	4.9%	45%	**47%**	82
S_A	0%	100%	79%	0%	100%	100%	47%	63%	0%	0%	**49%**	19
S_O	0%	0%	3.3%	1.1%	0%	100%	40%	0%	0%	0%	**14%**	90

Table 4.14. reveals that the individual transitivity parameters behave differently and, thus, are of different resolution power. The case marking of the A subjects seems to be mainly related to parameters action, telicity, volitionality, and agency, which display the most difference between the nominative and accusative case. On the other hand, all the mentioned parameters show equally high or even higher percentages with the S_A subjects. In fact, only the number of participants and the closely related affectedness and individuation of the object distinguish S_A clearly from A in LLCT. As far as the purely semantic properties are concerned, the S_A subjects of LLCT are, thus, as high or higher in transitivity than the verbs classified as transitive. So, Hopper and Thompson's scale tells the difference between A and S_A only because of the (syntactic) valency parameter that is incorporated in the object-related parameters. In sum, the parameters action, telicity, volitionality, and agency are decisive only when distinguishing the S_O subjects from other construction types. This

may be symptomatic of the inherent problems of Hopper and Thompson's model or of the atypical behaviour of Latin (of LLCT). Further comparative research is needed to answer these questions.

The observations made about the transitivity percentages of each construction type in Table 4.14. seem to corroborate the observations made on the basis of the accusative subject distribution across the transitivity continuum (Table 4.13. and Figure 4.10.). Indeed, in LLCT, the extension of the accusative into subjects more or less follows the transitivity degree: the observations recall the accusative percentages between subject case and construction type in the 3rd declension subcorpus seen in Table 4.6. in section 4.2.2. The percentages of Table 4.6. are reprinted here on the right of the transitivity degree values of each construction type in Table 4.15.[261] It is noticed that the A and S_A subjects are aligned with each other together with low accusative percentages (15% and c.10%, respectively) while the S_O subjects are set aside with a notably higher accusative percentage (S_O 31%, passive S_O c.30%). A theory-compatible reverse pattern arises from the distribution of these subjects across the transitivity degree.

Table 4.15. Dependence between construction type, transitivity degree, and accusative percentage.

Construction type	Transitivity degree	Accusative subjects
A	51%	15%
S_A	49%	c.10%
S_O	15%	31%

What does, then, the difference of 7 percentage points between the nominative and accusative occurrences of the A subjects in Table 4.14. mean as for the morphological alignment? Why is there no similar difference between the nominative and accusative occurrences of the S_A subjects or the S_O subjects in Table 4.14.? One might assume that if the transitivity degree, as defined according to Hopper and Thompson's model, aligns the A and S_A subjects as opposed to the S_O subjects, the A and S_A subjects should also pattern similarly in regard to case marking. As is seen in Table 4.15., the 15% and the c.10% of the A and S_A subjects are not very far from each other, but why then the mentioned difference of 7 percentage points?

Actually, even the S_A and S_O subjects should be expected to display a difference in transitivity between the nominative and accusative subjects, but in practice the pattern remains invisible, as the

---

[261] Treating together numbers that come from the case study sample (which is based on the whole LLCT) and from the 3rd declension subcorpus is not problematic in Table 4.15. because here the joint transitivity degree value does not take a stand on the accusativity/nominativity of the 2nd declension personal names.

major part of these subjects is really scattered across a very narrow transitivity span, as was seen above in Figure 4.6. It is because of the highly skewed distribution of the S_A and S_O subjects that it is not possible to show a statistically significant dependence between the transitivity degree and the internal accusative-percentage distribution of constructions S_A and S_O. Instead, the accusative percentage of the A subjects across the transitivity continuum (values 1 to 9) can be presented as a graph in Figure 4.11.

Figure 4.11. Accusative percentage of the A subjects across the transitivity continuum ($\chi^2 = 23.83$, $df = 8$, $p = 0.002$).
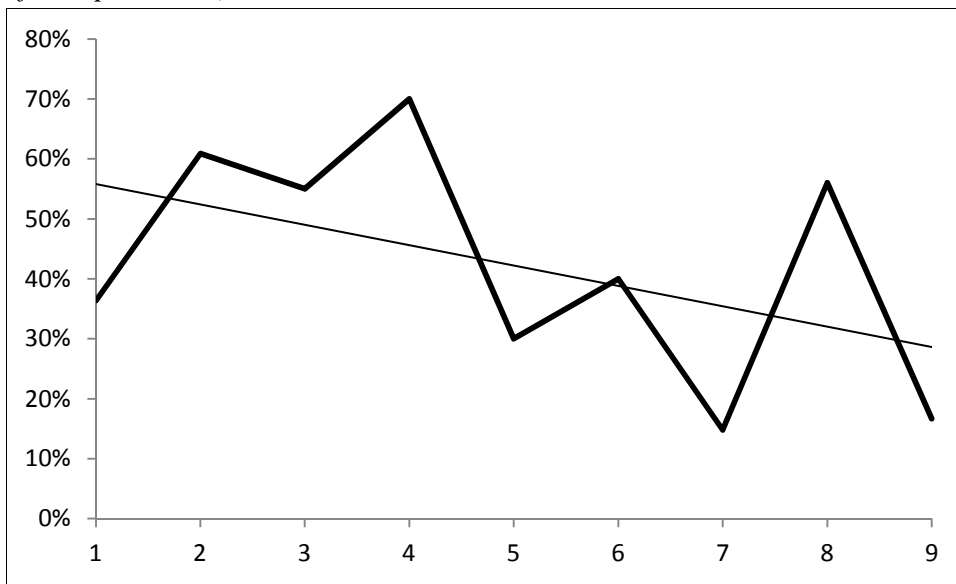
In sum, the difference of 7 percentage points between the nominative and accusative occurrences of the A subjects in Table 4.14. is probably only due to the fact that the internal transitivity variation of the A subjects is more substantial than the variation within the S_A and S_O subject categories. This can, indeed, be seen in the A graph in Figure 4.6.: the A subjects are more evenly distributed on the transitivity continuum than the S_A and S_O subjects. The wider transitivity distribution means that the A subjects assume rather equally all the values between 1 and 9. It is also due to this wider distribution that the difference between the nominative and accusative becomes visible, albeit a difference of 7 percentage points is not particularly large. In contrast, both the S_A and S_O subjects are quite homogeneous categories as far as transitivity is concerned, even though they are located at completely different levels (S_A around 50% and S_O around 15%). Because of this homogeneity of S_A and S_O, both the nominative and the accusative form subjects of these categories are mainly representative of the transitivity level that corresponds to the entire S_A or S_O categories.

In spite of the oscillation, the accusative subject trend seems to be descending in Figure 4.11. This corroborates the above conclusion: the A subjects are scattered across a transitivity continuum so that the higher the transitivity degree is, the lower the accusative percentage. This is, obviously, completely consistent with the pattern observed in the whole case study sample (Figure 4.10.). This conclusion has, however, important consequences relative to the generalisability to LLCT of the theoretical views presented in chapter 3. The theory of alignment dynamics suggests that the extended accusative encroaches first on the (inanimate) low-transitivity subjects and only after that on the (more animate) higher-transitivity subjects.

From this viewpoint, it is only normal that the A and S_A subjects of LLCT with almost equal transitivity degrees display similar accusative percentages: the percentages are, indeed, almost of the same size (15% and c.10% of the accusative for A and S_A, respectively). It is true that the mutual size order of the percentages 15% and c.10% does not comply with the expectations. This is, however, likely to be connected to the fact that was revealed in the above passages (Figure 4.6., Table 4.14.): the different construction types do not form a linear continuum but are located in a very peculiar way on the transitivity scale. In the next chapter, I shall have a look at some of the special transitivity features of the construction types.

Before proceeding to the analysis of A and S_A constructions, I shall summarise the discussion of this chapter. The chi-square tests that were performed in this chapter showed a statistical dependence between the following dependent/independent variable pairs (Table 4.11.):

- transitivity degree and construction type,
- transitivity degree and animacy class,
- case form of subject and transitivity degree.

As a consequence, all the null hypotheses H_0^1 to H_0^3 were rejected and the corresponding research hypotheses H_1^1 to H_1^3 accepted. In other words, the transitivity degree value produced by Hopper and Thompson's transitivity component analysis can be regarded as a reasonable tool for studying the subject case selection in the Latin of LLCT in spite of certain defects of the model that will be better discussed in the next section.

4.3.3. Hopper and Thompson's scale and the S_A and A subject constructions of LLCT

This section examines briefly the transitivity degree of certain construction types of LLCT: constructions with S_A and A subjects, i.e. unergative and transitive clauses. A detailed examination is to the purpose because, on several occasions, the construction types of LLCT have turned out to be not as prototypical in their transitivity as might be expected. The previous chapter revealed that the different construction types are located across Hopper and Thompson's ten-degree transitivity scale in a non-linear manner (Figure 4.6.). In addition, the internal transitivity distribution of the construction types was proved to vary considerably (Table 4.14.). As for the subject case marking, it was noticed that S_A and A subjects display almost equally low accusative percentages as well as almost equal transitivity degrees (Table 4.13. in section 4.3.2.). What was most striking is that the transitivity degree of the S_A and A subjects is only about 50%. In this section, the construction/verb classification 'transitive', 'unergative', 'unaccusative' will often be used instead of or along with the A, S_A, and S_O subject classification, as the focus is on the construction or the verb, not on the subject itself.

The S_O subjects of LLCT behave as expected, i.e. they cluster over the transitivity span 0–3. Hopper and Thompson's scale succeeds in distinguishing the S_O subjects from the other construction types, so the S_O subjects will not be discussed here. Instead, the unergative constructions of the S_A subjects of LLCT often have more in common with the transitive than with the unaccusative constructions, although syntactically the one-participant intransitive S_A and S_O constructions contrast with the two-participant transitive A constructions. The alignment of A and S_A is, of course, what is postulated by the model of semantically-based morphosyntactic realignment (see section 3.1.). It is, however, not theoretically or cross-linguistically obvious whether the presence of an agent is more important for transitivity than the presence of a patient.[262] In effect, the syntactic affinity of the intransitive S_A and S_O constructions and the syntactic disparity of the S_A and A constructions cannot be directly projected onto the semantic differences between these constructions. In this sense, the fourfold distinction of construction types that was introduced in section 3.1. and section 4.1.2. is nothing but a rough generalisation to be used as a tool. The LLCT-based observations verify these considerations.

I reprint here Table 4.13. as Table 4.15. As was just mentioned, both the unergative and transitive clauses have a sort of agent. Therefore, parameter 'agency' displays a rather high value with both the

---

[262] See Kittilä 2002, 80–90.

A and S_A constructions.[263] Agency is a fuzzy term that involves a bundle of control-related inherent semantic features of the subject and intersects at least with 'action' and 'volitionality'. The 78% of the S_A subjects is, however, higher than the 54% of the A subjects, and the same kind of difference in percentage is also found with the mentioned parameters 'action' and 'volitionality' (both 100% with S_A, but only c.65% with A).

Table 4.15. Transitivity component values of each construction type.

Construction type	2 participants	Action	Telicity	Punctuality	Volitionality	Affirmation	Realis mode	Agency	Affected O	Individuat. O	Σ mean	N
A	79%	65%	57%	2.7%	64%	90%	46%	54%	7.0%	44%	51%	187
S_A	0%	100%	75%	0%	100%	88%	44%	78%	0%	0%	49%	77
S_O	0%	0%	5.3%	3.4%	1.0%	97%	41%	0%	0%	0%	15%	207

According to Table 4.15., four of the object-unrelated parameters have a considerably higher percentage with S_A subjects than with A subjects. These four parameters are action, telicity, volitionality, and agency – they are all crucial. Only punctuality, affirmation, and the realis mode show a contrary tendency or, better, they do not contrast the construction types well enough. If the two-participant-related parameters are excluded, the following pattern results (Table 4.16.).

Table 4.16. The participant-number-insensitive transitivity component values of each construction type.

Construction type	Action	Telicity	Punctuality	Volitionality	Affirmation	Realis mode	Agency	Σ mean	N
A	65%	57%	2.7%	64%	90%	46%	54%	**54%**	187
S_A	100%	75%	0%	100%	88%	44%	78%	**69%**	77
S_O	0%	5.3%	3.4%	1.0%	97%	41%	0%	**21%**	207

Now, the mean transitivity degree of the S_A constructions is significantly higher (69%) than that of the A constructions (54%). Indeed, this can be deduced also from the differing distributions of S_A and A over the transitivity continuum in Figure 4.6.: the S_A subjects peak at value 6 instead of value 9 because the three parameters related to two participants display value 0 while almost all the other

---

[263] Ahearn 2010, 37–41.

parameters have value 1. This is certainly a fact that reduces the utility of Hopper and Thompson's transitivity scale in corpus linguistics. Moreover, the mean transitivity degrees of Table 4.16. (A 54% – S_A c.69% – S_O 21%) seem to reflect better, obviously in reverse order, the accusative percentage distribution of the construction types (A 15% – S_A c.10% – S_O 31%) that were seen in Table 4.6. in section 4.2.2. (and in Table 4.4. where the 2nd declension names are included).

It was said above that, of the object-unrelated parameters, punctuality, affirmation, and the realis mode do not contrast the construction types very well. On the other hand, these parameters cannot be deemed as important as the other parameters. As far as I know, no study has suggested that the affirmative/negative contrast would essentially affect the grammatical encoding of Latin. The same applies to the realis mode, which is likely to be tightly related to the communicative function of the data that, in this case, is exceptionally hypothetical with several conditions and stipulations proper to the charter genre.

In addition to the transitivity features indicated in the above tables, the S_A subjects of LLCT are highly individuated and often referential animate nouns. It is easy to imagine that they prefer the nominative case. The most frequent verb lemma with an S_A subject is *venire* 'to come' with 30 occurrences. This is 39% of the 77 S_A instances of the case study sample. In addition to *venire*, the unergative verbs include (*de*)*servire* 'to serve' (6 times), *ire* 'to go', *exire* 'to exit', *vadere* 'to go' (5 times), *appropinquare* (4 times) 'to approach', *reverti* 'to return' (3 times), and other sporadic motion verbs.

A considerable number of the S_A subject nouns (27, i.e. 35%) are personal names. Moreover, *homo* 'man', an individuated noun that will pass to the Romance in its nominative form (It. *uomo*, Rum. *om*, Ancient Fr. *on*), appears 13 times as the subject. It is not, however, always definite or individuated because it is preceded by e.g. *qui*(*s*)*cumque* 'whoever' (32). The other frequent subjects are: *partes* 'parties' (6 times), *filius/filii* 'son/-s' (6 times), *sacerdos/-tes* 'priest/-s' (5 times). It is noteworthy that low-animacy, i.e. inanimate, S_A subjects do not exist at all in the case study sample while they are not infrequent as A subjects (42 occurrences). This was already touched upon in the decision tree in Figure 4.3. in section 4.2.1. Phrases (31) and (32) present typical instances of S_A subjects:

(31) MED 437 (AD 820) *et <u>ille homo</u> qui in ipsa casa avitaverit semper ad* [= *ab*] *mandato nostro venire debeas* [= *debeat*] *ad iustitiam faciendo*

151

"and the <u>man</u> who will be living in this house must always come by our command to assist us in any litigation "

(32) MED 401 (AD 816) *nam si <u>alter quiscumque homo</u> venerit absque nostro conludio*
"if <u>any other man</u> will come [there] without causing any damage [to our interests]"

To summarise, both the verbs and subject nouns that usually occur in the S_A constructions belong to rather limited inventories: the verbs are mainly motion verbs and the subjects highly animate definite agents, including personal names. The A subject constructions show a more multifold picture, so it is not possible to present 'typical' instances of the A subjects of LLCT. The graphics of the previous chapter revealed that the internal transitivity variation of the A subjects is far greater than the variation within the S_A and S_O subject categories. This means that the A subjects occur quite evenly with all the values between 1 and 9 (Figure 4.6.). Figure 4.11. further showed that the A subjects depend on the subject case selection in terms of an internal transitivity continuum of their own.

Indeed, emphasising the border between transitive and intransitive constructions does not seem justified on the basis of evidence of LLCT if analysed with this specific model of transitivity. A solution might be to replace the fourfold construction type classification (A, S_A, S_O, S_O passive) with a comprehensive transitivity degree analysis that would be tailored to the special needs of Latin. Hopper and Thompson's model is aimed at recognising transitivity cross-linguistically and, therefore, some of the parameters are necessarily redundant or dysfunctional in language-specific contexts. The guidelines for this kind of customised transitivity scale will be proposed at the end of this section.

One of the reasons why many S_A subjects of LLCT exceed the A subjects in transitivity is bound, of course, to my definition of transitivity, which is based on the semantic valency of the verb (section 4.1.2. to section 4.1.4.). In this study, even those semantically transitive verbs that lack an overt object are classified as transitive although they behave syntactically intransitively. In a certain sense, the border between transitive and intransitive verbs is impermeable: the semantically transitive verbs *can* (but not *must*) occur with a genuine object while the intransitive verbs cannot. However, when Hopper and Thompson's model is applied, the seven parameters that are not connected to the number of participants easily surpass the influence of the participant-related parameters. Due to the very same parameters, it is technically impossible for the S_A subjects to score higher than 7. In this respect, it is even more striking how many of them do score 6, which is actually the maxi-

mum of the curve in Figure 4.6. The transitivity continuum is operational in its entirety only for those A subjects that have a realised object.

Rovai (2005) points out that most of the previously known Latin cases of the accusative subject with transitive verbs involve an atypical subject characterised by a low degree of animacy, as in (33) and (34). Rovai and Cennamo consider that the non-prototypical inanimate agent subjects of transitive clauses represent an intermediate stage between the prototypical animate agent subject and the prototypical inanimate patient object. The transitive constructions with low-animacy subjects are also seen as the pathway or the point of departure for the extension of the accusative from intransitive constructions to transitive ones.[264]

> (33) *si iumentum <u>morbum renalem</u> temptaverit* (*Mulomedicina* 55)[265]
> "if a <u>renal ailment</u> troubles the beast"

> (34) *<u>fontem</u> vero ubi testa saniam radebat quater in anno colorem mutat* (*Egeria*, *excerpta* 13.1)
> "indeed, the <u>fountain</u>, where he scraped the pus with a crock, changes its colour four times a year"

As was seen above, the S_A subjects are usually highly definite personal names or other animate nouns, and this is certainly one of the reasons behind their high transitivity degree. On the other hand, it is a question of construction type as well: those transitive verbs that allow a non-prototypical inanimate subject are often non-prototypical transitive verbs in so far as they display only a few high-transitivity parameters. They are typically experience or change-of-state verbs, as in (33) and (34). Conversely, several transitive constructions with a prototypical highly animate A subject are also prototypically transitive. In the corpus of Rovai (2005), these constructions do not display other accusative subjects than those with the controversial plural ending -*as*.[266] In LLCT, the highest-scoring accusative-form A subject of transitive construction is *genitore* in (35). Its transitivity degree is 8.[267]

> (35) MED 524 (AD 834) *quas ipsius* [= *ipsi*] *ecclesie offeruit <u>ipse genitore nostro</u>*

---

[264] Rovai 2005, 63, 69; Cennamo 2009, 324–326.
[265] Here, the accusative form may be related to the fact that the argument structure and meaning of the verb *temptare* is rather ambivalent. See Adams (2013, 247, 253) for the problems involved in the *Mulomedicina* text tradition.
[266] Rovai 2005, 86–87.
[267] In the case study sample, there is even one accusative-form A subject with degree 9, but it is a 2nd declension personal name (*Ansuartu*), so I omit it here.

"which <u>our father</u> donated to that church"

The transitive construction prototypically favours an agentive subject and an agentive subject is prototypically human or at least animate. Nevertheless, although several A subjects are highly animate and referential, the construction type also contains many low-animacy nouns contrary to the S_A constructions where the lexical pigeonhole is much narrower. For example, the construction with the inanimate subjects *lex* 'law' and *ratio* 'reason' in (36) scores only 1. Similarly, an S_O subject, such as *Paulus notarius* in (37), may be very high in animacy, but the construction still has a low score (here 2).[268] The A and S_O subjects can be any lexemes, while the S_A subjects are practically only personal names or other (highly) animate nouns.[269]

(36) CDL 255 (AD 771) *nec <u>lex</u> nec <u>ratio</u> contenit ut ipsa femina cum custodes ecclesie simul inhabitet*

"neither <u>law</u> nor <u>reason</u> allows that the woman lives with the church wardens"

(37) MED 564 (AD 840) <u>*Paulus notarius*</u> *domni imperatoris interfui*

"<u>Paulus, the notary </u>of the lord Emperor, was present"

On this occasion, I point out that Hopper and Thompson's parameter realis/irrealis mode correlates clearly with the highest degree of animacy of the subject (personal names) in the case study sample. Table 4.17. shows that constructions with personal name subjects are much more often in the realis mode than constructions with less animate/referential subjects. This reflects the fact that named human agents are the active participants of real world actions: their doings are reported objectively in the indicative mood (and mostly in affirmative sentences). The lowest mode percentage, i.e. 27% of the animate common noun subjects, may relate to the above-mentioned formulaic and rather 'indefinite' use of *homo* and other animate common noun S_A subjects in general stipulation and reservation clauses (see (31) and (32)). It is interesting that even parameter affirmation seems to follow the same distribution although the differences between the animacy classes are less conspicuous.

---

[268] The personal name S_O subjects are restricted to certain few verb lexemes, such as *adesse* 'to be present', *esse* 'to be', and *residere* 'to reside'.

[269] Note that, in the case study sample (see Appendix 4.3.), the subjects *Deus* 'God' and *Dominus* 'Lord' are quite frequent. Both are here interpreted as personal names, as their referents identify with (worldly) persons. *Deus* can, for example, possess churches, etc. in the charter formulae on a par with humans. Both *Deus* and *Dominus* are more often in the nominative than in the accusative case, which may be related to their special agentive status, but can also be influenced by the fact that *Deus* and *Dominus* often occur as the abbreviations *ds* and *dns*, respectively. The graphic appearance of the abbreviations is likely to have sustained the nominative ending in -*us*.

Table 4.17. Transitivity parameters affirmation and mode by animacy class of subject.

Animacy	Affirmation	Realis mode
inanimate	94%	32%
animate	87%	27%
personal	100%	80%

Next, I shall summarise briefly the observations that have arisen in the previous sections concerning Hopper and Thompson's transitivity scale and scalar transitivity models in general. First, I shall outline the successes and drawbacks of the utilised model. After that, I shall suggest how a better-suited case-specific model for (charter) Latin can be created.

Hopper and Thompson's model examines the subject case selection chiefly, but not exclusively, from the viewpoint of construction type properties. In section 4.2., I examined the case selection departing from the properties of the subjects, i.e. their animacy class. Both methods seem to indicate a presence of (the residues of) a semantically-based case marking in LLCT. It is theoretically important that the two methods produce results that are substantially in accord with each other. This cannot result from the overlap of the two analysis methods because no transitivity component is this strictly related to the animacy of the subject. In this respect, Hopper and Thompson's model does rather well.

In spite of this success, Hopper and Thompson's transitivity component analysis is far from ideal for the purposes of the present study. This is obvious, as Hopper and Thompson's scale is a tool for roughly evaluating the transitivity of various constructions from a typological viewpoint. It was seen that the two-participant parameters are more or less off the point in a study where a semantically-based morphological alignment is traced. In sum, transitivity is not distributed evenly across the conventional fourfold construction type classification although this is what is usually supposed to be the case.

The two-participant parameters are, indeed, one of the weaknesses of Hopper and Thompson's model. The authors consider, perhaps for justified cross-linguistic reasons, that O and the object-related properties are equally or even more important than the subject-related properties. It is not possible to go into the details of the theory of transitivity here, but it is enough to say that it is highly problematic to expect high individuatedness of both A and O. This does not reflect real-world situations where patients tend to be much less individuated than agents. Indeed, Naess (2007) em-

phasises that transitivity boils down to distinguishing maximally between A and O.[270] In this respect, a construction with both highly individuated A and O should be viewed as non-prototypical.

It also turns out that Hopper and Thompson's scale is best suitable for traditional school-book examples of high-transitivity. Conversely, it is not particularly appropriate to distinguish between low-transitivity constructions. This can possibly be fixed by completing the component analysis with additional parameters that concentrate on low transitivity domains. It might be useful, for example, to be able to distinguish between S_O constructions expressing a state and a change of state (e.g. between *esse* 'to be' and *fieri* 'to become'). Indeed, after their 1980 article, Thompson and Hopper themselves have ended up criticising transitivity studies for drawing on invented school-book material. As a reaction, their 2001 article is based on transcribed spoken texts.

On the basis of the previous considerations, it is best to tailor a custom transitivity scale for assessing the transitivity degree of Latin texts. This comprehensive multi-component scale must somehow combine all the so far discussed variables that have been examined in order to show their impact on the subject case selection in LLCT. First, the tailored model must take the inherent properties of the subject into consideration in a more systematic way than Hopper and Thompson's scale. In this way, it would be easier to assess the degree of control that the verb has over the verbal event, which was shown to be a cross-linguistically important factor in semantically-based alignments. The extended animacy hierarchy, which is the basis of the animacy/referentiality hierarchy utilised in this study (see section 4.1.1.), can be dismantled into animacy hierarchy and referentiality hierarchy and enriched with other relevant dichotomies/hierarchies. The revised model should include:

1) animacy proper (human, animate, inanimate),
2) subject-inherent referentiality (pronoun, proper name, common name),[271]
3) contextual referentiality (definiteness, specificity, anaphoricity),
4) person (1st, 2nd, 3rd).[272]

Second, the parameters relative to the verbal event must be revised. Parameter kinesis, which consists only of action/non-action, has to be opened up into more dichotomies. For example, a fourfold classification of aktionsart, according to Vendler (1957), i.e. state, action, achievement, and accom-

---

[270] Naess 2007, 22–24.

[271] Following Timberlake (1977, 162), the analysis of subject-inherent referentiality could be made more sophisticated by classifying common names according to the dichotomies concrete/abstract and count/mass.

[272] As for these hierarchies, I refer again to Croft 2003, 130. In this study, only the 3rd person is concerned: as personal pronouns are excluded from the basic query subset, only the lexical subjects remain.

plishment, may turn out to be useful. Parameter 'agency' can be omitted because the above-mentioned subject-inherent parameters take the agentivity of the subject sufficiently into account.

Finally, I propose that the parameters that involve O should not be taken into account when counting the transitivity degree of syntactically intransitive constructions, i.e. the one-participant constructions with unaccusative and unergative verbs. In this way, the syntactic factor would not demote the transitivity value of these syntactically intransitive constructions, and their transitivity value could be better compared with that of the syntactically transitive constructions, i.e. the two-participant constructions with semantically transitive verbs. The parameters involving O are to be included, however, in the analysis of transitive constructions, but they merit a more detailed scrutiny than in Hopper and Thompson's model. The parameter individuatedness of O can be replaced by the parameters that were proposed to be utilised above for analysing the inherent properties of the subject.[273]

4.4. Verbal semantics and the Latin charter genre
4.4.1. Verbal panorama of LLCT

The preceding observations on the relatively low transitivity degree of the LLCT clauses elicit the question about the quality of verbs that are typical of the charter Latin text type. It is intriguing that even the transitivity degree of transitive clauses with nominative-form subjects remains as low as 54% (Table 4.14.). It seems to be worth having an overview on the verbal landscape of the LLCT finite verbs in addition to the sample analysis performed in section 4.3. The immediately following passages seek to characterise charter Latin from the point of view of verbal semantics by analysing the most frequent verbs of the basic query subset. After that, some of the most frequent verbs and verb groups of LLCT will be discussed concerning their transitivity and other semantic features. Section 4.4.2. summarises what has been said about the connection of verb types of LLCT, their transitivity, and the charter genre.

Table 4.18. presents the top 30 verb lemmas of both transitive and intransitive verbs of the LLCT basic query subset. LLCT contains 648 transitive clauses whose infinite-form main verb is modified by a modal auxiliary verb in 131 cases. The finite main verbs plus the infinitive heads of the modal-verb constructions comprise 134 different verb lemmas. The 7 most frequent lemmas form 50% of

---

[273] A model that places subject-related, verb-related, and object-related parameters over a scale would probably be the only way to avoid an overlap between parameters as far as that is possible (Malchukov 2006, 330–334).

the total, while the 40 most frequent lemmas reach 80%. As many as 46% of the lemmas occur only once but, on the other hand, they correspond to only 9.4% of the total number of clauses. All this is symptomatic of the highly repetitive nature of the charters. Intransitive clauses (excluding passive clauses) are 702, and 85 of them contain a modal auxiliary verb. As the intransitive verbs consist of only 68 different verb lemmas, the lexical diversity of verbs is lower than with the transitive clauses: the most frequent lemma, *esse* 'to be', forms 45.2% of the total, and the 6 most frequent lemmas are responsible for as much as 80% of the total occurrences of intransitive clauses. However, 32% of the lemmas occur only once, and they correspond to only 3.8% of the 702 clauses.

Table 4.18. The 30 most frequent verb lemmas of transitive and intransitive verbs (basic query subset).

Transitive verbs				Intransitive verbs			
habere[274]	134	*reddere*	7	*esse*	317	*manere*	4
dicere	67	*accipere*	6	*adesse*	131	*obvenire*	4
facere	40	*deprecari*	6	*residere*	54	*persistere*	4
dare	36	*construere*	5	*venire*	29	*recedere*	4
donare	20	*emittere*	5	*permanere*	22	*accrescere*	3
confirmare	15	*inferre*	5	*habitare*	9	*deservire*	3
dirigere	15	*iudicare*	5	*oriri*	7	*exire*	3
decernere	14	*admittere*	4	*remanere*	6	*mori*	3
offerre	13	*aedificare*	4	*advivere*	5	*percurrere*	3
respondere	13	*agere*	4	*ire*	5	*pertinere*	3
venumdare	11	*eligere*	4	*interesse*	5	*servire*	3
continere	10	*intentionare*	4	*reverti*	5	*testificari*	3
possidere	9	*iubere*	4	*subscribere*	5	*advenire*	2
profiteri	9	*manifestare*	4	*apparere*	4	*complicere*	2
componere	7	*mittere*	4	*appropinquare*	4	*fieri*	2

It is to be emphasised that the numbers cannot be fully compared with those of other corpora because here, as everywhere in this study, by 'clause' a subject/verb combination is understood. Once there are several cases where there is more than one subject, the same verb is counted as many times as there are subjects attached to it. This especially skews the numbers concerning intransitive clauses, which typically display formulaic structures with several coordinated subjects.

---

[274] The English equivalents of these transitive verbs are (to) 'have', 'say', 'do', 'give', 'donate', 'confirm', 'direct', 'decree', 'offer', 'answer', 'sell', 'contain', 'possess', 'profess', 'pay', 'solve', 'receive', 'pray', 'construct', 'send', 'bring', 'condemn', 'admit', 'construct', 'act', 'choose', 'contest', 'order', 'manifest', 'send'; intransitive verbs (to) 'be', 'be present', 'reside', 'come', 'remain', 'dwell', 'rise', 'remain', 'live', 'go', 'be among', 'return', 'subscribe', 'appear', 'approach', 'remain', 'come', 'persist', 'recede', 'increase', 'serve', 'exit', 'die', 'run', 'pertain', 'serve', 'witness', 'come', 'please', 'become'.

Generally speaking, the verbal panorama of LLCT charter Latin is one of strikingly low transitivity. Many of the top 30 verb lemmas of transitive clauses are very low in transitivity, and the same applies, even more pronouncedly, to intransitive verbs (especially to the top 5 of them). The only relatively frequent intransitive-clause verb lemma with a higher transitivity degree is the unergative motion verb *venire* 'to come' with 29 occurrences. I shall examine now in more detail a few of the most frequent transitive and intransitive verbs. I start with the transitive verbs, where it will be seen that some of the apparently high-transitive verbs, such as *facere* 'to do', are actually rather low in transitivity.

Habere/possidere

Habere 'to have' is the most typical transitive verb of LLCT with its 20.7% share (134 occurrences) of all the transitive clauses. *Possidere* 'to possess' is far less frequent (9 occurrences). As the charters usually deal with possessing something, it is not surprising that 45% of *habere* refers to real possession. These verbs head objects that are concrete nouns (e.g. *casa* 'house', *hortus* 'garden', *solidi*, *tutor* 'carer'; see (38)) or referential relative pronouns (*qui*) with concrete noun antecedents (39). What is surprising, on the other hand, is that 55% of instances of *habere* do not refer to real possession. I call non-real, abstract possession the use of *habere* in which a) the object of *habere* is a non-referential, indefinite, or otherwise non-individuated pronoun (e.g. *quisque* 'anyone', *quantum* 'how much'; see (40)), or b) there is no object ('deleted object'; see (41)), or c) the object is an abstract noun that, together with the verb, constitutes a phraseological construction, such as *potestatem/licentiam habere* 'to be at liberty' (18 and 11 occurrences, respectively) (42). In cases like (40), *quantu*(*m*) could be viewed as an adverb, were it not to agree in gender with its feminine-gender antecedent where there is one (usually there is none) (43). The object distribution of *habere* can be seen in Figure 4.12.

(38) CDL 28 (AD 720) *ubi sibi abbas uel monachi iniui consistentis ortum uel pigmentarium hauire debeat*
"where the abbot and the monks that reside there have to have a garden or *pigmentarium*"

(39) CDL 253 (AD 771) *et ipsa uinditione quas eorum* [= *eis*] *Brittulo fecit*
"and the sales contract that Brittulo made to them"

(40) MED 303 (AD 802) *quantu ipsi quondam Prandulo ad manum sua abuet*

"as much as the late Prandulo had in his hands"

(41) CDL 139 (AD 759) *res illa* [--] *qualiter Baruccio ad manum suam habuit, uobis dedi ad resedendo et laborando*

"I gave you that property to be resided and cultivated [--] in the same manner as Baruccio had [it] in his hands"

(42) CDL 281 (AD 773) *tam tu quam et ille homo* [--] *potestatem abeatis uendere et dispensare medietate ex omni re mea pro anime meae remedio*

"both you and that man [--] are at liberty to sell and spend half of all my property for the salvation of my soul"

(43) CDL 283 (AD 773) *et omnem rem* [--] *mihi tradidisti ad laborandum* [--] *quantam Magnaris clericus et genitor eius Deusdedit abuerunt*

"and you gave me to be cultivated the whole property [--] as much as Magnaris, the clerk, and his father, Deusdedit, possessed"

Figure 4.12. Object distribution of verb *habere* 'to have' ($N = 134$) in the LLCT basic query subset.
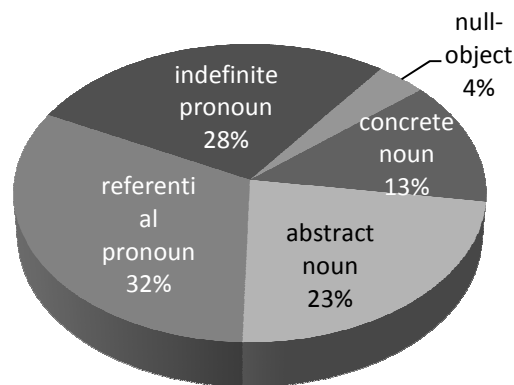
Even though *habere* is semantically and usually also a syntactically transitive verb, it expresses a state. At least, there is no real action involved, as 'having something' mainly characterises the subject, i.e. indicates its property, a feature considered to be typical of stative (intransitive) verbs. The difficulty of assigning the actor and undergoer macroroles to *habere* reveals the non-prototypical transitivity of the verb. The semantic role of the subject could be location while the object seems to be theme. Both are non-prototypical.

Dicere

Dicere 'to say', but also 'to speak', occurs typically in trial records (*notitia iudicati*) and in other documents involving verbal statements. As was stated in section 4.1.4., *dicere* and other speech verbs differ radically from prototypical transitive verbs as far as their syntactic behaviour is concerned: only 6% of the cases (see Figure 4.13.) have a nominal object (44) while 61% have as the object a clausal complement (45) or a direct quotation (46). (Instead, the Classical Latin standard method, the accusative and infinitive construction, is not found at all.) The rest, 33%, are elliptical, i.e. they do not have an overt object at all (47). The object distributions of *respondere* 'to answer' and *profiteri* 'to profess' are very similar to that of *dicere*. These verbs also occur mainly in *notitia iudicati*.

(44) CDL 84 (AD 745) *omnia in adpretiato in tregenta soledus tibi uendere uisus sumus*, *quem* [= *quod*] *tertio homo Deus timente dixere* [= *dixerit*]
"it is certain that I have sold you everything, appropriately priced, at thirty *solidi*, which the third God-fearing person will certify"[275]

(45) MED 309 (AD 803) *iurando dixerunt ut veritatem fuissent*
"they said under oath that it was the truth"

(46) CDL 255 (AD 771) *iterum pars ipsius infantuli dicebat*: '*si placet uobis*, *uolumus istum Petrum clericum ut nobiscum habitet*'
"the party of the infant said again: 'we want that Petrus, the clerk, lives with us if it suits you'"

(47) MED 539 (AD 838) *Savinus similiter dixit*
"Savinus spoke similarly"

---

[275] Apparently, a trustworthy, i.e. God-fearing, witness was needed in some cases. I suppose he was the 'third person' in addition to the buyer and the purchaser.

Figure 4.13. Object distribution of the verb *dicere* 'to say' ($N = 67$) in the basic query subset of LLCT.
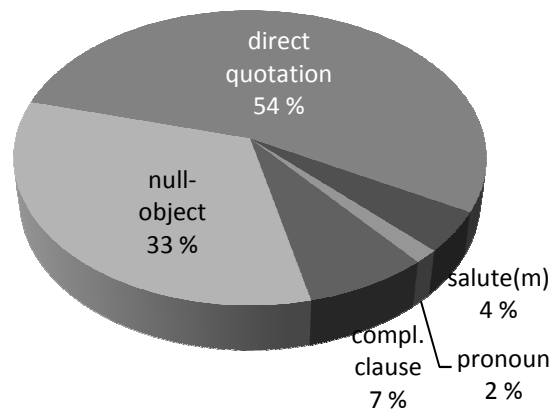
Semantically, the subjects of speech verbs are agents, but their objects are effected, i.e. non-prototypical. Because of their non-prototypical syntactic behaviour, speech verbs, such as *dicere*, usually seem to be considered intransitive, except when they have a nominal object. In many languages, there are different lexical items for intransitive 'speaking' and transitive 'saying', albeit the verb senses often overlap, as in Latin *loqui* and *dicere*. Thompson and Hopper (2001) view speech verbs (and many other verbs) as intransitive even when they have a clausal object complement. In fact, this interpretation is related to their denying the very existence of clausal object complements (at least in phraseological contexts). However, semantically clausal object complements do belong to the argument structure of the verb, as they are part of the semantic valency.

Facere

The verb *facere* 'to do/make' ($N = 40$) is potentially high in transitivity because the subject is usually agentive and high in volitionality. On the other hand, the object of *facere* is often not prototypical but an effected one. However, even the agentivity and volitionality of the subject are also often diminished by the fact that the verb is used in a weakened sense 'to cause' in several highly common phrases, such as *voluntatem facere* 'to make sb's will' (both concrete and abstract), *iustitiam facere* 'to do justice', but also *placitum/cartulam facere* 'to make an agreement/charter' (not denoting the physical act of compiling a charter, but the abstract, juridical deed) (48). *Facere* is almost reduced to an auxiliary that, when combined with an abstract or concrete noun, forms a new phrasal verb construction (not in the English-grammar sense) with abstract meaning. Yet, 'to do' is not a void verb in these contexts, as it conveys its causative contribution to the new construction. These con-

structions are extremely frequent in several languages: e.g. It. *fare le scale* 'climb the stairs' or Eng. *make a decision* 'to decide'. *Facere* also occurs in many other figurative connections, such as (49). In (50) and (51), *facere* is used as a clear causative auxiliary verb. Intriguingly, the only clearly high-transitive *facere* with a concrete (effected) object is found in (52).

(48) CDL 247 (AD 770) *in presentia suprascriptorum testium* [--] *fecerunt inter se placitum Gaudiosus presbiter et Humulus clericus*

"in the presence of the above-mentioned witnesses [--] Gaudiosus, the priest, and Humulus, the clerk, compiled a *placitum*"

(49) CDL 125 (AD 757) *Deus fecit hominem ad imaginem sue similitudinis* (4 times)

"God made man in His own image"

(50) CDL App. (AD 786) *ipse domnus Iohannes episcopus ante se uenire fecit ipsum Alpertulum clericum*

"Lord Iohannes, the bishop, made Alpulus, the clerk, come before him"

(51) MED 539 (AD 838) *ipse episcopus eam pertinentem episcopatui sui faciebat*

"the bishop incorporated it into his bishopric"

(52) CDL 85 (AD 746) *cunsuetudo ipseius case quas* [= *quam*] *parentis nostris ficerunt*

"the rent of that house which our relatives used to pay"

Dare/donare/offerre

Verbs denoting 'giving', such as *dare* 'to give', *donare* 'to donate', *offerre* 'to offer' (*N*= 69), are the largest group of relatively high-transitive verbs of LLCT, together with the few instances of *aedificare* 'to build' and *construere* 'to construct'. The subjects of these verbs are agents, but the objects are themes, not patients. Thus, the objects, albeit concrete, individuated, and referential, are not completely affected, which is one of the high-transitivity criteria. The group is rather uniform and can be illustrated exhaustively by the following three examples (53), (54), and (55).

(53) MED 246 (AD 794) *id est casa et res mea illa* [--] *quas ipse Periteu mihi in monganicaput dede*

163

"i.e. my house and property [--] that Periteu gave me as a morning gift"

(54) MED 460 (AD 824) *quondam Filipertus clericus avius meus per cartulam donavit mihi medietatem ex omnibus rebus sui*
"my late grandfather Filipertus, the clerk, donated me half of all his property through a charter"

(55) CDT 36 (AD 793) *qui supra Ursiperto clericus* [--] *oblazones meas munera offero*, *non quanta deveo sed quanta baleo*
"I, the above Ursiperto, the clerk [--] donate my humble gift, not as much I should, but as much as I can"

Other common transitive verbs include, for example, verbs that denote ordering (*iubeo* 'to order', *decerno* 'to decree') and confirmation (*confirmare* 'to confirm'). The performative verbs and the verbs that report performative utterances or actions are discussed below after the intransitive verbs because they include both transitive and intransitive verbs.

Esse

I now turn to the most frequent exponents of the intransitive verbs of the LLCT basic query subset. There is much less to say about individual intransitive verbs. As was stated above, their frequency top is constituted of very few verb lemmas that are mostly of very low transitivity: *esse*, *adesse*, *residere*, *permanere* (317, 131, 54, and 22 occurrences, respectively). *Venire* (29 occurrences) and *oriri* (7 occurrences) are the most frequent active or dynamic verbs and, thus, higher in transitivity. In fact, the number of active intransitive verbs is higher within the less frequent verb lemmas, but the total number of clauses of this kind remains insignificant in respect to the huge number of the most frequent, low-transitive ones.

Esse 'to be' is a copula verb that connects the subject and the non-verbal predicate. The predicate can be an adjective, a noun, or a prepositional phrase. In this study, the predicates that include *esse* are not further divided up in subgroups according to the quality or sense of the predicate part. What is common to all the *esse* predicates is that they denote a stative, non-dynamic state. Cases of *esse*

as an auxiliary verb of certain verbal constructions (*esse* + present/future participle) have been excluded from the analysis.[276]

Metaphorical motion verbs[277]

It was stated that *venire* 'to come' is one of the most transitive of the intransitive basic query subset verbs. The motion involved in *venire* is real in (56). Many other potentially high-transitivity intransitive verbs, such as *oriri* 'to rise', *ire* 'to go', *reverti* 'to return', or *obvenire* 'to arrive/come toward' appear to be used figuratively, so that the motion the verbs seem to express is not real, but metaphorical: e.g. the abstract sense 'to originate' of *oriri* is preferred to the concrete 'to rise' (57). The same drive towards figurativeness can be perceived in examples (57), (58), and (59). Additionally, the verbal process is non-volitional and the subject low in control in (57) and (59), as the subjects are inanimate. This semantic property of the subjects is the decisive criterion between S_O and S_A subjects and, thus, between unaccusative and unergative clauses: (57) and (59) are unaccusative clauses while (58) is unergative.

(56) MED 437 (AD 820) *si* [--] *ipsi homo qui in ipsa casa resederit ad mandato nostro venerit ad iustitiam faciendo*
"if [--] the man who resides in the house comes to the court by my order"

(57) CDT 31 (AD 790) *unde nullo tempore hurietur* [= *oriatur*] *intentio aut altercatio*
"so that there will never arise a contention or altercation"

(58) CDT 46 (AD 798) *et nullus de heredibus* [--] *aliquando contra hanc nostra decritione ire* [= *iret*]
"and no-one of the heirs [--] should ever rise against this our decree"

(59) CDL 232 (AD 769) *post uero meum obitum omnis iam dictus meus adquisitus reuertatur in potestate praefatae Dei ecclesiae*
"after my death, all the mentioned equipment of mine shall return in the possession of the above-mentioned God's church"

---

[276] On the verbs *permanere* and *esse* in classical prose, see Spevak 2010, 178–193.
[277] Metaphorical or fictive motion has to be understood here more broadly than the similar, but rather narrow-scoped term coined by Talmy (1996).

There are still two further groups of verbs that I want to deal with briefly. The first of these includes verbs that I call writing act verbs. This group does not align with the transitive/intransitive split, but has members in both construction types. I describe this group here in order to sketch as extensive as possible a picture of the charter Latin verbal semantics even though the group does not have direct importance to the transitivity discussion at hand. The second group that will be discussed is modal auxiliary verbs.

'Writing act' verbs

Austin (1962) defines as speech act or performative verbs those verbs whose uttering is, or is part of, the performing of a certain kind of action. The classical examples include expressions, such as (60). Not surprisingly, charter Latin that documents legally binding juridical acts favours speech act verbs, e.g. (61). These 'real' speech act verbs are not, however, pertinent to this study, as they by definition have as their subjects the 1st person singular/plural pronoun (or pro-drop) and, consequently, are excluded from the basic query subset. Moreover, it is all about written language in LLCT. Therefore, I introduce here the term 'writing act verb' that is intended to denote verbs carried out by means of writing them (in a charter) whether they be in the 1st or 3rd person. Since even 'writing act verbs' usually operate with pronouns (or pro-drop), there are not many examples of 1st person cases. See, however, (62) which is a real writing act verb and differs from the speech act verbs in so far that it has not been uttered aloud at the juridical act, but fulfils its juridical function in the very performing of its writing. Phrase (62) has the same symbolic value as the sign of the cross, *signum sanctae crucis*, that the illiterate were asked to draw in place of the subscription (63). Both (62) and (63) are formulaic phrases that occur with minor modifications in a wide variety of charters.

(60) *I <u>name</u> this ship the 'Queen Elizabeth'*[278]

(61) CDL 108 (AD 753) <u>*manifestus sum*</u> *ego Pertifuns quia deuitor* [= *debitus*] *sum dare tibi domno Uualprand episcopo soledus propter casa Auderad*
"I, Pertifuns, <u>make it manifest</u> that I am obliged to give you, Lord Walprand, the bishop, one *solidus* for the house of Auderad"

---

[278] Austin 1962, 5.

(62) CDL 113 (AD 754) *ego Teuderadu presbiter in ach cartula cumutationis facta a Gualprandu episcopus ed Alpertu duc<e?> propia manus mea me teste <u>supcripsi</u>*

"I, Teuderadu, the priest, <u>subscribe</u> in my own hand, as a witness, this exchange record that was made by Gualprandu, the bishop, and Alpertu, the duke"

(63) MED 186 (AD 782) *signum + manus Angeli filio quondam Burichi testis*

"sign + [= Holy Cross] of the hand of Angelus, son of the late Burichus and witness"

As said, the actions performed by 'writing act verbs' are very often reported in the 3rd person. Of course, the verbs are not then performative any longer. However, reporting these actions constitutes an essential part of the juridical probativity and validity of the agreements recorded in charters. In this respect, verbs such as *profiteri* 'to profess', *manifestare* 'to manifest', and *testificari* 'to testify' are a special characteristic of charter Latin, both in the 1st and 3rd persons. Closely related are also *iudicare* 'to judge', *iubere* 'to order', and *decernere* 'to decree'. The following phrases exemplify these verbs.

(64) MED 309 (AD 803) *et dum prefati sacerdotes taliter <u>testificati fuissent</u>*

"and when the above-mentioned priests <u>had witnessed</u> in this manner"

(65) MED 736 (AD 857) *Guntelmus filio bone memorie Gumperti per cartulam iudicati <u>decrevit</u> in potestate mea*, *id est parte sua de res illa*

"Guntelmus, son of the late Gumpertus, <u>ordered</u> his part of that property to my possession through an adjudication"

Indeed, even *dicere* 'to say' and *respondere* 'to answer' can be included in this group as far as they are reporting performative utterances: it is important to record the juridical statements that the participants made, for example, at trial. This type of speech verbs was discussed above in section 4.1.4.

The role of modal verbs

Next, I touch briefly on modal verbs. I shall not, of course, examine the semantics of the modal verbs themselves, but concentrate on their influence on the semantics of the main verb they modify. In Hopper and Thompson's transitivity scale, modal verbs are mainly related to volitionality and

mode (Table 4.15.). The irrealis mode conveyed by modal verbs usually lowers the volitionality degree of the whole verb construction, as in (66) and (67), where (*non*) *posse* 'can (not)' and *debere* 'to have to/must' deprive the subject of part of its independence and, thus, control over the verbal process: doing something forced cannot be genuinely volitional.[279] Once Hopper and Thompson's model gives separate parameters for volitionality and mode, I have decided to interpret the volitionality as a property of the main verb in the transitivity case study of section 4.3. Thus, the influence of the modality modification is evaluated by the realis/irrealis mode parameter, instead.

(66) MED 172 (AD 778) *neque nullo homine* [--] *possit molestare aut resubtragere mea offerta*
"nor can anyone [--] contest and reverse my donation"

(67) MED 437 (AD 820) *et ille homo* [--] *semper ad mandato nostro venire debeas* [= *debeat*] *ad iustitiam faciendo*
"and that man [--] always has to come to the court by our order"

Modal verbs occur with 14.6% of the finite verb constructions of the LLCT basic query subset (19.2% with transitive, 11% with intransitive constructions). Although no comparable Classical Latin corpus data are available, especially the percentage of the transitive constructions seems to be considerable. The most frequent modal verb is *debere* 'to have to/must' (Table 4.19.). This is only predictable, as charter language understandably imposes various obligations to the contracting parties. The second most frequent verb lexeme is *videri*, which in LLCT is interpreted as a modal verb expressing evidentiality, as explained in section 4.1.4. Typical of juridical language are also the solemn verbs *dignari* 'to condescend' (68), *praesumere* 'to dare' (69), and *mereri* 'to merit/deserve', which emphasise obedience to authority, everlasting or temporary, thus reinforcing the hierarchies of society.

(68) CDL 248 (AD 770) *si mihi Dominus respicere dignatus fuerint et filios aut filias procreauero*
"if the Lord deigns to regard me again and if I create sons or daughters"

(69) CDL 35 (AD 724) *et numquam nos uel posterus noster te de hanc dicto loco molestari praesumat*

---

[279] Rovai 2005, 63; Hopper & Thompson 1980, 252. On *debere*, see also Cuzzolin 2010, 254.

"nor should we or our heirs ever dare to expel you from this mentioned place"

Table 4.19. Frequencies of modal verbs in the LLCT basic query subset.

Modal verb	N	Modal verb	N
debere[280]	68	*quaerere*	8
videri	60	*mereri*	4
dignari	22	*nolle*	3
praesumere	16	*valere*	3
velle	16	*solere*	1
posse	12	∑	213

Modal verbs have a more extensive field of application in LLCT than in Classical Latin. The verb *debere* has come to be used as a future periphrasis or at least its obligative modality has been considerably weakened in several formulaic phrases of charter Latin (70).[281] Moreover, modal verbs are very often in the conjunctive or in the future perfect, which increases the irreality of the verbal constructions. *Potuerit* in (71) conveys a hypothetical tone to the clause. The question is no longer about being able to realise the action described by the main verb. The modal verb can be here seen as an extra hypotheticality marker that replaces the potential conjunctive form *introeat* or the future perfect *introierit* that were obsolescent for morphophonological reasons.

(70) CDL 211 (AD 767) *ipse mea offerationem ferma et istabile diueam* [= *debeat*] *permanire*
"my donation will stay firm and stable"

(71) MED 187 (AD 782) *et si quicumque homo vobis in ipso monasterio* [--] *in qualivet portionem introire potuerit*
"and if any man whosoever might interfere with [--] any portion in the monastery to your disadvantage"

To summarise, modal verbs seem to be characteristic of charter Latin. They have several important discourse-related functions within the inventory of formulaic means of expression of the charter genre. In general, modal verbs decrease the transitivity degree of the genre.

---

[280] These verbs are in English (to) 'have to/must', 'seem/be apparent', 'condescend', 'dare', 'want', 'can', 'want/try', 'deserve', 'not want', 'can', 'be used to'.
[281] Väänänen 1981, 132.

4.4.2. Verbal semantics, genre, and transitivity

A detailed examination of the behaviour of individual verbs corroborates the first-sight impression that many of the clauses of LLCT are rather low in transitivity. It is perhaps predictable that the great majority of the intransitive clauses is very low in transitivity, but this applies even to most of the verbs classified as transitive. Some of the most typical charter Latin transitive verbs do not involve real transitive action, but describe states or properties. Others have non-prototypical objects and/or appear in constructions with figurative meaning.

It is interesting to combine the results of the qualitative study of section 4.4.1. with those obtained in section 4.3. from the quantitative transitivity component analysis. Figure 4.14. shows how many clauses there are for each Hopper and Thompson's transitivity component value in the sample of those 471 clauses. It is again seen that the bulk of the clauses is located towards the low-transitivity end of Hopper and Thompson's scale, while there are only a few exponents at the high-transitivity end: no clause scores 10 points and only 6 clauses score 9 points.

Figure 4.14. Distribution of transitivity component values inside the sample.
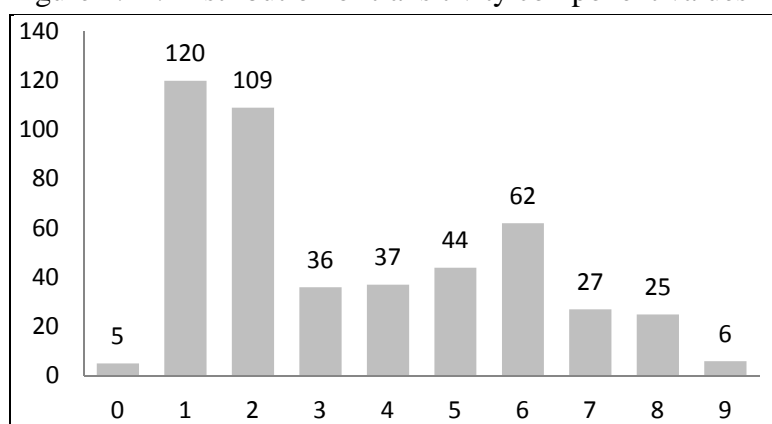
In fact, only clauses of the kind "Jeff killed John" really score 100% in Hopper and Thompson's transitivity component analysis. In most texts, these school-book examples of transitive clauses are extremely rarely found. Intuitively, one would think that perhaps the most natural context of highly transitive clauses are narratives and various reports, e.g. news texts (written or oral), where reporting actions and events is of the essence.

Indeed, studies have proved that transitivity is text-type specific. This relates to the idea that transitivity is usage-based: the speakers/writers adopt a linguistic register appropriate to the current situa-

tion, and the modes of representation vary accordingly.[282] For example, Thompson and Hopper (2001) show that the transitivity degree of modern conversational English is low:

> "[T]he low Transitivity in our conversational data is to a considerable extent determined by the kinds of things we are doing when we talk with friends and acquaintances. We do not seem to talk much about events, let alone actions [--] but rather, our talk is mostly about 'how things are from our perspective'. Our data show that we describe states, reveal our attitudes, ascribe properties to people and situations, and give our assessments of situations and behaviour."[283]

The difference seems to be especially marked between spoken and written registers. Thompson and Hopper (2001) report a 27% transitive-clause portion for conversational English. Vázquez and García-Miguel (2009) show that the corresponding number for modern Spanish is 40%, while the transitive-clause percentage of other, written, genres fluctuates around 44% and 45%. As yet, other relevant genre/transitivity studies on any language, let alone Latin, do not exist. Unfortunately, even the mentioned studies fail to examine the construction types in a more detailed way. Moreover, the criteria on which they classify transitive and intransitive verbs differ from each other.[284] In the LLCT basic query subset, 46.7% of the clauses are classified as transitive, but the percentage cannot be directly compared to the above numbers because I have classified the transitive and intransitive clauses according to the semantic valency of the verbs.

In this chapter, I summarise what has been said in the preceding chapters by relating it to the concept of genre. Texts are classified into genres on the basis of their communicational intent. Since genres are recognisable communicative conventions on how texts are written, they are affected by extra-linguistic features including audience, communicative purpose, and function. These extra-linguistic factors influence the choice of linguistic text type features that are likely to change over time in response to language-users' socio-cultural needs.[285]

---

[282] Vázquez & García-Miguel (2009) put it in a very functionalist way: "The transitivity lexico-grammatical system is conceived here as a resource that allows speakers to select the event's participants they want to talk about, which is crucial for discourse construction. Besides, if syntactic (in)transitive patterning is a long-term consequence of the preferences in usage, it can be assumed that the various contexts of use and the speakers' various communicative goals trigger differences in the relative frequency of each construction."

[283] Thompson & Hopper 2001, 53 (quoted also by Vázquez & García-Miguel, 2009).

[284] Vázquez & García-Miguel (2009) count constructions with complement clause objects as intransitive constructions (thus also Thompson 2002).

[285] Lehto 2013, 237; Taavitsainen 2001, 139–141.

Vázquez and García-Miguel claim that, at least in modern Spanish, the varying transitivity degree of different genres is not primarily a result of the genres applying different construction types (as proposed by Thompson and Hopper (2001)) but of their different degree of subjectivity: the spoken language employs more 1st person subjects than other genres. The whole point is, according to Vázquez and García-Miguel, that the very frequency of speaker/writer subjectivity turns certain mental verbs, such as 'to think', into epistemic discourse markers, such as 'I think'. Thus, Vázquez and García-Miguel offer an explanation to the phenomenon of the phraseologisation of the 'schematic epistemic/evidential/evaluative fragments' discussed by Thompson and Hopper (2001).[286]

Even though Vázquez and García-Miguel's interpretation of the epistemic discourse marker phenomenon seems to be plausible, the results of my transitivity component analysis contradict their claim about the indifference of the construction type. It was seen in section 4.3.2. (see Table 4.15.) that different construction types score very differently in transitivity component analysis. Granted, oral conversation in a modern language is far from the written language of early medieval charter Latin. However, I suggest that the observation that there is a dependence between genre and transitivity also holds good, *mutatis mutandis*, for the genre of charter Latin (in comparison to other Latin genres). It is an issue that has not yet been studied, and unfortunately cannot be studied within the scope of this study. Nevertheless, I am ready to apply the basic ideas of Thompson and Hopper (2001) and Vázquez and García-Miguel (2009) to the charter Latin of LLCT.

I do not suggest that charter Latin would have something important in common with the spoken language. What I suggest is that charter Latin shares some significant features with non-narrative texts. In LLCT, there are practically no accounts of connected events, i.e. narrative parts reporting what someone did and what then happened. The reporting is limited to the speech and writing act verbs that record the juridically important components of the legal act. The most transitive clauses of this type denote giving and confirming something (8 transitivity points) while the 5 clauses rating 9 transitivity points contain the verbs *aedificare* 'to build' and *construere* 'to construct' (see (72)).

(72) MED 231 (AD 790) *quondam Sichipertus construxit ecclesiam in onore Dei et beati sancti Vitalis Christi martiris*

---

[286] Thompson & Hopper 2001, 27–28, 53. The grammaticalisation of clauses, such as 'I think' or 'I don't know', as discourse markers has been related to the progressive subjectivisation of their meaning, which would also explain the weakening of their valency pattern (= less transitive) as a consequence of the loss of their capacity to encode events (Vázquez & García-Miguel 2009; Traugott 1989, 35–36).

"the late Sichipertus constructed a church in honour of God and saint Vitalis, martyr of Christ"

Instead of being narrative, the text type of charters is something that I shall call 'dispositive' language. It is a language type that relates to "bringing about the settlement of an issue or the disposition of property".[287] Thompson and Hopper's (2001) conversational data basically conveys the speakers' stances towards the issues and claims at hand. Charter Latin expresses the commitment of the contracting parties to a shared objective as well as epistemic stances concerning various states of things (e.g. *videri* in (17) and (18) in section 4.1.4.). In this respect, both conversational and dispositive texts are about negotiating an agreement. In both text types, there is no need for narrative, as the underlying facts and the relevant contextual data are already familiar to the participants. Verb choices, of course, are adjusted to the shared efforts. As was noticed earlier, the verbs that denote 'having something' are a speciality of charter Latin and they, too, contribute to lowering the general degree of transitivity.

The main objective of section 4.4. was to survey the verbal panorama of the charter Latin of LLCT and, through this survey, to conceive how the semantics of these text-type specific verbs and constructions is related to the low transitivity degree of the LLCT clauses in general. This was done in order to find out whether there is something special on the lexical level of charter Latin that ought to be taken into account when studying the possible case alignment change. Moreover, it seemed to be useful to clarify whether there is a connection between the transitivity degree of LLCT and the charter genre in general.

One of the most important conclusions that can be made about this section and on the statistics of the previous section is that because of its verb-type distribution, it is actually not reasonable assume that LLCT could reflect very well a possible alignment change: the highly transitive indicator verbs are missing. In other words, it is not possible to know how the subject case selection would work in the highest domains of transitivity because there are no verbs representing those domains. Another, more universal, conclusion is that on the grounds of these corpus data, the traditional dichotomy of transitive and intransitive clauses, based on the number of (syntactic or semantic) participants, is not opportune in all the corpora or in all the genres: the verbs classified as transitive are very heterogeneous and, seemingly, often lower in transitivity than many intransitive verbs.

---

[287] Oxford English Dictionary, s.v. *dispositive*.

4.5. Summary

Chapter 4 examined subject case selection as far as semantic variables are concerned. The main intention was to find out whether there is a statistically significant relationship between semantics and the case form of subject. Another intention was to figure out which one of the discussed variables yields the statistically most reliable theory-compatible model. The statistical testing of the variables that were presented at the beginning of the chapter in Table 4.2. showed that there is a statistically significant dependence between animacy and the subject case as well as between construction type and the subject case. In LLCT, the chi-square tests of the decision trees in Figure 4.1. and Figure 4.2. revealed that the split is more significant when effected by animacy, i.e. the inherent properties of the subject, than when effected by construction type, i.e. the properties of the verb/verbal event. It also turned out that the 2nd declension singular personal names misled the analysis. Hence the need to resort to the 3rd declension singular imparisyllabic subcorpus when it is necessary to exclude all the possible distracting factors. On the other hand, the 2nd declension singular personal names are an essential part of charter Latin and, therefore, it is usually possible and even recommended to use the entire LLCT as long as the influence of the 2nd declension personal names is taken into account.

On the basis of the theoretical considerations of section 4.2.1., it is possible to conclude that traces of a semantically-motivated case marking system are attested in LLCT although the accusative percentage patterns of Table 4.3. and Table 4.4. are perhaps not as clear as one could expect were the conditions optimal. On the other hand, the fact that the classes do not display maximal differences on an ideal synchronic continuum, such as A 0% – S$_A$ 50% – S$_O$ 100%, is likely to result from several features of written language and the mechanics of language change in general. First, the evolution of the spoken language is always reflected imperfectly in the conservative written code. Second, it is not clear in what kind of distributions a diachronic change is supposed to surface.

As Halla-aho (2009) states, language change is usually a slowly emerging tendency that is visible for a long time only as a statistical preference for one variable. As a consequence, the distribution patterns change as the evolution continues. It is, however, often impossible to tell whether a certain evolution is still underway or whether it has already reached its end point. In this respect, the attested extended accusatives of LLCT can be taken to reflect a rather advanced stage on the alignment continuum that proceeds from a nominative/accusative system to a semantically-based system, to an ergative/absolutive system, and finally to a neutralised system: even though in LLCT the focus is

likely to be on the semantically-based phase, the data seem to involve traits of ergative/absolutive alignment (accusative S_A) as well as of an already neutralised alignment (accusative A). On the other hand, personal pronouns (e.g. *ego* in section 2.5.1.), which were excluded from the study, typically occur as highly agentive A subjects and, therefore, conserve the nominative marking, representing, thus, a nominative/accusative alignment. They can be thought to constitute the 0%-tail of the above-mentioned imaginary accusative percentage continuum. In sum, the morphological alignment is best considered synchronically to be a continuum with different concomitant and overlapping evolutionary stages.

The transitivity component analysis of section 4.3. did not provide a practical measure for predicting the subject case. What was shown was that the fourfold construction type classification is problematic because it combines arbitrarily criteria of semantic and syntactic origin. It was suggested that a tailored transitivity scale for (charter) Latin would be needed where only the relevant features are taken into account. This scale should also better address animacy and the properties of the subject in general. I recognise here a potential object of a future study. The relevance of section 4.3. can be seen in that it has pointed out weaknesses of the available methods and also proposed some solutions to them.

Finally, section 4.4. uncovered the close connection between the genre and the transitivity degree in the Latin of LLCT. This connection has also been demonstrated in certain modern languages. As the charter genre appeared to be rather low in transitivity, it was concluded that, in the end, LLCT is not an optimal material for studying case alignment. The low overall transitivity degree is likely to make the accusative subject distribution far narrower (i.e. percentages between 10 and 50) than it might look were the entire transitivity scale in use. The latter might be the case, for example, in narrative material, which, however, follows the Classical Latin standards even in the Late Latin period (historiography) or does not survive from Italy (hagiographic narratives). In any case, the transitivity of the Latin textual genres will be worth examining in a future corpus study.

5. Syntax and subject case selection in LLCT

This chapter will show that an empirical study of case alignment cannot ignore the syntactic context. After analysing verbal semantics and its influence on the subject case in chapter 5, I shall discuss some seemingly syntactic factors that may interfere with case selection. Even though the case marking of charter Latin appears to be partly semantically-based, some syntactic, or more broadly,

structure-related factors seem to have been involved in the process of eventual case selection of individual subjects or subject groups. Cennamo, for example, considers syntactic and pragmatic features, such as clause type and subjectisation of topics, to have conflated with semantic features in the formation of the semantically-based alignment.[288] This is how factors that are not directly related to semantics may have come into play. It is not obvious, however, in which sense those syntactic factors of LLCT that will be dealt with in the following are related to the semantically-based alignment whose residues were examined in chapter 4.

Section 5.1. examines the structural complexity that derives from the coordination of subjects (a phenomenon sometimes called 'complex subjects'), and section 5.2. is about the attributes that are attached to the subject NP head. Section 5.3. focuses on the subject position within the clause. First, section 5.3.1. surveys in general how the LLCT subjects are located relative to the finite verb. Then, section 5.3.2. examines how the distance between the subject and the verb affects the case assumed by the subject.

Table 5.1. summarises the dependent and independent variables that will be discussed in chapter 5. Each variable is also provided with its categorical levels. The abbreviation 'Hyp.' stands for the hypothesis involving the mentioned variables and 'Ch.' for the chapter in which that hypothesis will be tested, respectively.

Table 5.1. Dependent and independent variables with their categorical levels.

Dependent variables		Independent variables		Hyp.	Ch.
Variable name	Categorical levels	Variable name	Categorical levels		
case form of subject	nominative or accusative	structural category	category (A1, A2–4, B1–2)	H_1^1	5.1.
case form of attribute		syntactic attribute position with respect to its head within the subject NP	distance in dependency levels (pos. integer)	H_1^2	5.2.
		attribute position with respect to its head within the subject NP	distance in word positions (neg. or pos. integer)	H_1^3	
case form of subject		length of subject NP	number of words (pos. integer)	H_1^4	
		subject position with respect to the verb	distance in word positions (neg. or pos. integer)	H_1^5	5.3.2.

---

[288] Cennamo 2009, 327–328, 340–341.

It should be noticed that, in this study, the word position within the subject NP or sentence is treated as a nominal, i.e. a categorical, variable although it would be natural to interpret the distance as an interval variable. (In chapter 4, the transitivity degree variable was also operationalised as categorical.) When an interval variable is categorised, the analysis inevitably loses some of the information available. Nonetheless, the categorisation of the distance variables appears to be an adequate way to find out meaningful correlations in the LLCT data, as will be shown below. Encouraged by the theory-compatible results, I have exploited even in this chapter the easy and illustrative decision tree model. The decision tree technique, based on the chi-square test, categorises continuous and interval variables automatically into statistically significant categories.[289] The algorithm then partitions the data on the basis of these categorical levels. Thus, the most conspicuous interactions of the independent variables can be estimated, although the decision tree is not a real multivariable method.

The fact that the decision tree algorithm groups statistically similar categories automatically is likely to alleviate the defects caused by the categorisation of the distance variable. With the grouping method, it is possible to find out which categories are in statistical terms more connected to each other than to the other categories. Indeed, this study does not view the position that a word can occupy within a linguistic string as a genuinely free choice. This is because distance is not really free to assume an arbitrary position, i.e. any integer between -∞ and +∞, but tends to cluster around certain hotspots instead. These hotspots can be extra-linguistically motivated: in LLCT, they may be related to formulaicity. This issue will be discussed further in relevant places.

As was said above in section 4.2.2., it is not always possible to treat the 3rd declension subcorpus separately in chapter 5 and 6, where the syntactic variables are discussed, because the size of the subcorpus does not necessarily enable chi-square tests. In any case, I shall exploit the subcorpus to support the argumentation whenever it is possible to conduct the chi-square tests, even at the expense of the uniformity of categorisation of the independent variable (Table 5.6.). In some cases where the subcorpus can be fully utilised, I shall use only it alone and not the entire LLCT data (Figure 6.4.). To conduct a more balanced statistical analysis, a larger treebank would be required.

As mentioned earlier, I shall classify the phenomena discussed in this chapter preliminarily under the title 'syntactic'. However, it remains unclear whether these factors are really to be described as

---

[289] The next step would be to utilise logit regression or other multivariable methods that were beyond the scope of the present study.

syntactic within LLCT or whether the extra-linguistic, i.e. psychological or mnemonic, element must sometimes be considered to be more dominant. It can also be asked whether 'syntactic' and 'mnemonic' are ultimately only two names for a single phenomenon. Psycholinguistics has been extensively studying processing time as well as the working and the short-term memory, but this study can only touch on the most relevant findings of the field. It is evident that syntactic complexity is related to the amount of processing difficulty, i.e. the cognitive effort required for producing or parsing an expression. It is, however, not so evident what this syntactic complexity consists of. In any case, processing difficulty has been proved to correlate with differences in the working memory capacity and, consequently, can be measured by way of the latter.[290]

Another difficulty is that psycholinguists have studied much more the comprehension than the production of written texts, let alone the production and comprehension of speech. Moreover, historical written texts differ essentially from texts, spoken or written, that are produced in controlled experimental settings with testees. The extendibility of the results of this kind to the interpretation of the complexity in charter Latin remains doubtful.[291] When writing, the writer is able to design the sentences in advance and, in the case of an early medieval scribe, to make recourse to his mental repository of suitable formulaic phrases. On the other hand, a scribe could not easily cancel or emend what was written or change the sentence structure on the run. He was also continuously forced to interrupt his work because he had to dip the quill in ink or sharpen his nib every now and then.

Nevertheless, it is necessary to formulate a definition of what is understood by syntactic complexity in this study. Connecting complexity with structure length is intuitive and seems to be a worthy choice. Indeed, several memory-based theories suggest that comprehension difficulty reflects dependency lengths, i.e. the distance between words that are dependent on each other, such as a verb and its subject.[292] However, it is not clear how exactly this dependency length ought to be measured. It has long been suggested that processing difficulty, if measured as processing time, is not

---

[290] Jaeger & Tily 2010, 324.
[291] On the difference between the processing in written and spoken texts, see Raumolin-Brunberg 1991, 135; Beaman 1986, 45–46; Ellis & Beattie 1986, 211–227.
[292] Jaeger & Tily 2010, 324. Hawkins (2004, 31–34) examines essentially the same phenomenon from the viewpoint of 'constituent recognition domains', which mean the shortest strings of words within which all the possible children of a syntactic constituent can be identified. As for typology, Hawkins proposes that word orders which minimise constituent recognition domains are processed more efficiently. Vice versa, the less minimised the domain is, e.g. the more words it contains, the slower the processing and the higher the syntactic complexity.

only dependent on the number of words, but on the number of constituents of the sentence or on the number of the propositions, i.e. chunks of (new) information, that the text contains.[293]

The theoretical importance of constituents or propositions is mainly related to the fact that the closed-class grammatical words, such as prepositions and conjunctions that bind the propositions together, are easier to process than the open-class items that introduce the propositions into a text. Moreover, superordinate propositions, such as the main-clause predicate and its arguments, are likely to be more easily processed than subordinate propositions, which merely modify the superordinate elements.[294] To summarise, comprehension difficulty is conceptually not only related to the plain number of words, but also to what these words are comprised of. Nonetheless, Szmrecsányi (2004) has shown that measures, such as intervening words, syntactic nodes or phrases, and new discourse referents, are all highly correlated.[295]

In this chapter, I shall adopt the view that comprehension difficulty is essentially memory-related and reflects dependency lengths. Moreover, I shall utilise simple word counts as the measure of dependency length. For this study, it is enough that there is the above-mentioned correlation between the word count and the more sophisticated measures. I also assume that the role of the linear length of the word string may be more accentuated in the highly formal and repetitive charter language than with other texts. Whatever the optimal measure of dependency length is, in corpus linguistics, where vast masses of words are at play, it can be expected that the differences between various measures are increasingly neutralised. The following generalisation is likely to apply: the longer the word chain is, the more nodes, phrases, and chunks it will have and, on average, the more syntactic complexity it will contain.

This study is, however, not only about the syntactic complexity of word string length, but also about the syntactic complexity of coordination. In section 5.1., the dependency length between the head and the attribute of the subject NP will be measured both as word string length and as dependency levels. As for coordinated subjects, it seems to be generally accepted that coordination as such induces rather mild syntactic complexity in comparison to, for example, subordination that has been

---

[293] Kintsch 1977, 358–360; Singer 1990, 143–144. There are also differences between the processing times of open-class and closed-class words, the latter being more rapidly processed (e.g. Gordon & Caramazza 1985, 96–98).
[294] Hawkins 1983, 98–104; Kintsch 1977, 360–362. See Raumolin-Brunberg (1991, 132–135) for a good synthesis of these problems in historical linguistics.
[295] Szmrecsányi 2004, 1031–1038; Jaeger & Tily 2010, 324; Kälviäinen 2013, 177–179.

extensively studied.[296] I emphasise, however, that coordination is not only about equaling or listing items but can have complex hierarchising functions in classifying propositions, as will be shown in section 5.1. In this study, I shall consider coordination a part of structural complexity although its role is multifold. On this occasion, it is also to be noted that the subjects of main clauses and subordinate clauses will be treated alike in this study.

Most of the conclusions that will be drawn in this chapter are based on the premise of the accusative as the default case in the Latin of LLCT. In short, a subject that occurs in a non-prototypical environment is inclined to slip into the default case which is assumed to be the accusative at the time of LLCT. In earlier Latin, the nominative is likely to have been the unmarked default case, and the markedness turn to an unmarked default accusative may have taken place sometime in Late Latin. The prototypicality of the subject environment is also essentially related to the dependency length between the subject and the verb. The question of the default case will be touched upon several times on the following pages. I refer here especially to section 6.2., where the default case will be discussed in more detail.

5.1. Subject case and complex coordinated structures

This section and the following two sections will refer to 'structural categories', in which all the subjects of LLCT appear organised according to their dependency path, i.e. the coordinating nodes that link the subjects to their verbal heads (Figure 5.1. and Figure 5.2.). Thus, in this study the structural categories are always about coordination. At the same time, the structural categories are technical units exploited by the treebanking and query procedures.[297]

Rich use of coordination (by way of conjunctions or asyndeton) is characteristic of charter Latin and legal language in general. In LLCT, many entities appear as lists: names of rogators and respondents, names of witnesses, items sold or bought, descriptions of those items, various conditions, and so on. Beyond the exigencies of practical issues, listing seems to be an inherent feature of

---

[296] Griffin & Crew 2012, 420–421; see also Beaman 1984, 59–61. Lehto (2013, 254) notes that frequent coordination does not necessarily increase complexity if it is systematic and intended to make the contents more intelligible. This is certainly true but does not mean that coordination would not increase the processing effort required for decoding the syntactic structure. Besides, in charter Latin, the coordination rarely renders the contents more intelligible.

[297] The PML-TQ query language recognises syntactic relations as combinations of head/dependent relations and pertinent syntactic function tags (see section 2.4.).

the legal genre, as is seen in a typical charter formula (1).[298] Example (2) contains 7 coordinated structures, many of which are redundant as far as the conveyed information is concerned. The structural categories can be utilised to define the dependency paths between any nodes: between verbs and their subjects and objects in example (2) or between verbs and their objects in examples (7) and (8). In this study, obviously, the relation between the verb and its subject(s) receives the most attention.

(1) CDL 178 (AD 764) *in qua eclesia pro fascinora do, duno, trado donatoque esse uolo*
'I give, donate, concede and want [it] to be donated to that church for my sins'

(2) CDL 139 (AD 759) *similiter promitto ego Peredeus in Dei n(omine) episcopus uobis Gumfrid et Baruncio germani de casa seo et res illa quem uobis ad resedendo et laborando dedi in loco Saltuclo, casa cum curte, orto, uineis, terris, siluis, oliuitis, qualiter Baruccio ad manum suam habuit, uobis dedi ad resedendo et laborando, in tali tenure, ut per omne annum mihi reddere debeatis uno sold(o) bono expendibile et medietate uino et angaria*
'in a similar vein, I, Peredeus, bishop in the name of God, promise you Gumfid and Baruncio, brothers, out of that house and land that I gave you to be resided and cultivated in Saltuclo, the house with court, garden, vineyards, fields, woods, olive groves, just as Baruccio had them in his possession; I gave [it] you to be resided and cultivated in such terms that you must pay me every year one fine *solidus* of full weight and half of the vine and the corvée'

In theory, the number of possible structural coordination categories is infinite but, in practice, the six most common categories A1, A2, A3, A4, B1, and B2 correspond to 99.25% of all the subjects of LLCT. The 12 subjects that constitute the remaining 0.75% have been omitted for technical reasons.[299] The numbers of the basic query subset subjects for each mentioned structural category can be seen below in Table 5.2. It is to be noticed here that these categories are not about how many mutually coordinated subjects or verbs there are on the same level of coordination structure, but

---

[298] Coordination in terms of listing has been typical of legal Latin from very early on. Clackson & Horrocks (2007, 107) consider that repetition and the exhaustive enumeration of options in the early Latin law texts derived originally from religious language. It enhanced memorisation but also conveyed the seriousness of the business by distancing the language from everyday usage. For the language of ancient Roman laws, see also Poccetti & al. 1999, 197–204. Lehto (2013, 244–253) has studied coordination in Early Modern English proclamations from the viewpoint of syntactic complexity.

[299] These 12 stray subjects belong to rare and complicated structural categories. I have omitted them in order not to strain the treebank queries unreasonably. The 12 subjects include 1 B4-type subject (a coordinated subject that is attached to coordinated verbs through two levels of coordinators) and 11 C1-type subjects (subjects of several types of elliptic structures that are difficult to deal with in PML-TQ query language).

how many levels of coordination there are between the subject and the verb, i.e. the length of the coordination path. In sentence (4) below, more than 20 subjects are coordinated with each other, and yet they form a simple A2 coordination.

The structural categories that are utilised in this study are described in terms of dependency relations as follows:

- Category A1 represents a non-coordinated subject attached directly to its non-coordinated verbal head. This is the simplest subject/verb relation.
- Categories A2–4 contain coordinated subjects that are linked to their verbal head through one (A2), two (A3), or three (A4) subsequent coordination nodes (conjunctions or commas).
- Categories B1 and B2 contain non-coordinated (B1) or coordinated (B2) subjects that are linked to two or more mutually coordinated verbal heads through one or two coordination nodes.

Figure 5.1. Tree structures of categories A1, A2, A3, and A4.
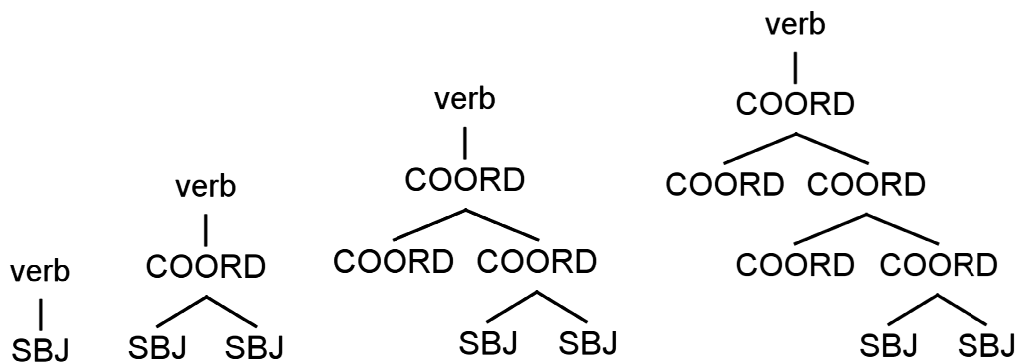
Figure 5.2. Tree structures of categories B1 and B2.
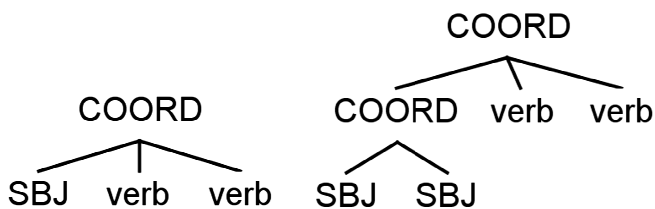
For category A1, the simplest example is MED 539 (AD 838) *Perio dixit* 'Perio said'. Instances of other structural categories abound below among the example sentences of this section: e.g. (4) belongs to A2, (7) contains an A3 coordination, (5) belongs to B1, and (6) to B2. It should be remembered that I define as subjects only the head of the subject NP, i.e. the noun (or substantival adjec-

tive) that ranks the highest on the animacy hierarchy (see section 4.1.1.). For example, in sentence (3) only *clerici*, the head of the subject NP is defined as the subject, as can be seen in Figure 5.3.

(3) MED 539 (AD 838) *isti omnes prefati clerici* [--] *unianimiter testati sunt*

'all these aforesaid clerics [--] witnessed unanimously'

Figure 5.3. Tree structure of sentence (3).
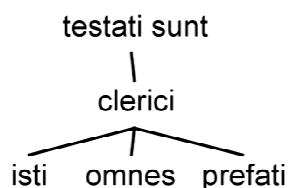
This NP-head-only definition is due to the fact that, in charter Latin, the eventual attributes can be in a case form that differs from that of the NP head: the subject NP cannot be dealt with as one entity. I shall discuss the internal variation of the subject NP and the factors involved therein in section 5.2. Moreover, coordinated subjects of categories A2–4 and B2 are counted in the statistics as independent subjects and not, for example, as one complex subject NP, as is sometimes done in other studies.[300] This would be impossible since the subjects that are coordinated with each other not infrequently apply different case forms. For example, in category A2, 12% of the coordinated subject combinations include subjects in two different case forms. Due to the above-explained principles, the numerical data extracted from LLCT are not completely comparable with those of other studies on Latin. On the other hand, most studies of Latin linguistics do not even provide numerical data.

Here, as everywhere in this study, a single subject/verb combination is understood by the term 'clause'. As for categories A2–4 and B2 with several coordinated subjects, the same sentence is counted as many times as there are different subjects attached to its verbal head. It is also to be noticed that the ramification of the subject branch is far more common than the ramification of the verb branch. There are only 118 sentences involving coordinated verbs (B1–2) in LLCT against the 479 sentences involving non-coordinated verbs (A2–4). Although the coordination of verbs differs technically from the coordination of nouns, both are dealt with as parallel phenomena in this study.[301] As said, the coordinated subjects are all taken into account: for example, in sentence (4), all the 25 coordinated subjects are included in the numerical analysis. As for coordinated verbs,

---

[300] E.g. Raumolin-Brunberg 1991, 115.
[301] Torrego (2009) is a comprehensive descriptive study on Latin coordination. See also Pinkster 1990, 257–258.

only the verb nearest to the subject is counted. In the B1-type sentence (5), only verb *suscripsi* 'I subscribed' is acknowledged, while *vidi* 'I saw' is left out. When the distance between the subject and the verb is measured in section 5.3., only the positional interval between *Istefanacis* and *suscripsi* is counted.[302]

Occasionally, two or more coordinated verbs that have one or more subjects in common differ semantically from each other, as is the case with the B2-type sentence (6). *Residere* 'to reside' (*debeamus*) is an inactive intransitive verb, while the other verbs, *gubernare et tegere seo meliorare* 'to govern', 'to roof', 'to ameliorate' (*debeamus*), are transitive. Blends of this kind are infrequent and are necessarily excluded from the study of semantic conditioning of subject case selection that insists on one unambiguously classifiable semantic value for each subject/verb combination. The coordinated verbs that I include in the study must be essentially of the same semantic value, as the transitive *suscripsi* and *uidi* in (5).

(4) CDT 45 (AD 812) *ubi aderant nobiscum Thomas diaconus*, *Otus presbyter*, *Chunifrid presbiter*, *Roschisi presbiter*, *Iohannis clericus*, *Hildipertus subdiaconus*, *Grimpo clericus*, *Sanitas notarius*, *Rachipertu clericus*, *Petro clericus*, *Baucherat vassus domni regi*, *Willardo gastaldus*, *Istabili marepas*, *Gumpertus filius quondam Ursi*, *Teodingo*, *Pertualdo*, *Amulo*, *Allo*, *Gumprando et Walprando germanis*, *Grauso clericus*, *Suliprando*, *Nandifrid*, *Lautpertu et alios plures*
'where were present with us Thomas, the deacon, Otus, the priest, [--] Lautpertu, and several other persons'

(5) CDL 56 (AD 736) *Istefanacis uir clarissimus* [--] <u>*suscripsi*</u> *et eum signum facientem* <u>*uidi*</u>
'(I), Istefanacis, *vir clarissimus*, <u>signed</u> and <u>saw</u> him make his sign'

(6) MED 575 (AD 841) *tali tinore ut nos aut homines nostros in ipsa casa* <u>*residere*</u> *debeamus et tam ipsa casa quam et ipsa curte seo ipso fenile bene* <u>*gubernare*</u> *et* <u>*tegere*</u> *seo* <u>*meliorare*</u> *debeamus*
'in such terms that we or our dependants must <u>dwell</u> in that house and <u>govern</u> well and <u>roof</u> and <u>ameliorate</u> that house and that court and that barn'

---

[302] This rule also applies to the (few) postverbal subjects of coordinated verbs: the positional interval is counted between the last verb and the subject.

According to Torrego (2009), groups of coordinators stay cross-linguistically at the same hierarchical level if they form a repetitive pattern (for example, *et – et*). In contrast, if a combination of different coordinators occurs in the same context, the coordination pattern extends over various hierarchical levels.[303] Nevertheless, in charter Latin, hierarchical differences are commonplace even between NPs coordinated by the same coordinators, as in example (7). On the other hand, different coordinators do not necessarily imply different hierarchical levels, as sentence (8) reveals.[304]

(7) CDL 166 (AD 762) *per omnem annum iustitia ipsei case reddere debeam porco uno ualente tremisse uno et uno pullo et quinque ouas et camisia una ualente tremisse uno et uno animale in mense magio ualente tremisse uno, uinum et labore secundum consuetudinem ipsei case et angaria secundum consuetudinem de ipsa casa*

'I have to settle every year as the rent of that house one pig worth a *tremissis* and one hen and five eggs and one shirt worth a *tremissis* and in May one sheep [?] worth a *tremissis*, vine and corn according to the convention of the house and the corvée according to the convention of the house'

(8) CDL 178 (AD 764) *in primis fondamentum infra ciuitate, ubi ipse eclesia beati sancti Arcangeli Mihahelis fondata est, cum curte et puteum, cum granario et ipsa sala, comodo ipse istaffili positi sunt, seo et orto, comodo sepis circumdatu fuerit, seo et casa Rachuli in Sexto una cum ipso Rachulo et casa in Uersilia, qui regitur per Sirola massario homine liuero, portionem meam in integrum, et casa Magnipertuli de Asilacto una cum ipso Magnipertulo et quarta portione de sala et de granario seo et finile*

'firstly, that plot outside the city where the church of Saint Archangel Michael stands with its court and well, with the store room and the hall, as the boundary markers demarcate it; and the garden, as it is surrounded by the fence; and the house of Rachulus in Sexto with Rachulus himself and the house in Versilia that is governed by Sirola, the leaseholder, free man, my share [of it?] in its entirety; and the house of Magnipertulus of Asilacto with Magnipertulus himself and one-fourth of the hall and the store room and the barn'

In example (7), the objects *vinum* 'vine' and *labore* 'corn' have been tied together more tightly than the other coordinated items because the modifier *secundum consuetudinem ipsei case*, "according to the convention of the house", modifies only these two words. Even *angaria* 'corvée' seems to be

---

[303] Torrego 2009, 475–476; Dik 1968, 45–47.

[304] It can be questioned whether *et* and *seo et* can be viewed as different coordinators.

slightly disconnected from the preceding members of the coordination because of its modifier. Figure 5.4. presents one possible interpretation of the coordination structure of example (7). *Porco*, *pullo*, etc. are coordinated with each other as an A2 coordination, while *vinum* and *labore* form an A3 coordination. In fact, *pullo et quinque ovas* 'a hen and five eggs' may also form a more tightly connected phrase, in which case they ought to be interpreted as an A3 coordination. The whole syntax tree that example (7) is part of can be seen in Appendix 2.2.

Figure 5.4. Tree structure of example (7).
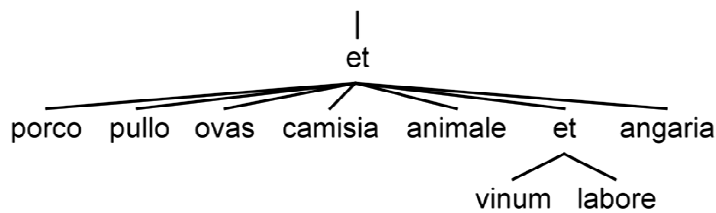
In a similar manner, at the end part of example (4), the cluster *Gumprando et Walprando* has been taken out from the lengthy A2 coordination by the attribute *germanis* 'brothers'. The attribute ties the two names with each other, thus raising them to an A3-type coordination within an A2 coordination. On the other hand, even *Allo* and *Amulo* may be brothers of *Gumprando* and *Walprando*. This is an example of how arbitrary the interpretation sometimes is. While example (4) is about subjects, in (7) and (8) objects are coordinated. In (8), each coordinated item is defined by a relative clause or by a prepositional phrase. Here, the coordinators are different, *seo et* and *et*, but the coordination is of the simple A2 type. The coordination includes seven coordinated items and an interjective apposition, *portionem meam in integrum* "my share in its entirety", which comments on the preceding item *casa in Versilia* "house in Versilia".

I argue that, just as coordination in itself increases the complexity of the clause, strengthening the union of two or more items within a coordinating structure further augments its syntactic complexity. Examples of these unions that "deepen" the coordinating dependency structure are the abovementioned *Gumprando et Walprando* in (4) or *vinum et labore* in (7). In this kind of coordination, it is always about an unequal grouping of coordinated items inside a coordinating structure. The *vinum et labore* node in Figure 5.4. clearly complicates the grouping of the coordinated items, thus contributing to the overall complexity of the structure. These deeper-level subunits are likely to require more time not only when hearing or reading the sentence, but also when compiling it: the writer is more prone to slip into text-organisational disorder, which may be expressed as discontinuous structures or, for example, as the use of "wrong" grammatical forms. Eventually, this study

does not focus so much on the coordination itself but on how the coordinated nodes group into clusters that are located at different levels of the sentence-internal hierarchy according to their relative distance to the verbal head.

I shall now proceed to the numerical analysis. As will be argued in more detail in section 5.2., it can be postulated that

H_1^1: the more complex the structural category is, the more often is the subject expected to occur in the accusative.

The corresponding null hypothesis is as follows:

H_0^1: The structural category has no significant effect on the case form of the subject.

Hypothesis H_1^1 is based on the assumption that, in ambiguous environments, the default-case accusative often slips in and replaces the nominative subject. Because the deepness of coordination is interpreted as syntactic complexity, there is expected to be a statistically significant connection between the case form of the subject and the structural category the subject is in. Table 5.2. presents the percentages of accusative subjects in each structural category of LLCT.

Table 5.2. Number of subjects and accusative subject percentage across structural categories in LLCT.

	Nominative		Accusative		N
A1	660	67.3%	321	**32.7%**	981
A2–4	284	59.3%	195	**40.7%**	479
B1–2	94	80%	24	**20%**	118
$\sum$	1,038	65.8%	540	34.2%	1,578
Chi-square	$\chi2 = 20.04, df = 2, p < 0.001$				

Table 5.2. shows that the increase of the accusative percentage from category A1 to A2–4 conforms to the above-postulated direction of growth but the percentage of category B1–2 does not. The accusative percentage of category B1–2 is lower than that of both the other categories although category B1–2, the most complex of all, would be expected to have the highest accusative percentage of the three. This unexpected pattern is explained, however, by the fact that the relatively small category B1–2 (118 occurrences) is not only complex in structure, but also extremely formulaic. The cate-

gory seems to be too formulaic to reflect real linguistic tendencies.[305] This becomes apparent when category B1 (non-coordinated subject) and B2 (coordinated subject) are scrutinised separately. In B2, only 13% of the subjects are in the accusative form because the category is heavily occupied by occurrences of (crystallised) *res* 'property' and (animate) *filii*(*s*) 'sons', *posteri*(*s*) 'descendants', and *homo* 'man' (23 in 60) that are invariably in the nominative. Sentences (9), (10), and (11) exemplify these typical formulaic expressions.

(9) MED 414 (AD 818) *si nos vel ille <u>homo</u>, cui nos eam dedissemus aut dederimus, vobis eam intentionaverimus aut retraxerimus per quolibet ingenium*
"if we or the <u>man</u> to whom we shall have given or shall give it will contest or dispossess it for whichever reason"

(10) MED 447 (AD 822) *predicta casa et <u>res</u> in mea sint et permaneant potestatem*
"[that] the aforesaid house and <u>land</u> be and remain in my possession"

(11) CDT 88 (AD 813) *si nos ipsi vel nostri <u>filiis</u> aut eredis contraverimus aut ab omne homine non putuerimus defendere*
"if we or our <u>sons</u> or heirs contest [it] or cannot defend [it] from whosoever"

The same words are also found in category B1 but to a lesser degree. Owing to the high formulaicity rate of the most complex structural categories, the classification into categories seems to be somewhat controversial as far as the subject case is concerned. Nevertheless, the dependence between case and coordination category remains statistically significant in the two large categories A1 and A2–4 when B1–2 is left out: $\chi2 = 8.99$, $df = 1$, $p = 0.003$.

In section 5.1., it has been shown that coordination may be an instance of syntactic complexity. It has also been shown that the degree of coordination, i.e. the independent variable whose effect on the dependent variable was tested, affects subject case selection to some extent in the manner predicted by hypothesis H_1^1. This is, however, not without exception since the accusative percentage does not rise anymore in the most complex structural category. Based on the pattern between categories A1 and A2–4, it can be suggested that syntactically complex structures favour realisation of accusative, the by then default case of Latin, as the subject case form. This suggestion still requires

---

[305] See also Table 5.3. (section 5.2.) and the related discussion on the relationship between structural category and relative subject position.

further proof, which cannot be provided by the LLCT data. On the other hand, section 6.1. will show that structural category is closely interrelated with construction type, which, as is known, affects subject case selection. This may explain the fact that subject case selection is in part dependent on structural category.

5.2. Complexity of the subject NP

The previous section studied the dependence between coordination and subject case selection. This section discusses three other dependences that may be relevant to the case marking of the subject NP. The dependent and independent variables can be seen in the following table, which reprints part of Table 5.1.

Dependent variables		Independent variables		Hyp.
Variable name	Categorical levels	Variable name	Categorical levels	
case form of attribute	nominative or accusative	syntactic attribute position with respect to its head within the subject NP	distance in dependency levels (pos. integer)	H_1^2
		linear attribute position with respect to its head within the subject NP	distance in word positions (neg. or pos. integer)	H_1^3
case form of subject		length of subject NP	number of words (pos. integer)	H_1^4

First, I shall examine the dependence between the case form of the subject NP head and the case form of its noun/adjective attributes. This will be realised by measuring their distance in dependency levels, i.e. the number of edges that separate the attribute from the head.[306] Second, I shall investigate how the linear distance between the subject NP head and its attributes affects the case selection of the attribute. Finally, I shall study whether the total length of the subject NP correlates with the subject case selection. It is important to notice that by attributes I refer to all those adjective modifiers, determiners (demonstratives and possessives), and appositive noun modifiers that agree with their head by number and case, with the adjectives and determiners agreeing even by gender. This definition excludes genitive attributes.[307]

---

[306] The links between the nodes of a dependency tree are called edges.
[307] A more detailed classification of attributes would have been possible but unnecessary in this kind of study. An accurate study of the NP can be found in Spevak 2014 and Spevak 2010, 223–265.

In this chapter, I examine mainly NPs that involve subjects of category A1. The reason for this is twofold: first, the A1 subjects are most likely unmarked (as will be argued in section 5.3.1.) and, therefore, drawing on this category may exclude some linguistic and extra-linguistic sources of error. The other reason is of economic nature: the attributes are attached to their heads in nearly an infinite number of ways, and the analysis would become highly challenging were the queries run even for the attributes of the A2–4 and B1–2 subjects.[308]

I now turn to analysing quantitatively how the case form of the subject NP head and its attributes depend on each other. I shall test the following hypothesis:

H_1^2 : The greater the dependency level distance between the subject NP head and its attribute is, the more often is the attribute expected to occur in the accusative.

The corresponding null hypothesis is:

H_0^2 : The dependency level distance has no significant effect on the case form of the attribute.

In charter Latin, whichever NP can have a considerable number of attributive modifiers that can be related to each other in various different ways: one or more attributes may be directly linked to the head or they may be chained, i.e. modifying the preceding attribute. By chained attributes I mean attributes that are hierarchically dependent on each other so that the lower-level attribute directly modifies the higher-level attribute that again directly modifies a higher-level attribute or alternatively the head. For example, in (12) *Iacobo*, *humilis*, and *episcopus* all modify the head, *ego*. *Iacobo* modifies *ego* directly, whereas *humilis* and *episcopus* modifies it indirectly. Yet they all (should) agree in case with their head. Figure 5.5. presents the dependency structure of (12).

(12) *ego Iacobo humilis episcopus*
"I, Iacobo, humble bishop"

---

[308] In this study the following attribute categories of A1 subjects are included: A1.1 (non-coordinated attributes attached directly to the head), A1.1.1 (non-coordinated second-grade attributes of attributes attached directly to the head), A1.1.1.1 (non-coordinated third-grade attributes of second-grade attributes of attributes attached directly to the head), A1.2 (coordinated attributes attached to the head), A1.2.1 (non-coordinated second-grade attributes of attributes attached to the head) and A1.1.2 (coordinated second-grade attributes of non-coordinated attributes attached directly to the head). They cover almost all the attributes of the A1 subjects.

Figure 5.5. Dependency tree of example (12).

```
    ego
     |
   Iacobo
     |
 episcopus
     |
   humilis
```
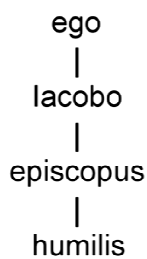
Note that the NP of example (12) is headed by a pronoun, *ego*, whereas hypothesis H_1^2 is tested only on NPs that have a noun (or substantival adjective) head. Hypothesis H_1^3 (see below), instead, will be tested only on NPs that have a pronoun head. What is essential for this study is that these attributes are quite often in a case form that differs from that of their head. The attributes that modify the same head can also be in a different case form from each other. I underline that I will now examine the dependency relations within the syntactic structure, not the linear word order. The latter will be discussed later in this chapter.

Table 5.3. presents the case form of the noun/adjective subject NP head and its attributes. These are significantly fewer than those with a pronoun head (mainly *ego* or *nos*). Head signifies the head of the subject NP, the first-grade attribute is the first attribute of the chain, directly attached to the head, and the second-grade attribute is the second attribute of the chain, attached to the first attribute. In the case of coordinated attributes, the grade number of the attribute is counted leaving aside the coordinators. For example, in Figure 5.6., the attributes *sanctus* and *venerabilis* are treated as first-grade attributes, i.e. equally to the directly attached attributes, such as *Iacobo* in Figure 5.5.

Figure 5.6. Dependency tree of *sanctus et venerabilis locus* "holy and venerable place".

```
      locus
        |
       et
      /    \
sanctus   venerabilis
```

191
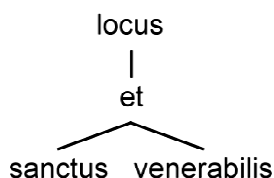
Table 5.3. Case distribution of the subject (noun or adjective) and its attributes in LLCT.

Position	Accusative		Nominative		Σ
	N	%	N	%	
head	101	32.0%	215	68.0%	316
1st-grade attribute	114	36.1%	202	63.9%	316
2nd-grade attribute	20	c.70%	10	c.30%	30
Chi-square	$\chi^2 = 14.50, df = 2, p < 0.001$				

Table 5.3. shows that the first-grade attribute is more often in the accusative (36.1%) than the subject head itself (32.0%), while the leap between the first-grade and second-grade attributes is more substantial (about 30 percentage points). The second-grade attributes are, however, few with respect to other classes. The dependence between attribute position and case is statistically significant ($\chi^2 = 14.50$, $df = 2$, $p < 0.001$). The percentages seem to show that the further off the head the attribute is located, the higher is its accusative percentage. This is well in accordance with the rather logical assumption that the length of the subject NP contributes to its syntactic complexity: the larger the syntactic distance between the head and the attribute is, i.e. the longer the attribute chain is, the more easily the case form of the attribute slips into the accusative, the default form (see section 6.2.). A parallel interpretation was proposed for coordination in the preceding section.

This said, subject NPs with noun/adjective head that present only one case form (nominative or accusative) comprise 272 out of 316 (86.1%), which nonetheless proves a relatively good consistency in case form usage across the subject NP. In total, 32% of these same-case chains are in the accusative form. It is also important to examine whether there is a difference in consistency between the accusative-head and nominative-head chains. It turns out that both the accusative-head and nominative-head chains change the case form in 14% of the cases (the accusative-head in 14 cases out of 101 and the nominative-head in 30 cases out of 215).

I examine here only subject NPs with noun or adjective head, as defined in the basic query subset. If subject NPs with a pronominal head (mainly highly referential *ego* and *nos*) were taken into account, the results would become distorted because c.99% of all the pronominal subjects are in the nominative form. Additionally, when the head is pronominal, the second position is usually occupied by 2nd declension personal names, which are famously often in the accusative (cf. example (12) *ego Iacobo humilis episcopus*) (see section 4.2.3.). With these names, the accusative percentage would rise disproportionately high in the second position of the subject NP. Indeed, only 240 (35.6%) in 674 subject NPs with a pronominal head present only one case form (nominative or ac-

cusative). This is considerably less than with subject NPs with noun or adjective heads (86.1%): the 2nd declension personal names in *-o* cut off the chain that has begun in the nominative form.

The previous section examined the dependency relations inside the subject NP. Next, I shall have a glance at the linear distance between the syntactic head of the subject NP and its attributes: the following paragraphs investigate whether the linear position that the attribute occupies within the subject NP affects the case selection of that attribute. This is done in order to find out whether the subject case selection inside the subject NP, in general, is influenced more by the internal syntactic structure of the phrase (dependency relations) or by the word order, i.e. the phrase-internal linear distance between the head and the attribute. This is worth studying because the linear modifier order of the subject NP does not necessarily match the order in which the words are attached to each other in the dependency structure. As is well known, Latin word order enables variation within the NP, as modifiers can precede or follow the head. This study does not claim anything general about the modifier order of (charter) Latin, but concentrates exclusively on the case forms of the attributes.[309] It has to be kept in mind that the formulae seem to be responsible for several unexpected modifier constructions, which sets an extra challenge for this study.

I address here only category A1 for the same reasons as above, where the dependency order was under examination. This time, however, only the subject NPs with pronominal head are examined. This is possible, and even necessary, because the (predominantly nominative) pronouns do not skew the statistics, as only the case forms of the attributes are analysed. Discussing subject NPs with a pronominal head on a par with NPs with a noun/adjective head would be harmful because noun heads often have preposed modifiers while pronominal heads prefer almost exclusively postposed modifiers. It is of no use to align two categories with conflicting characteristics.

As said, this study takes into account the attribute categories listed in note 304. The case form of each attribute of each A1-type subject of LLCT is examined independently, regardless of the case form of the head or the other possible attributes of the phrase. The distance of the attribute from the syntactic head is measured as word counts. Here, as everywhere hereafter, by linear position I mean the position a word occupies inside an NP or a sentence. For example, in example (12), reprinted here as (13), *ego* occupies position 0, *Iacobo* (+)1, etc. A negative integer would imply that the at-

---

[309] The head/modifier order is amply studied from a typological viewpoint. For Latin, see Spevak 2010, 223–265; Adams 1976a, 88–90; Bauer 2009, 256–265. Obviously, the case form of the (subject) NP modifier has not been studied earlier because the question is pointless for standard Latin, where no variation in the case form is allowed within an NP. LLCT is one of the few corpora where this type of study is possible.

tribute precedes its head, but there is only one instance of that with the pronoun-head subject NPs in LLCT. Thus, the distance from subject *ego* to attribute *episcopus* is 3 word positions. It is worth noting that the dependency distance of the same words, *ego* and *episcopus*, would be 2 (hierarchy levels), however, as the adjective *humilis* depends on its syntactic head *episcopus* in the dependency hierarchy (see Figure 5.5.).

(13) *ego Iacobo humilis episcopus*
"I, Iacobo, humble bishop"

The hypothesis to be tested here is presented in H_1^2 and the corresponding null hypothesis in H_0^2.

H_1^2: The greater the linear distance between the subject NP head and its attribute, the more often is the attribute expected to occur in the accusative.

H_0^2: The linear distance has no significant effect on the case form of the attribute.

Figure 5.7. presents the accusative percentage of the attributes of A1 subject NPs as a function of distance from the NP head. The trend line drawn on the graph seems to be slightly ascending. This may imply that the further off the head the attribute occurs, the higher is its accusative percentage, so the above null hypothesis can be confuted. The result looks parallel to that received above with the previous variable (see Table 5.3.). On the other hand, the frequencies are very small after value 4 even though the relation of case and attribute position is statistically significant on the entire range. Values 9 and 10 (10 occurrences) were combined with value 8 to make the chi-square test possible.

Figure 5.7. Accusative percentage and frequency of the attributes of A1 subject NPs as a function of linear distance from NP head ($\chi^2 = 31.92$, $df = 7$, $p < 0.001$).
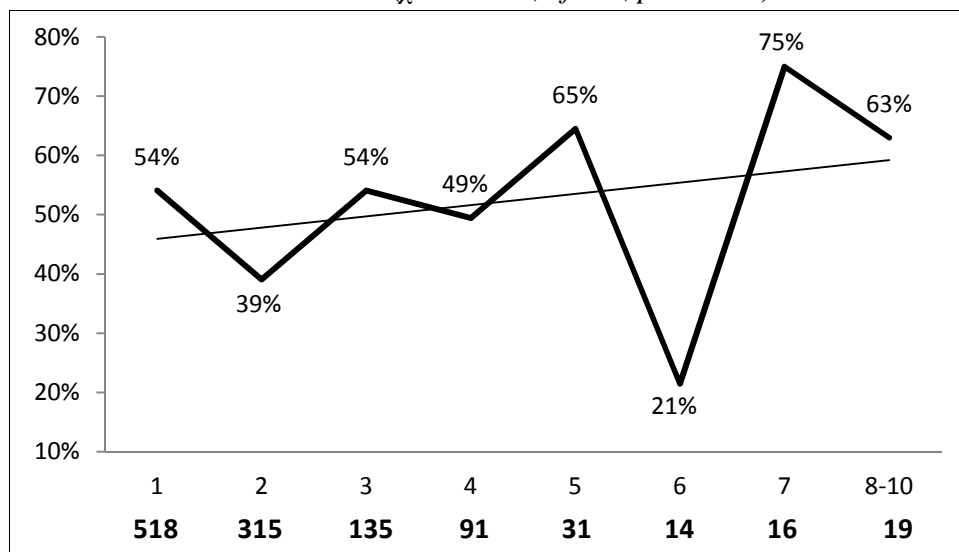
In spite of the statistical significance, the percentages in Figure 5.7. do not form as clear a pattern as might be postulated on the basis of hypothesis H_1^3. It is not possible to explain the radical dive at value 6; it is probably due to simple random variation. The adjusted standardised residuals, which can be found with the cross-tabulation in Appendix 5.1., do not help in interpreting the oscillation of the curve. In sum, one has to be cautious about the numbers in Figure 5.7. However, as the trend that seems to occur there is in line with that of the syntactic position variable (Table 5.3.), it can be accepted until the contrary is proved.

The first two of the three independent variables discussed in this chapter, i.e. the syntactic and linear distance between the subject head and its attribute, seem to correlate positively with the dependent variable, i.e. the case form of the subject NP attribute. However, these two independent variables are not likely to be independent *of each other*. On the contrary, they are very likely to interplay heavily because the word order usually follows more or less straightforwardly the syntactic dependency structure. As the trend seems to be more apparent with the syntactic distance (Table 5.3.) than with the linear distance (Figure 5.7.), it can be legitimate to conclude that the syntactic dependency variable (cf. dependency length above) is the primary *explanans*, which the word order variable imperfectly reflects. These estimations must, however, be regarded with some reservation until further proof is provided. What all this may signify regarding the accusative case and its having become the default case will be discussed in more detail below in section 6.2.

The last task concerning the subject NP is to examine whether the total length of the subject NP affects subject case selection. As was stated above, the NP length is probably related to syntactic complexity because processing long chains of words usually (but not always) takes more time than processing only a few words. Indeed, it has been demonstrated for written historical English that the NP length correlates with certain syntactic and/or pragmatic phenomena, such as the position of the NP within the sentence.[310]

I have measured the size of each subject NP by counting the number of its syntactic descendants, i.e. the words attached to each subject as modifiers (adjective attributes, appositional attributes, genitive attributes, and relative clauses). This time, all the categories of the basic query subset (A1, A2–4, B1–2) are included. The most frequent cases are subjects with one modifier and subjects with no modifier at all (see the frequencies at the bottom of Figure 5.8.). All the frequencies between 0 and 10 can be seen in the histogram of Appendix 5.2.

The influence of the subject NP length on the case form of the subject NP is tested with hypothesis H_1^4. The corresponding null hypothesis is H_0^4.

H_1^4: The longer the subject NP is, the more often the subject NP head is expected to occur in the accusative.

H_0^4: The length of the subject NP has no significant effect on the case form of the subject NP head.

Figure 5.8. shows the accusative percentage of the subject NP head as a function of the subject NP length. What is seen here is that the accusative percentage hovers around 35% notwithstanding the number of descendants. Although the trend seems to be slightly descending because of the 28% of the values ≥ 8, the chi-square test does not reveal a statistically significant dependence between the case and descendant number variables within any range. Thus, the null hypothesis cannot be rejected. It is to be noticed that value ≥ 8 includes 68 sporadic occurrences ranging from between 8 and 35. These values had to be combined to enable the chi-square test.

---

[310] For written texts of Early Modern English, see Raumolin-Brunberg 1991, 120–121, 135–140.

Figure 5.8. Accusative percentage and frequency of subject NP head as a function of subject NP length in LLCT ($\chi^2 = 7.05$, $df = 8$, $p = 0.531$).

	0	1	2	3	4	5	6	7	≥ 8
%	38%	31%	37%	33%	35%	36%	31%	36%	28%
n	**471**	**549**	**234**	**105**	**62**	**22**	**29**	**39**	**68**
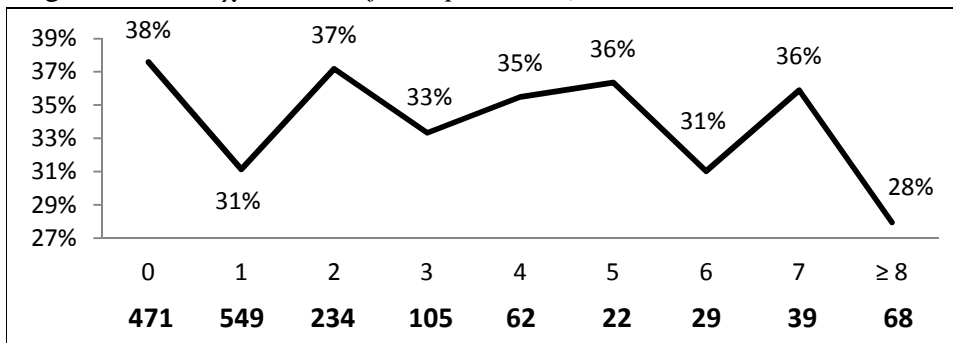
Thus, there is no statistically significant dependence between the total 'mass' of the subject NP and the case form of its head. This, however, is not necessarily surprising because the mentioned modifier masses can be located either between the subject NP head and the verb of the clause or outside this crucial subject/verb bond. When the word masses are located between the verbal head and its subject dependent, the dependency length, which was discussed at the beginning of chapter 5, increases. Thus, even the processing time is expected to extend. In this sense, the result of this last test may be seen as indirect support for the view that dependency length is, indeed, a more important component of syntactic complexity than the sheer number of words. It would be possible to study this in more detail by combining the relative position to the current study. This would require, however, a complicated multivariable analysis, which is outside the scope of this study. Other factors, too, should be taken into account: the length of the NP is often related even to its absolute position within the sentence, with the longest focus-like NPs towards the end of the sentence.[311]

To sum up, this chapter has detected a statistically significant dependence between the variables of the following dependent/independent variable pairs:

- the case form of the attribute and the syntactic position of that attribute within the subject NP,
- the case form of the attribute and the linear position of that attribute within the subject NP.

The third tested dependence, i.e. that between the case form of the subject and the length of the subject NP, appeared on the other hand to be statistically non-significant.

---

[311] Raumolin-Brunberg 1991, 137–139.

5.3. Subject case and linear position

Another syntactic feature that awaits to be studied in LLCT is how the linear distance between the subject and its verbal head affects subject case selection. First, it is necessary to survey in general how the LLCT subjects are located with relation to the finite verb. Linear position means the position a word occupies within a sentence. As mentioned above, it could be linguistically more rewarding to measure distance lengths as the number of constituents, but this was not possible for the present because of the annotation style followed in LLCT. Therefore, the word-based indicators presented below are inevitably rough. Here, as in several other chapters of this study, only noun and adjective subjects are taken into account; pronouns, albeit typical subjects, are left aside. In addition, it always has to be remembered that clearly continuous topical subjects are usually omitted in (Classical) Latin. Due to this pro-drop, sentences with explicit subjects already constitute a specific selection.[312]

5.3.1. Linear position of subjects in LLCT

As stated above, in this and the following chapter, the graphs and statistics are usually based on the whole LLCT data. In section 4.2.2., it was shown that the 3rd declension imparisyllabic subcorpus displayed certain expected semantic dependences better than the entire LLCT. With linear position, the difference between the whole corpus and the subcorpus is not particularly striking and the small size ($N = 392$) of the subcorpus often impedes statistical testing. Therefore, numbers based on the 3rd declension imparisyllabic subcorpus will be presented here only when one of the variables under examination is the subject case. This is the case mainly in section 6.1., in which I shall analyse the interaction of the independent variables.

---

[312] Bolkestein 1996, 9.

Figure 5.9. Frequency of subjects as a function of distance from verb in LLCT (-20 to +20).
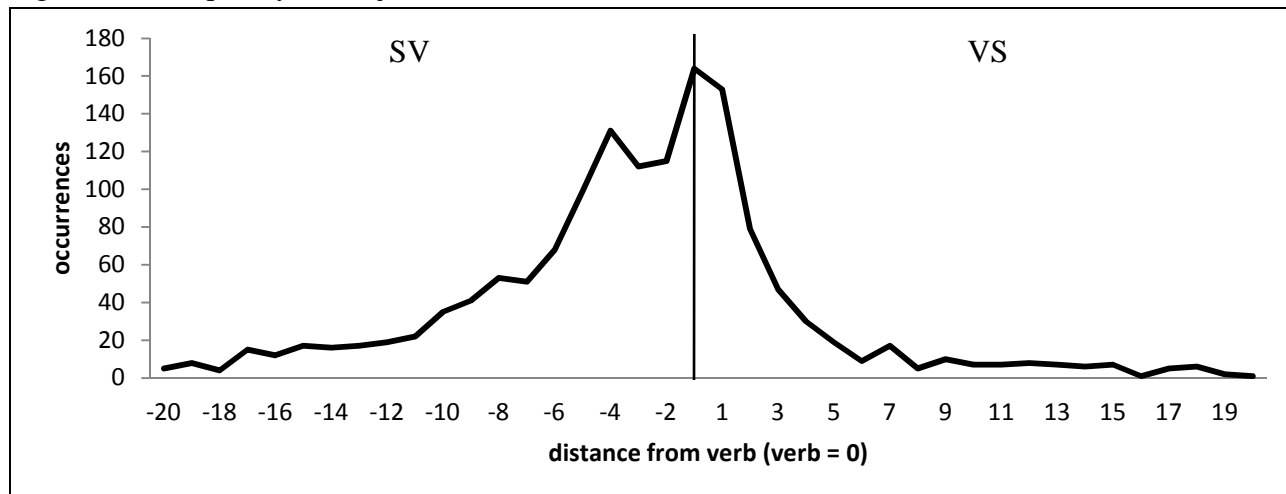
The graph in Figure 5.9. presents the frequency of occurrences of subjects (nominative and accusative) as a function of the distance from the verb in LLCT. All the subjects of structural categories A1, A2–4, and B1–2 of the basic query subset are included. The units of the horizontal axis are integers that correspond to positions occupied by the words of the sentence. The verb occupies position 0, which is, of course, not found on the graph axis. The border between positions -1 and +1 is, instead, indicated with a vertical line. For example, a subject that immediately precedes its verbal head is assigned value -1, whereas a subject that comes immediately after the verb occupies position +1. Figure 5.9. shows only the range -20 to +20 that covers the frequency peak.[313]

It is noteworthy that the majority of the subjects is located rather close to their verbal head. In other words, there seems to be a tendency to maintain the cohesion of the verbal nucleus by an immediate realisation of the primary argument. About 65% of the subjects (1,025 out of 1,583) are found within a range of 10 word position units: 8 places before and 2 places after the verb. In Charter Latin, several long and complex constructions recur, which sometimes places the subject at some distance from the verb: the most remote positions in LLCT are -63 and +102. Nevertheless, the main tendency is to keep the subject near to the verb.

Figure 5.9. also reveals that most of the subjects are located before the verbal head. Thus, the SV order dominates and the preverbal subject is usually rather close to the verb. This can be seen better in the cumulative graph in Figure 5.10. By position -1 (just before the vertical line), the function has reached the value of 1,068 and, after that, increases only by 515 (up to the value of 1,583). Thus, approximately 67.4% of the clauses with noun/adjective subjects in LLCT are SV-oriented.

---

[313] The highest occurrence numbers of the peak (range -9 to +3) can be seen in the lower graph of Figure 5.14.

Figure 5.10. Cumulative graph of the occurrences of subjects as a function of distance from verb in LLCT.
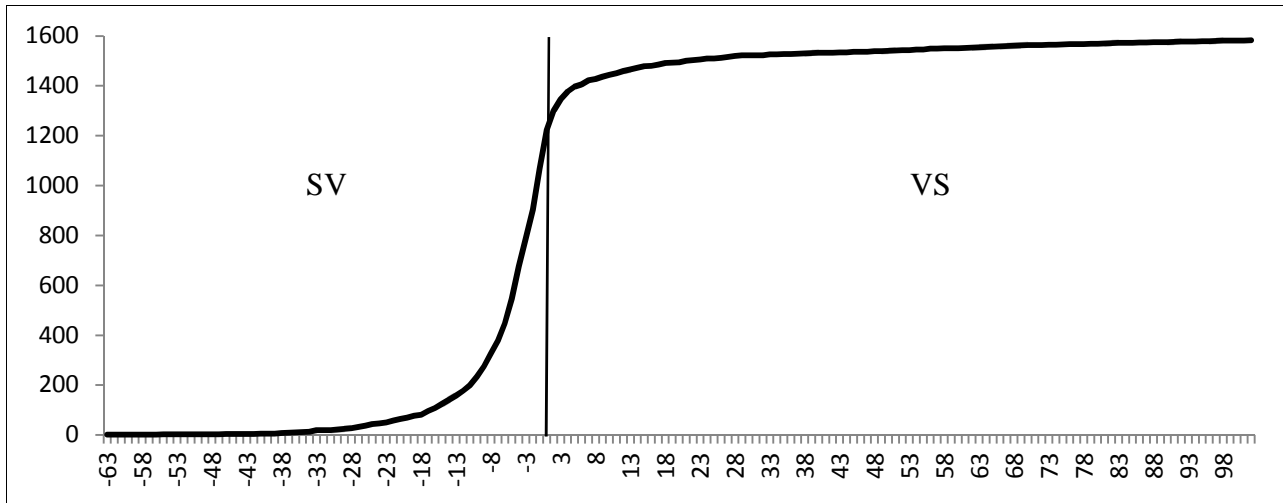
The skewing effect of the recurring formulaic phrases of charter Latin can be estimated when the corpus is examined by the structural categories A1, A2–4, and B1–2. Differences appear when the location of the subject and the verb is viewed within the scope of the entire sentence. In Table 5.4., the distance is measured as the number of word positions that separate the subject and the verb. The distance is here always a positive number and does not indicate the preverbal/postverbal status of the subject.

Table 5.4. Distance of the subject and the verb from the outset of sentence and subject/verb order in structural categories of LLCT.

Category	Average position from outset of sentence		distance	Word order (%)		N
	SBJ	verb		SV	VS	
A1	23.2	26.5	4.2	74.0%	26.0%	981
A2–4	30.4	24.5	17.5	50.2%	49.8%	480
B1–2	22.0	28.3	9.9	82%	18%	122
all	25.3	26.0	8.7	67.4%	32.6%	1,583

First, the figures in Table 5.4. reveal that the sentences of charter Latin are relatively long: on average, the verb and its subject are located in positions between 20 and 30 (words) from the beginning of the sentence.[314] While the subjects of LLCT are predominantly preverbal, category A2–4 differs from others in that its subjects are almost as often postverbal as they are preverbal. This is likely to

---

[314] Note that one sentence can include several subjects. Remember also that these statistics omit pronominal subjects.

arise from certain common types of formulaic phrases that happen to belong to category A2–4. These phrases have two (or more) coordinated subjects, which multiplies their influence with respect to the non-coordinated A1 subjects. These constructions include long lists of personal name subjects, mainly witnesses, that are often attached to a sentence-initial verb *aderant* or *erant* (see example (14)). These two verbs are responsible for 143 out of 239 postverbal subjects in category A2–4. *Venerunt* is a parallel case with 9 postverbal occurrences (example (15)). The SV order will be discussed in more detail below.

(14) MED 742 (AD 857) *erantque nobiscum Offo*, *Minto*, *Liutperto*, *Rumualdo*, *Gisperto*, [--], *Iohannes et reliqui multis*
'with us there were Offo, Minto, Liutperto, Rumualdo, Gisperto, [--], Iohannes, and several other persons'

(15) CDT 45 (AD 812) *venerunt ibi ante nos Arnulfus vicedomoi nec non et Rotprandulu*, *Aspertulu clericus et Perticausulu germanas personas altergatione habentes*
'there came up before us Arnulfus, the vice-count, as well as Rotprandulu, Aspertulu, the clerk, and Perticausulu, the brothers who had an altercation'

The length of the coordinated structures is also the reason for the exceptionally high distance between the subject and the verb in A2–4 (17.5). Category A1, instead, can be viewed as an unmarked neutral category, where a non-coordinated subject is directly attached to a non-coordinated finite verb. This is the prototypical subject/verb relation. In A1, the subject tends to precede (74.0%) the verbal head more often than in category A2–4 (but less often than in category B1–2). In addition, the subjects of category A1 are closer to their verbal head than the subjects of the other categories, the distance being only 4.2 word position units on average.

One of the reasons why the distance between the subject and the verb is longer in the coordinated categories is that, since the A2–4 and B2 subjects are in coordination, the coordinated member that is closer to the verb and the member that is more distant to the verb are separated by a coordinating conjunction or comma (as with *aut* in (16)). Of course, there can also be other elements, such as attributes, between the coordinated subject nodes. Moreover, some coordinated structures can be very long and complex (see (17); cf. (7) above).

(16) (category A2) CDL 57 *et si eum Taso aut filiis eius menare uolueris* [= *voluerint*]

'and if Taso or his sons intend to chase him'

(17) (category A3) CDL 131 *sic tamen ut homenis qui in casas massaricias meas nunc presenti*
habitant [--] *aut eorum filiis uel nepotes uel quis ex eorum germine procreati fuerint uel*
procreati sunt [--] *persoluant redditum case*
'so that the people who live currently in my leasehold houses [--] or their sons or grandsons or
anyone who will be born from their family line [--] settle the rent of the house'

Figure 5.11. and Figure 5.12. present the number of subjects as a function of the distance from the
verb for the three structural categories. There are two figures because, owing to different magni-
tudes, it would be difficult to read the three curves in one chart. It is to be noticed that, in category
B1–2 where the verbs are coordinated, the distance is always counted starting from the verb closest
to the subject.

Figure 5.11. Frequency of A1, A2–4, and B1–2 subjects as a function of the distance from the verb
(-20 to +20).
Figure 5.12. Frequency of A2–4 and B1–2 subjects as a function of the distance from the verb (-20
to +20).
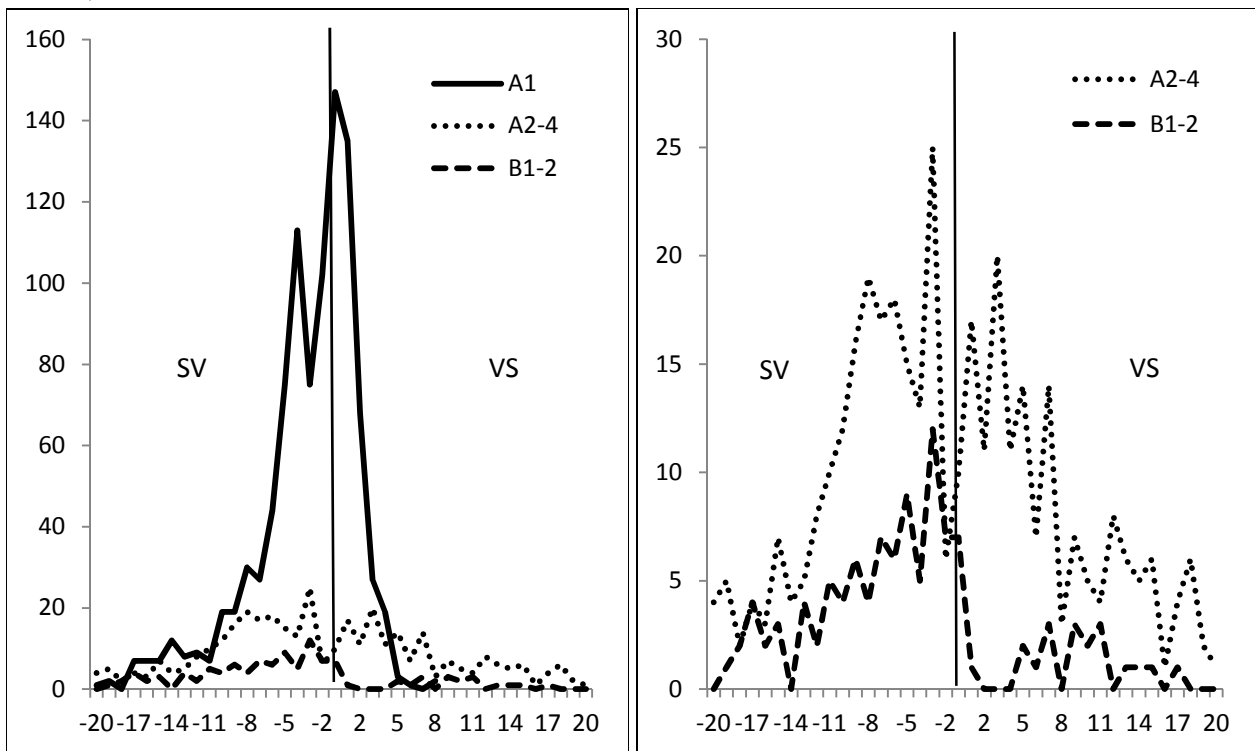
Figure 5.11. and Figure 5.12. illustrate the patterns that were noticed in Table 5.4.: the subjects of
categories A1 and B1–2 are decidedly preverbal while those of category A2–4 are divided more

equally. Category A2–4 is also special in the sense that its subjects are dispersed on a wider range than those of the other categories: several postverbal occurrences remain outside the range displayed in the above figures (-20 to +20), as can be foreseen from the distance value 17.5 of Table 5.4. As already mentioned, this is a consequence of the complicated and extremely long coordinated lists of the type presented in sentences (14) and (15).

The subjects of both categories A1 and B1–2 are limited to a more restricted range. Especially the A1 subjects come to a sudden end at value +5 and vanish completely after +8. As said above, category A1 seems to be a "natural", unmarked category where the subject is located before or just after the verb.[315] It is not surprising that the shortest sentences of LLCT belong typically to category A1. This is one of the reasons why the range of the most frequent A1 subject positions is relatively narrow. A1 represents a neutral linkage: if there is a need to say something about something, the most straightforward way is to take one verb and one subject and put them together. Therefore, A1 includes a vast variety of different verb lemmas. Because of this and the large size of A1, the distracting influence of formulaicity is not likely to skew the linguistic data extracted from that category. Thus, it may often be a good idea to test hypotheses against this "neutral" category.

Contrary to most of the A1 subjects, several clearly postverbal (coordinated) subjects of categories A2–4 and B2 are likely to be marked instances of charter language. They may be signals of charter style where large amounts of formulaic information are often integrated within a single sentence. Though the purpose of this study is not to examine the problems relative to the Late Latin word order, the theme will be touched in passing in the subsequent passages. A detailed analysis of word order is not possible because the annotation style does not allow an easy isolation of constituents.

As far as the frequencies are concerned, categories A2–4 and B1–2 seem to display more variation than A1. This is perceived as the saw-edge pattern of the A2–4 and B1–2 graphs in Figure 5.11. and Figure 5.12. These "wrinkles" result from the fact that the subjects are fewer in these categories than in A1, but also from the fact that the sentences of these categories are more prone to the twists of formulaicity in general. One major factor behind the saw-edge pattern is that categories A2–4 and B2 consist of coordinated subjects that are separated from each other by a coordinating conjunction or comma, as was mentioned above. The recurrence of coordinated formulae of this kind accentuates the peaks and troughs of the curve: the peaks are places occupied by the subjects while

---

[315] The mode of the occurrences of category A1 is -1 and the median -2.

the troughs are places occupied by the coordinating conjunctions or commas. The other possible intervening elements can, for example, be recurring attributes that do not show in this analysis, where only the highest representatives of the subject-function NPs are tagged as subject nodes of the treebank (see animacy hierarchy in section 4.1.1.).

There seems to be a considerable dip on the peak of the A1 curve in Figure 5.11. and the general LLCT curve in Figure 5.9. In fact, what is interesting is not so much the pit but the other summit at value -4 next to the highest point (-1). The frequency leap at position -4 can be explained by the fact that the very position -4 is often occupied by the subjects of common formulaic phrases, such as (18) (23 times), (19) (23 times), and (20) (7 times). In addition, personal name subjects with three postponed (adjective, noun, or genitive) attributes seem to be particularly frequent. The subject heads of these NPs are often found at position -4 (see (21)).

(18) MED 394 (AD 815) *ubi Agiprandus clericus rector esse videtur*
"where Agiprandus, the cleric, seems to be the rector" (formula "where [name] [title] seems to be the [appointment]")

(19) MED 477 (AD 826) *quantu Leupulu ad manu sua habuit*
"as much as Leupulu had in his possession" (formula "how much [name] had in his possession")

(20) CDT 45 (AD 796) *Aurulus germanus eius similiter dixit*
"Aurulus, his brother, said the same" (formula "[name] [kinship term] said/confessed in a similar manner")

(21) MED 774 (AD 865) *Alpertus notarius domni imperatoris interfui*
"(I) Alpertus, the notary of the Lord Emperor, was present"

5.3.2. Dependence between linear position and subject case

This chapter examines how the distance between the subject and the verb affects the case form assumed by the subject. This will be completed in section 6.1. by investigating, among other relationships, the dependence between construction type and relative subject position (preverbal/postverbal) and between construction type and structural category. The only dependent and independent varia-

bles that will be tested on each other in this chapter can be seen in the following table, which reprints part of Table 5.1.

Dependent variables		Independent variables		Hyp.
Variable name	Categorical levels	Variable name	Categorical levels	
case form of subject	nominative or accusative	subject position with respect to verb	distance in word positions (neg. or pos. integer)	H_1^5

The influence of the subject/verb distance on the subject case is tested with hypothesis H_1^5. The corresponding null hypothesis is H_0^5.

H_1^5: The distance between the subject and the verb affects the case form of the subject.

H_0^5: The distance between the subject and the verb has no significant effect on the case form of the subject.

This time, the formulation of the hypotheses differs from those of the previous chapters, where it was reasonable to expect a more or less linear correspondence between the dependent and independent variable, e.g. the larger the syntactic distance between the head and the attribute, the higher the accusative percentage. Here, instead, the influence of the distance is not likely to be linear because the preverbal and postverbal domains are known to behave differently in syntactic terms.

Figure 5.13. presents the accusative percentage of subjects as a function of distance from the verb in LLCT. The graph does not cover the whole scope of LLCT (i.e. -63 to +102) because the frequencies are extremely low at the outer imits of the span. Even in this graph, which ranges from -20 to +20, the saw-edge of both the fringes (outside the rectangle) reveals the variation caused by small frequencies. The behaviour of the subject case variable is statistically significant only within the range from -12 to +12 ($\chi^2 = 55.52$, $df = 23$, $p < 0.001$), but even there certain frequencies are so low that a percentage does not seem to be a fitting measure for describing the variation.

Figure 5.13. Accusative percentage of subjects as a function of the distance from the verb in LLCT (-20 to +20).
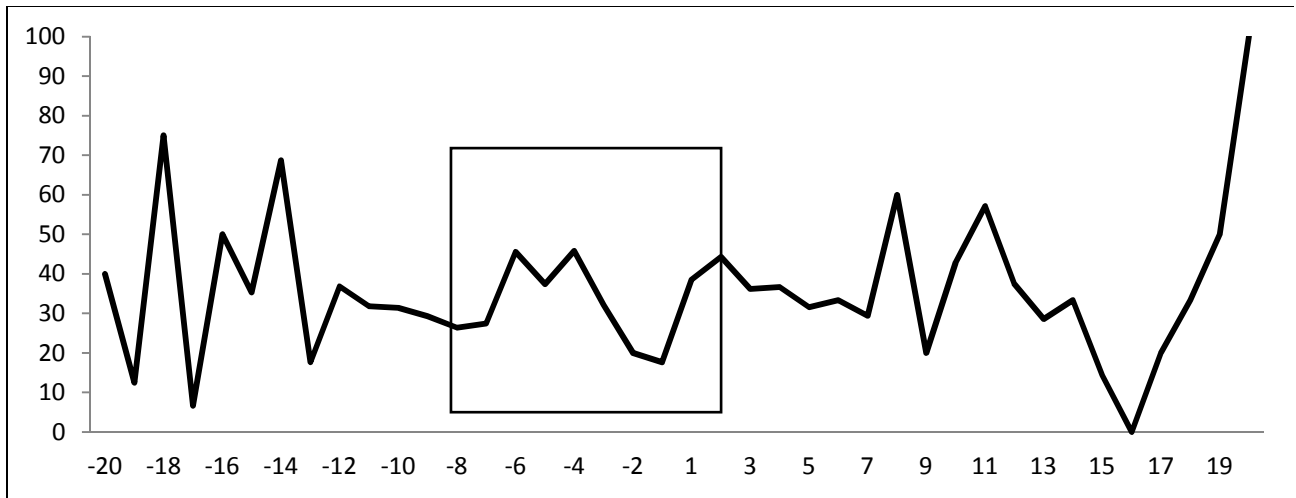
As was seen in Figure 5.9. and in the other above figures and tables, the frequency peak of the subject graph is located within a rather narrow range, outside which the frequencies grow smaller and smaller and, thus, are insufficient to meaningful percentage representation. The fewer the occurrences are, the greater the oscillation of the percentages and the lower their descriptive power. Therefore, the rectangle that overlays the curve in Figure 5.13. zooms in on the positions with more than 50 occurrences. The result of this cropping, i.e. the accusative percentage graph of the range from -8 to +2, is presented in the upper part of Figure 5.14. The lower part shows, for comparison, the combined frequency of the nominative and accusative subjects on a range from -9 to +3. Range -8 to +2 comprises 1,018 subjects, which corresponds to 64.7% of all the subjects of the basic query subset.

Figure 5.14. Accusative percentage and frequency of subjects as a function of the distance from the verb in LLCT (-8 to +2) (χ^2 = 49.69, df = 9, p < 0.001).
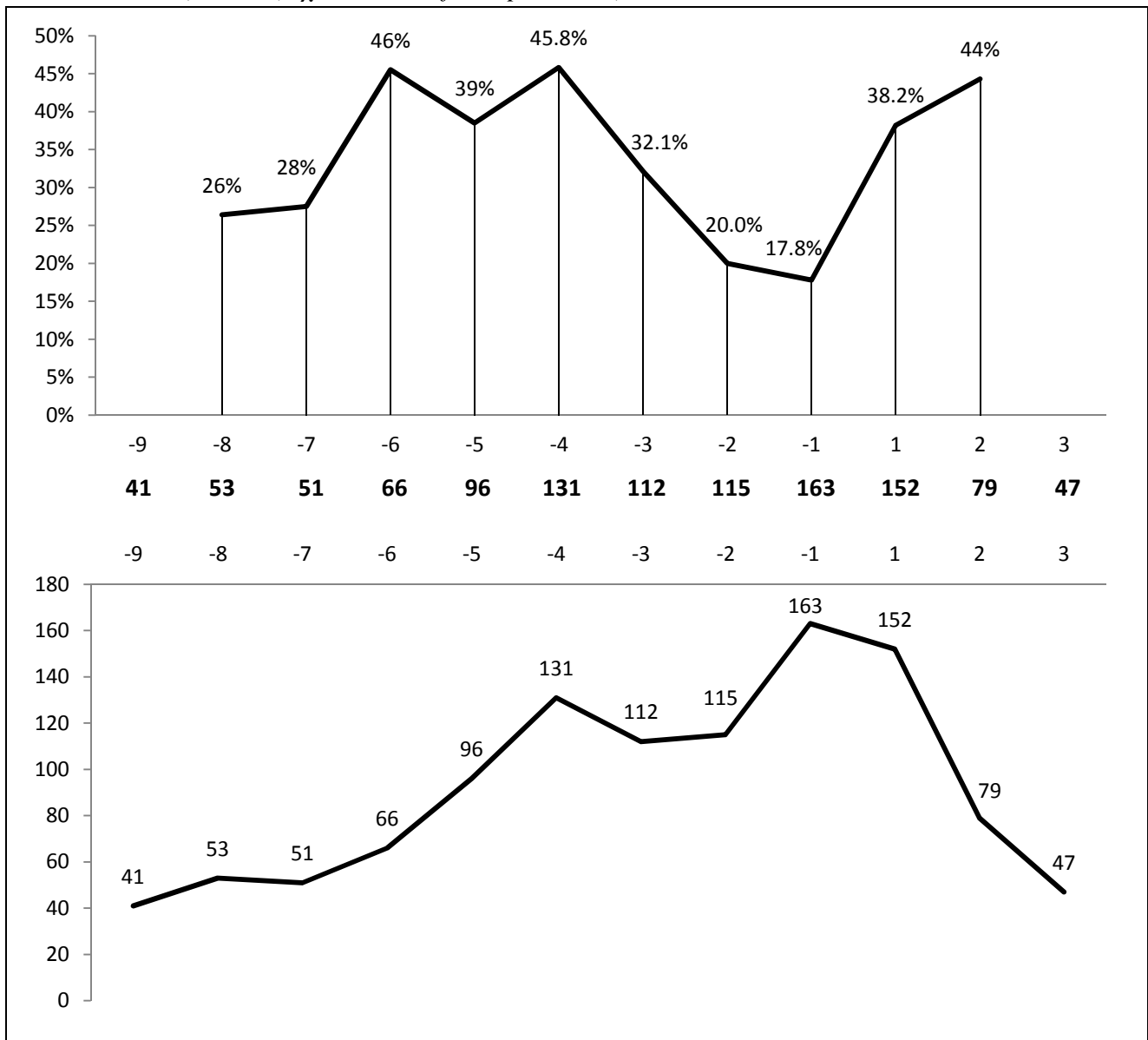
The cross-tabulation of the variables of Figure 5.14. can be seen in Appendix 5.3. The adjusted standardised residuals of the chi-square test (see Table 5.5.) reveal that statistically significant values cover positions -6, -4, -2, -1, and +2: the residuals with absolute values of 2.0 or higher are the crucial ones as far as the dependence between case and distance is concerned.[316] Moreover, there are no statistically significant residuals outside the cropped range in Figure 5.14. The residuals will be further interpreted below after Figure 5.18. The decision tree of Appendix 5.4. groups the categorical levels in statistically significant nodes that reflect practically the same pattern that is re-

---

[316] For the residuals, see section 4.2.1. and Agresti 2007, 38–39.

vealed by the graph in Figure 5.14.: the node with the lowest accusative percentage includes the positions -2 and -1.

Table 5.5. Adjusted standardised residuals of chi-square test for the data of Figure 5.14.

Variable		Distance									
case		-8	-7	-6	-5	-4	-3	-2	-1	1	2
accusative	residual	-1.0	-0.9	**2.2**	1.2	**3.3**	-0.2	**-3.1**	**-4.5**	1.5	**2.2**

In order to show a further piece of evidence for the validity of the pattern that arises out of Figure 5.14., I present here Table 5.6., which presents an essentially similar pattern for the 3rd declension imparisyllabic subcorpus. The 3rd declension imparisyllabic subcorpus was originally introduced in section 4.2.2. Due to the relatively small size of the subcorpus ($N = 392$), the dependence between the subject case and the linear position is not statistically significant on the whole range -8 to 2 that was presented in Figure 5.14. Instead, the decision tree algorithm can be used to classify the categorical variable into statistically significant groups, i.e. nodes, which appear to match rather well the pattern observed in the upper graph in Figure 5.14. The decision tree for subject case and distance can be seen in Appendix 5.5., while Table 5.6. presents the accusative percentage of the subcorpus subjects across the linear position variable that is classified into statistically significant categories ($\leq$ -3, -2 to -1, 1, 2 to 3, $\geq$ 4) on the basis of the decision tree ramification.

Table 5.6. Accusative percentage of subjects as a function of the distance from the verb in the subcorpus.

Case		Distance (categorised)					Total
		$\leq$ -3	-2 to -1	1	2 to 3	$\geq$ 4	
nominative	N	111	99	46	12	33	301
	%	70.3%	89%	78%	50%	83%	76.8%
	residual	**-2.5**	**3.7**	0.2	**-3.2**	0.9	
accusative	N	47	12	13	12	7	91
	%	30%	11%	22%	50%	18%	23%
	residual	**2.5**	**-3.7**	-0.2	**3.2**	-0.9	
Total	N	158	111	59	24	40	392
Chi-Square		$\chi^2 = 23.80, df = 4, p < 0.001$					

As mentioned earlier, the pattern that arises from Table 5.6. is rather similar to the pattern of the whole LLCT data in Figure 5.14. It is again positions -2 and -1 that have the lowest accusative percentage (11%) – and a significant adjusted standardised residual (-3.7). This residual means that the category that consists of positions -2 and -1 is clearly responsible for the highly significant statisti-

cal dependence found between the variables of the above contingency table. The same cannot be said of category ≥ 4 although it also happens to display a low accusative percentage (18%): residual -0.9 suggests that this percentage, which is not very theory-compatible, is not essential. All in all, this time the data of the subcorpus seem to support, by and large, the pattern extracted from the entire LLCT.

I shall now turn back to the analysis of the entire LLCT. The structural categories (A1, A2–4, B1–2) differ from each other, so it is necessary to examine them individually. Figure 5.15. presents the accusative percentages of category A1 subjects as a function of the distance from the verb. The rectangle again zooms in on the positions with more than 50 occurrences. This range from -5 to +2 can be seen in Figure 5.16., where the respective number of occurrences is indicated under each position.

Figure 5.15. Accusative percentage of A1 subjects as a function of the distance from the verb (-15 to +4) ($\chi^2 = 54.41$, $df = 18$, $p < 0.001$).
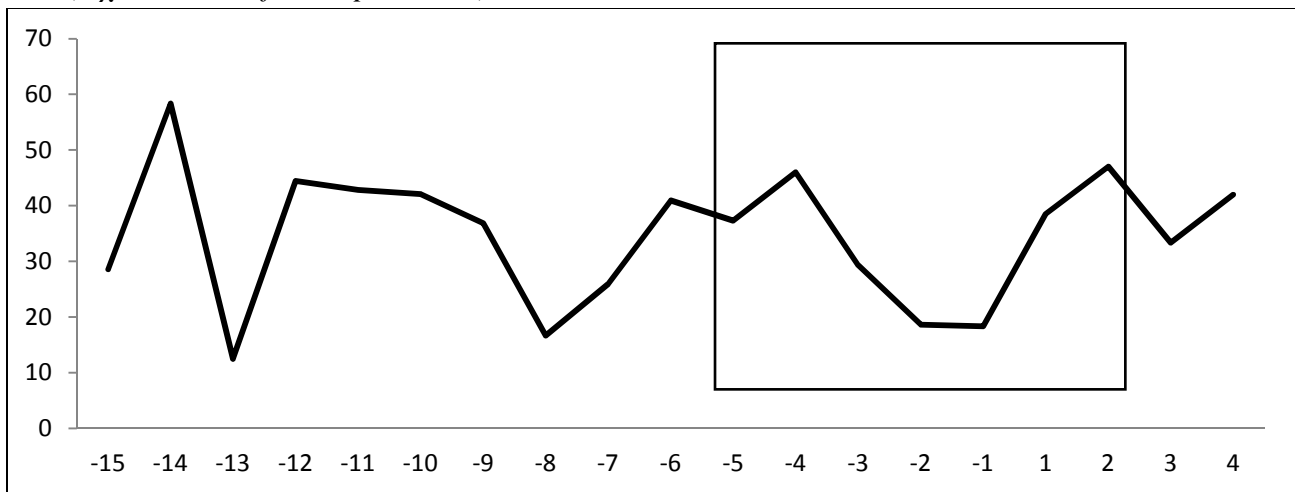
Figure 5.16. Accusative percentage and frequency of A1 subjects as a function of the distance from the verb (-5 to +2) ($\chi^2 = 41.72$, $df = 6$, $p < 0.001$).
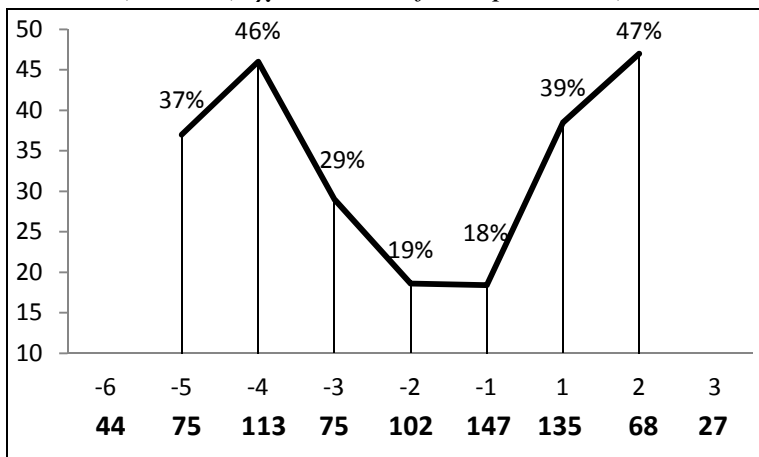
The adjusted standardised residuals of the variables (see Table 5.7.) reveal the same statistically significant positions -4, -2, -1, and +2 as in Figure 5.14. (where the entire LLCT data was examined on a range from -8 to +2). There are no statistically significant residuals outside the cropped range in Figure 5.15. (not even -6). The cross-tabulation of the A1 variables is found in Appendix 5.6.

Table 5.7. Adjusted standardised residuals of chi-square test for the data of Figure 5.16.

Variable		Distance						
case		-5	-4	-3	-2	-1	1	2
accusative	residual	1.0	**3.4**	-0.6	**-3.2**	**-4.1**	1.7	**2.7**

In sum, the pattern of category A1 is almost similar to that of the whole LLCT in Figure 5.14. How do then the other two categories, A2–4 and B1–2, relate to the whole and/or to category A1? With A2–4 and B1–2, the low frequencies frustrate the analysis. In category A2–4, the frequencies are so low that I have collapsed three adjacent positions into one histogram column in Figure 5.17. although it is questionable to merge adjacent frequencies into each other in structural categories with coordination, where certain positions are predominantly occupied by coordinators, not by coordinated subjects (see above). The A2–4 case distribution is not statistically significant within any range, which is probably due to the just mentioned merger of formulaically disproportionate positions. The only reason for presenting the A2–4 and B1–2 accusative percentages separately in the following figures concerns what they seem to have in common with category A1: the low accusative percentage at position -1.

The frequencies remain very low even in the merged frequency classes, as can be seen from the frequency figures under each column in Figure 5.17. Outside the presented range, the frequencies plunge even lower. What is noteworthy here is the column that receives the value 29%, a combination of the adjacent positions from -1 to +2. It includes the lowest accusative percentage of a single position, i.e. -1 with only 10 occurrences of which 2 are in the accusative (20%).

Figure 5.17. Accusative percentage and frequency of A2–4 subjects as a function of the distance from the verb (-10 to +8) ($\chi^2 = 6.96$, $df = 5$, $p = 0.224$).
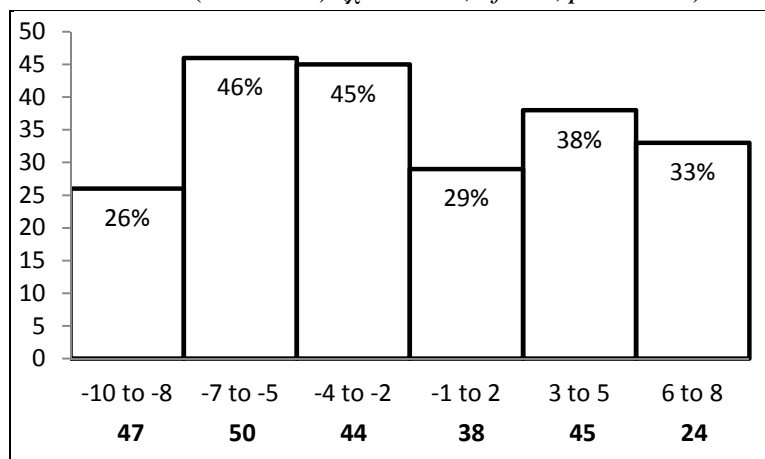
The very low B1–2 frequencies under the histogram columns in Figure 5.18. tell that the percentages are even more meaningless than those in Figure 5.17. Nor is the B1–2 case distribution statistically significant either. Nevertheless, even here the column that has the lowest accusative percentage (15%) includes the often-cited position -1. This position is comprised of 7 nominative and 0 accusative subjects, so its accusative percentage is 0. The presented range (from -12 to -1) is the largest possible one, as there are practically no occurrences before and after these positions. As was seen in Table 5.4., category B1–2 is markedly preverbal.

Figure 5.18. Accusative percentage and frequency of B1–2 subjects as a function of the distance from the verb (-12 to -1) ($\chi^2 = 2.83$, $df = 3$, $p = 0.418$).
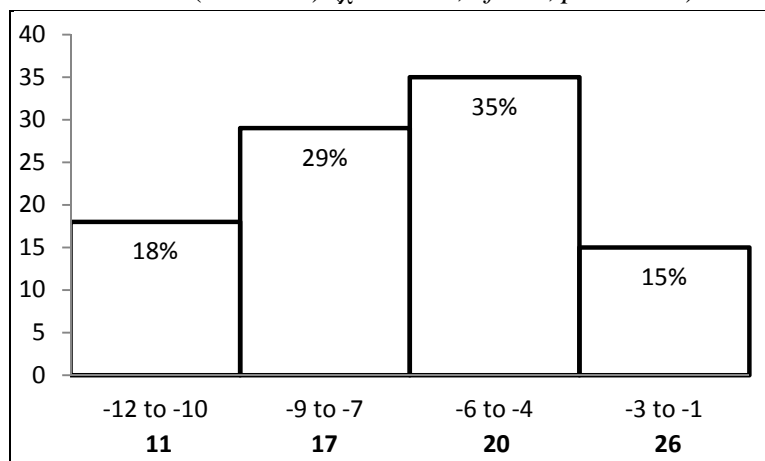
When all the categories are shown together in Figure 5.14., the heterogeneity of the formulaic categories A2–4 and B1–2 is lost under the high frequencies of category A1: the frequencies of categories A2–4 and B1–2 are too small to confuse the general trend laid down by A1. Nevertheless, alt-

hough most of the uneven patterns in Figure 5.17. and Figure 5.18. are missing in Figure 5.14., the peculiar position -1 that is clearly present in category A1 remains unchanged. As the different patterns of categories A2–4 and B1–2 are related to their different formulaic nature, these categories can be viewed as mere distractions compared to category A1, which is dominant and provides statistically significant results.

I shall next turn to this phenomenon common to all the categories, i.e. the relatively low accusative percentage immediately before the verbal head (positions -2 and, especially, -1). These positions contain most of the occurrences, so the observation can be deemed particularly reliable. Moreover, the adjusted standardised residuals for the accusative of both the whole LLCT and category A1 (Table 5.5. and Table 5.7.) show the strongest correlation between the case and position variables at -1. The high negative residuals -4.5 and -4.1 indicate that the number of accusative occurrences is significantly smaller than would be expected were the null hypothesis true. The same goes for position -2, too, albeit less pronouncedly. Even the other significant residuals at positions -4 and +2 corroborate the interpretation: together with those at positions -2 and -1, they delineate a neat, statistically significant dive in the curve.[317]

The other dives in the curves, on the other hand, such as the one at -7 and -8 in Figure 5.14., result most probably from mere random variation, as the occurrences responsible for that plunge are relatively low: 51 and 53, respectively. Moreover, their residuals are not significant (see Table 5.5.). In sum, only the systematically recurring dive at positions -1 and -2 seems to be a real, linguistically significant pattern.

This relatively low accusative percentage before the verbal head is, indeed, a remarkable observation. It is also an observation that has never been done earlier because charters are the only substantial Latin material with considerable deviations from the standard. Moreover, several scholars have studied charter texts for decades, but they have not realised this phenomenon because they have not utilised the (computational) corpus method. I assume that the immediate preverbal position should be considered the most natural, unmarked subject position of an SV-oriented language. I suggest that the scribes managed to produce the nominative subject, required by the classical norm, more easily in this kind of canonical position than in other positions. This is psycholinguistically reason-

---

[317] This does not apply so much to position -6 of Figure 5.14. (whole LLCT). It is to be noticed that position -6 is not statistically significant in the data of category A1, so it may be related to the random interference of the formulaic categories A2–4 and B1–2.

able: at position -1, the predicate, the verb that controls its arguments, is imminent. Thus, it has the maximal power to impose the case agreement on the preceding word, which in this case is the subject. That is to say, the writer is capable of realising optimally the syntactic cohesion required by the verbal nucleus when he is already anticipating the immediately following predicate at the moment of writing the subject. The further off from the predicate a word is, the weaker the control that is felt by anticipating the verb: the cohesion that imposes a certain case form on a certain argument is loosened. Moreover, the fact that position +1 has a considerably higher accusative percentage than position -1 suggests that it is not only the dependency length, i.e. the distance, but also the relative order SV that is decisive for the subject case selection. In other words, it is more the *immediate preverbal position* that counts than its being only immediate or only preverbal.

The above explanation is a linguistic one. However, the phenomenon may be partly extralinguistically, i.e. mnemonically or psychologically, motivated. Charter Latin is not a natural language in the sense that most of the formulae were obligatory. The scribes memorised the outlines of documents and then reproduced the formulae, hence the variations resulting from lapses of memory or from the overlapping of two mental images. The sentences were also long and burdened with subordinate clauses and parenthetical specifications. Indeed, the average sentence length in LLCT is 24.5 words and, although the most frequent length is 9 words, sentences with 50 to 100 words are not rare. The number of embedded structures is high as well.

It is, then, obvious that temporary blackouts occurred when writing lengthy sentences. If the word that was intended to be the subject and its verbal head were distant from each other, the cohesion of the structure was easily lost and the subject ended up in a "wrong" case form. On the other hand, the syntactic and psychological explanations are likely to interfere with each other to varying degrees. The question is whether 'syntactic' and 'mnemonic' are after all only two names of a single phenomenon. I suggest that the two terms can be seen as the two poles of a continuum: the longer the physical distance (time or word positions) between the subject and the verb is, the stronger the role of the mnemonic factor. With shorter distances, the psychological and syntactic motivations are likely to be blended.

5.4. Summary

Chapter 5 showed that an empirical study of case alignment cannot ignore the syntactic context even though semantics is the driving force of the intransitivity split in all likelihood. Moreover, cer-

tain factors that can be classified as syntactic seem to be so closely connected with the working memory that they can be characterised as mnemonic. As the basis of the analysis, I accepted a psycholinguistic theory that considers syntactic complexity and dependency length to be one and the same thing. This dependency length reflects performance preferences (Hawkins 2004) and is usually measured as word counts in this study.

Section 5.1. revealed that the depth of coordination, which was used as a rough equivalent of dependency length/syntactic complexity, affects subject case selection to some extent. Section 5.2. showed that the distance between the subject NP head and attribute, measured both as dependency levels and word-positions, seems to affect the case selection of the attribute: the further off from the subject NP head the attribute is located, the more easily the case form of the attribute slips into the accusative and the higher is its accusative percentage. The relationship between the subject case and the dependency level variable is statistically more significant, which suggests that the fundamental factor is really the dependency length, whereas the word count reflects it less accurately and may be, thus, epiphenomenal.

I also briefly discussed the word order of LLCT in section 5.3.1. There appears to be a tendency to maintain the cohesion of the verbal nucleus, i.e. to keep the dependency lengths short, by an immediate realisation of the subject argument before the verbal head (SV(O) order dominates). It is especially the structural category A1 that can be viewed as an unmarked neutral category in which a non-coordinated subject is attached directly to a non-coordinated finite verb, thus representing the prototypical linkage between the subject and the verb.

In section 5.3.2., I examined how the linear distance between the subject and its verbal head affects subject case selection. The most important pattern that arises both from LLCT as a whole and from the 3rd declension imparisyllabic subcorpus is the exceptionally low accusative percentages at the immediately preverbal positions -2 and -1. Indeed, subject positions -2 and, especially, -1 display the lowest accusative percentages in all the structural categories A1, A2–4, and B1–2. The immediate preverbal position is the position where the predicate is imminent, the anticipated verbal control and agreement are at their strongest, and where the realisation of the marked form, i.e. the nominative, is thus more probable than in other, less prototypical positions. This is why the nominative has the highest frequency in the immediate preverbal position.

The above explanation merges the linguistic (syntactic) and external (psycholinguistic) motivations. Sentences of charter Latin are often long and heavy with embedded structures and parenthetical specifications. As the charter formulae were produced from memory, it is obvious that the mnemonic and syntactic factors were intertwined. It is clear that some variations resulted banally from lapses of memory or from the overlapping of two mental images, while some were more related to the subtle interplay of language performance and dependency length. I suggest that, as regards the written charter Latin, the psycholinguistic/mnemonic factors and the syntactic factors can be seen as the two poles of a continuum: the longer the physical distance between the subject and the verb is, the stronger the role of the banal mnemonic factor. With shorter distances, psycho-syntactic explanations are likely to be blended.

To conclude chapter 5, I present those dependent/independent variable pairs that displayed statistically significant dependences when examined in this chapter (cf. Table 5.1.):

- the case form of the subject and the structural category in which the subject occurs (section 5.1., H_1^1),
- the case form of the attribute and the syntactic position of that attribute within the subject NP (section 5.2., H_1^2),
- the case form of the attribute and the linear position of that attribute within the subject NP (section 5.2., H_1^3),
- the case form of the subject and the subject position as a function of the linear distance from the verb (section 5.3.2., H_1^5).

6. Interaction of syntax and semantics

Chapter 6 examines the interaction of the syntactic and semantic variables that have been discussed in chapter 4 and chapter 5. In section 6.1., the relationships between several syntactic and semantic variables will be tested both with contingency tables and decision trees. The dependence between the variables will be again defined with the chi-square tests, as explained in section 4.2.1.

6.1. Relative subject position, construction type, and structural category

In this chapter, my aim is to detect dependences between various variables that have already been touched on in the previous chapters. The aim is to find out how those variables interact that were

shown to affect the subject case selection in the chapter 5. Relationships between the variables are first examined with contingency tables and, at the end of the section, the interaction of the most central syntactic and semantic variables is tested on the subject case by utilising the decision tree classification.

For the first part of this section, I leave aside the word-position distance between the subject and the verb and concentrate mostly on the relative position of the subject in relation to the verb. The relative position is about whether the subject is located before the verbal head (SV) or after it (VS). I shall not engage in a detailed study of the subject/verb order of LLCT, but only touch upon the field as far as it is relevant for understanding morphosyntactic alignment in LLCT. After that, I seek to find out to what extent relative position, construction type, structural category (A1, A2–4, B1–2), animacy, and linear position interact. The relationships that will be tested in this section are as follows:

Table 6.1. Dependent and independent variables with their categorical levels (section 6.1.).

Dependent variables		Independent variables		Hyp.
Variable name	Categorical levels	Variable name	Categorical levels	
case form of subject	nominative or accusative	relative subject position	word position (preverbal/postverbal)	H_1^1
relative subject position	word position (preverbal/postverbal)	construction type	A, S_A, S_O, and S_O passive	H_1^2
construction type	A, S_A, S_O, and S_O passive	structural category	category (A1, A2–4, B1–2)	H_1^3
structural category	category (A1, A2–4, B1–2)	relative subject position	word position (preverbal/postverbal)	H_1^4
construction type	A, S_A, S_O, and S_O passive	subject position with respect to verb	distance in word positions (neg. or pos. integer)	H_1^5
animacy class	personal, animate, inanimate	subject position with respect to verb	distance in word positions (neg. or pos. integer)	H_1^6
case form of subject	nominative or accusative	formulaicity degree	free/formulaic	H_1^7

Based on what was theorised about the cohesion of the verbal nucleus at the end of the preceding section, I shall now examine whether syntactic cohesion is stronger when the subject is preverbal and the nominative, thus, more frequent in the preverbal position than in the postverbal position in LLCT. This hypothesis is based on two related assumptions: 1) that Latin was predominantly

SV(O) oriented at the time of LLCT,[318] 2) that the preverbal position is the canonical subject position, whereas the postverbal position is the canonical position of the object. These issues were already touched on preliminarily in section 5.3.2.

I shall test the relationship between the relative (preverbal/postverbal) subject position and the case form of the subject with hypothesis H_1^1. The corresponding null hypothesis is H_0^1.

H_1^1: The relative position of the subject affects the case form of the subject.

H_0^1: The relative position of the subject has no significant effect on the case form of the subject.

This time, I test hypothesis H_1^1 on the data of both the entire LLCT and the 3rd declension imparisyllabic subcorpus. Table 6.2. points out that there is a statistically significant dependence (χ^2 = 5.29, df = 1, p = 0.021) between the case form of the subject and its relative position in the entire LLCT: the combined accusative percentage of preverbal subjects (32.4%) is 6 percentage points lower than that of postverbal subjects (38.3%).[319]

Table 6.2. Dependence between subject case and relative subject position in the entire LLCT.

Subject position	Case			N
	nominative	accusative	% accusative	
preverbal	722	346	32.4%	1,068
postverbal	318	197	38.3%	515
Total	1,040	543	34.3%	1,583
Chi-square	χ^2 = 5.29, df = 1, p = 0.021			

In the 3rd declension imparisyllabic subcorpus, no statistically significant dependence is attested (χ^2 = 0.79, df = 1, p = 0.374). This probably results from the fact that the subjects with 'distracting' ambiguous endings, such as -o, were removed from the subcorpus. Although the subjects of the

---

[318] Ledgeway 2012, 225–234, 335–336; Bauer 2009, 269–271; cf. Pinkster 1991, 79–80. Classical Latin was predominantly SV(O)-ordered but the VS order was not alien either. Bolkestein (1996) reports a percentage of c.20% for the declarative main clauses with explicit subjects in a sample of Caesar and Cicero's letters. She analyses the marked, 'emotive', VS order by way of several parameters including discourse-pragmatic topicality and focality, the semantic valency of the verb, and the relative number of coherence/cohesion markers. Bortolussi (2013, 3–4) discusses the semantic parameter by highlighting the role of unergative verbal semantics. Thus, he connects the discussion to the syntactic-semantic primitives (A, S_A, and S_O) that Ledgeway (2012) considers important in the formation of (Late) Latin and Romance word order.

[319] Note that because distance is defined as either positive or negative in this study, it incorporates the variable 'relative position' (preverbal/postverbal). Therefore, Table 6.2. can be viewed as a further test of hypothesis H_1^5 of section 5.3.2.

subcorpus are with certainty either nominative or accusative forms, they are only representative of the 3rd declension subjects, not of the whole Latin declension. This imbalance is likely to skew those distributions that the statistically significant dependence of Table 6.2. is based on. For example, as most of the personal names were excluded along with the other 2nd declension subjects, a major part of the A constructions was lost, and it is the A constructions that often have a preverbal subject, as will be explained below in Table 6.3.

Indeed, the difference between the whole LLCT material and the subcorpus is considerable. Note also that the statistically significant dependence that is observed in Table 6.2. may result essentially from the fact that the positions -1 and -2 with an exceptionally high nominative percentage are preverbal (see Figure 5.14.). That the dependence between relative subject position and subject case selection is not clear in the 3rd declension subcorpus may be, indeed, further proof for the conclusion that the immediately preverbal position is what matters and not so much the relative position by itself.

I shall leave the subject case for a while in order to discuss construction type and its relation to relative subject position. This is needed to analyse the connections between the independent variables that have been used to explain the variation in the subject case variable. Other studies have, indeed, demonstrated that verbal semantics (construction type) and relative (preverbal/postverbal) subject position are related to each other. These studies prove that the subject of a transitive verb is preverbal more often than the subject of an intransitive verb in Classical Latin prose and poetry. Similarly, the S_O subject of a passivised transitive verb is postverbal more often than the subject of the same active verb.[320] Moreover, it has been shown that there is a relationship between VS word order and the unaccusative construction type in Classical Latin. It is, however, not only the unaccusative verb that counts, but the whole morphosyntactic alignment. When discussing the VS order, Ledgeway (2012) suggests that the same factors may be at play in Late Latin that cause the postposition of the S_O subjects in the modern, otherwise SVO-oriented, Romance languages:

---

[320] Ledgeway 2012, 335–336; Bolkestein 1996, 19–20; Bortolussi 2013, 3–4; Spevak 2010, 145; Adams 1976b, 121–126, 1976a, 95; Siewierska 1988, 47. See also chapter 3 of Spevak (2010), which is an excellent study on the placement of the verb and its arguments in declarative sentences of classical prose as far as their pragmatic values are concerned (Spevak 2010, 115–193).

"[T]his SVO order masks in most modern varieties an active/[inactive] alignment, where S and O are to be understood more broadly as A/S_A and O/S_O, respectively."[321]

I shall test the dependence between the relative (preverbal/postverbal) subject position and the construction type with hypothesis H_1^2. The corresponding null hypothesis is H_0^2.

H_1^2: The construction type is related to the relative position the subject assumes with respect to its verbal head.

H_0^2: The construction type has no significant effect on the relative subject position.

Table 6.3. presents the cross-tabulation and the result of the chi-square test. The subjects are classified here again as subjects of transitive constructions (A), subjects of unergative constructions (S_A), and subjects of unaccusative constructions (S_O). The subjects of passive constructions (also S_O) are analysed separately from other subjects of unaccusative constructions.

Table 6.3. Dependence between relative subject position and construction type in LLCT.[322]

Relative position		Construction type				Total
		A	S_A	S_O	S_O pass.	
preverbal	N	558	66	337	99	1,060
	%	**75.6%**	76%	**53.2%**	86%	67.4%
	residual	**6.5**	1.7	**-9.8**	**4.4**	
postverbal	N	180	21	296	16	513
	%	24.4%	24%	46.8%	14%	32.6%
	residual	-6.5	-1.7	9.8	-4.4	
Total	N	738	87	633	115	1,573
Chi-square		$\chi^2 = 101.51$, $df = 3$, $p < 0.001$				

Table 6.3. shows that null hypothesis H_0^7 can be rejected: there is a statistically significant dependence between the subject position and the construction type ($\chi^2 = 101.51$, $df = 3$, $p < 0.001$) in the entire LLCT. The A subjects of transitive verbs are preverbal in 75.6% of the cases while the same

---

[321] Ledgeway 2012, 335–336, 231–234, with examples from modern Italian and Catalan; cf. Sornicola 1995, 74–83; Adams 1976a, 95–98; Adams 1976b, 122–129. On the semantic (and discourse-pragmatic) motivation of the Italian subject inversion, see Maiden 1998, 203. See Cennamo 2001b, 15–16, for the role of word order in certain extended accusative constructions of the *Lex Curiensis*.

[322] The total numbers of preverbal and postverbal subjects differ between Table 6.2. and Table 6.3. because the ten subjects that are attached to two (or more) verbs of different semantic type have been subtracted from the numbers of Table 6.3. (see section 5.1.).

percentage for the S_O subjects of unaccusative verbs is only 53.2%. As they do in several other contexts (see chapter 5), the S_A subjects of unergative intransitive verbs align with the transitive and not with the other intransitive verbs. This is, of course, in keeping with the semantically-based alignment, as suggested by Ledgeway (above). The passive verbs, on the other hand, behave against expectations.[323] The most important categories for this study are, however, the A and S_O subjects that have the highest numbers of occurrences and that are at the opposite ends of the transitivity continuum (see section 4.2.1.). The adjusted standardised residuals show that, technically speaking, all the construction types other than S_A (±1.7) are relevant for the high chi-square value of the test.

As already mentioned, the observations made on the basis of Table 6.3. (and Table 6.4.) seem to support Ledgeway's suggestion (above) about the A/S_A and O/S_O split of the Late Latin SVO order. If the passive is left aside, the observations reflect a tendency for actor and undergoer subjects to be structurally distinguished, i.e. to occupy different positions, according to the semantically-based alignment. So, the semantically-based alignment is not only visible in the case marking of the subject, but also in the linearisation of the elements within the clause. Of course, in order to prove this in a more binding manner, a profound study of the linear positions of LLCT objects would be needed.[324] This is, however, beyond the scope of this thesis. Here, it is enough to state that the relative position of the subject is related to the construction type in which the subject occurs.

Next, I shall discuss the relationship between construction type, structural category (A1, A2–4, B1–2), and relative subject position. I find it useful to check how the structural category to which the subject belongs is related both to the construction type and to the preverbal/postverbal ratio of the subjects. By doing this, it is possible to analyse how the formulaicity of charter Latin affects the relative position data. It was shown in section 5.1. that because the formulaic expressions of charter language are distributed differently across structural categories, the categories reflect heavily the grammatical properties of those formulaic patterns that happen to be frequent in each. The construction type is included in the discussion in order to find out whether it is more the construction type or the structural category that affects the relative subject position. The plain dependence between structural category and the relative subject position without the construction type can be seen below in Figure 6.1.

---

[323] The S_O passive subjects ought to align with the other S_O subjects. The passive constructions of LLCT are, however, usually crystallised relative clauses of the type *unde brevem scriptum ante nos legebatur* "of which the record was read before us".

[324] According to a preliminary study, the most common object categories are unexpectedly slightly more OV-oriented than VO-oriented. This needs to be verified, however.

The cross-tabulation of Table 6.4. presents the dependence between the construction type and the structural category for both the preverbal and postverbal subject positions. Hypothesis H_1^3 is tested in order to find out whether there is a dependence between structural category and construction type. The corresponding null hypothesis is H_0^3.

H_1^3: The construction type is related to the structural category to which the subject belongs.

H_0^3: The structural category has no significant effect on the construction type.

Table 6.4. Dependence between structural category and construction type in preverbal and postverbal subject positions in LLCT.

Structural category		Relative position									
		preverbal					postverbal				
		Construction type									
		A	S_A	S_O	S_P	Total	A	S_A	S_O	S_P	Total
A1	N	365	40	252	67	724	105	6	128	15	254
	%	**50.4%**	6%	**34.8%**	9%	100%	**41.3%**	2%	**50.4%**	6%	100%
	residual	**-2.1**	-1.4	**3.1**	-0.1		**2.9**	**-2.0**	**-3.3**	**3.6**	
A2–4	N	122	21	72	24	239	61	10	166	1	238
	%	**51.0%**	9%	**30%**	10%	100%	**26%**	4%	**69.7%**	0%	100%
	residual	-0.6	1.9	-0.6	0.4		**-4.2**	0.1	**5.1**	**-3.3**	
B1–2	N	71	5	13	8	97	14	5	2	0	21
	%	**73%**	5%	**13%**	8%	100%	**c.70%**	c.20%	**c.10%**	-	100%
	residual	**4.3**	-0.5	**-4.1**	-0.4		**3.1**	**4.7**	**-4.6**	-0.8	
Total	N	558	66	337	99	1,060	180	21	296	16	513
	%	**52.6%**	6%	**31.8%**	9%	100%	**35.1%**	4%	**57.7%**	3%	100%
Chi-square		$\chi^2 = 24.70$, $df = 6$, $p < 0.001$					$\chi^2 = 59.37$ (Fisher's exact), $df = 6$, $p < 0.001$				

The lowest line (Total %) shows that in LLCT, transitive clauses (A) favour the preverbal position (52.6%). On the other hand, the preferred subject type in the postverbal position is the S_O subject of unaccusative clauses (57.7%). Accordingly, it can be seen that the A subject rate is about 17 percentage points higher in the preverbal position (52.6%) than in the postverbal position (35.1%). The corresponding difference between the S_O subjects of unaccusative constructions (31.8% vs. 57.7%) is about 26 percentage points.

There are considerable differences between the structural categories, however. In category B1–2, the preverbal position favours the A subject more clearly (73%) than in the other categories and, even in the postverbal position, the A subject (c.70%) outnumbers the S_O subject (c.10%), unlike in the other categories. There are only 21 postverbal occurrences though. In category A2–4, the direction of the difference between the A and S_O percentages is similar to that in A1, but is more accentuated. This behaviour is most likely to result from the different degree of formulaicity of the categories and is, therefore, partly extra-linguistic. For example, the high postverbal S_O subject percentage (69.7%) of category A2–4 is probably due to the frequent witness lists with a sentence-initial verb *aderant* or *erant* (see example (4) in section 5.1. and example (14) in section 5.3.1.).

The above null hypothesis H_0^8 can be rejected because there seems to be a statistically significant dependence between the structural category and the construction type of the subject both in the preverbal and postverbal positions (in spite of the fact that category B1–2 does not practically distinguish preverbal and postverbal subjects). The adjusted standardised residuals support the interpretation that it is mainly the A and S_O subjects that create the statistically significant dependence in Table 6.4. They also form the greatest semantic and, obviously, even syntactic contrast between each other.

At this stage, it is interesting to return temporarily to section 5.1. and to review, in the light of the numbers of Table 6.4., Table 5.2., which presents the dependence between the subject case and the structural category. In Table 5.2. in section 5.1., it was noticed that the subjects of category B1–2 have a notably low accusative percentage (20%), whereas A1 has 32.7% and A2–4 40.7%. The construction type distribution of the above Table 6.4. suggests that this bias towards the nominative is related to the exceptionally high A subject percentage both in preverbal (73%) and postverbal (c.70%) positions in category B1–2 and probably has little to do with the structural complexity of the category. Moreover, both category B1–2 and construction type A are preverbal (see Table 5.4. and Table 6.3.). This fact and Figure 5.18., which shows how the accusative percentage is as its lowest (15%) precisely on the range that includes the often mentioned position -2 to -1, seems to prove that the low accusative percentage (20%) of category B1–2 is more due to other (interrelated) variables than to a real structural complexity of category B1–2.

The above observations can be seen as evidence against the use of structural categories as indicators of structural complexity. The complexity of the coordination structures may play a role in the sub-

ject case selection, but it is likely to be outweighed by other factors, such as the skewed distribution of formulaic expressions in a specific structural category. In my estimation, there is no doubt that the accusative percentage of the A1 subjects (32.7%), which is considerably lower than that of the A2–4 subjects (40.7%) in Table 5.2., results partly from the lower complexity degree of category A1. However, this relatively low 32.7% of the A1 subjects also seems to be related to the fact that the A1 subjects are, on average, more agentive than the A2–4 subjects: in category A1, there are more A subjects and fewer S_O subjects than in category A2–4.[325]

Table 6.4. presented the dependence between the structural category and the construction type separately for the preverbal and postverbal subject positions. The relationship between the structural category and the relative position can, however, be seen more illustratively in the decision tree in Figure 6.1., which is created to test hypothesis H_1^4. The corresponding null hypothesis is H_0^4.

H_1^4: The structural category to which the subject belongs is related to the relative position of the subject.

H_0^4: The relative position has no significant effect on the structural category.

---

[325] This can be seen from the numbers of the above Table 6.4.: the A percentages of category A1 (50.4% and 41.3%) are higher, if treated together, than those of category A2–4 (51.0% and 26%), while the S_O percentages of category A1 (34.8% and 50.4%) are lower, if treated together, than those of category A2–4 (30% and 69.7%).

Figure 6.1. Decision tree for the structural category (dependent) and the relative subject position (independent).
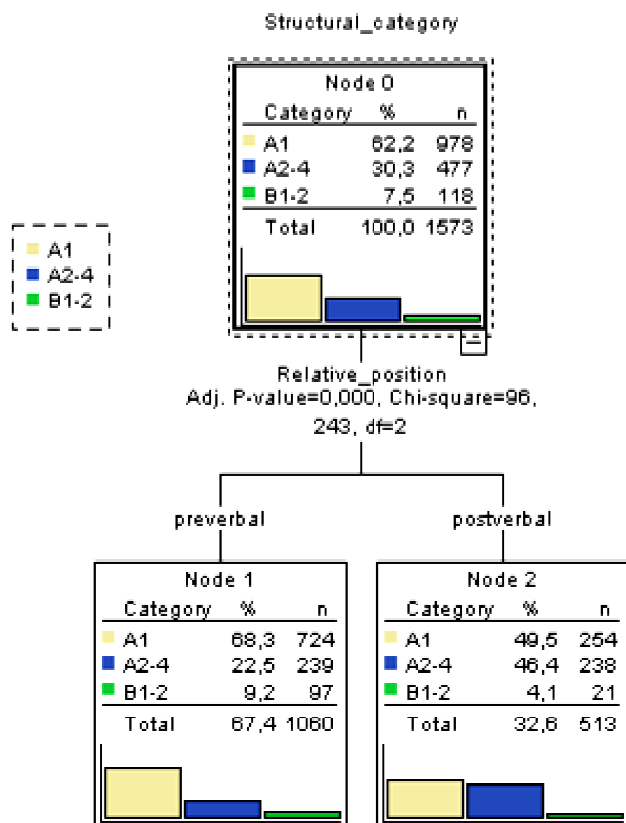
Figure 6.1. shows a highly significant dependence ($\chi^2 = 96.24$, $df = 2$, $p < 0.001$) between the varia-bles, so null hypothesis H_0^4 can be rejected. I highlight here category A1, which was argued to be a "neutral" or "unmarked" structural category in section 5.3.1. This interpretation is well in accord-ance with the fact that, in a predominantly SVO-ordered language, such as Late Latin, the preverbal subject position is usually the unmarked one: indeed, 68.3% of the preverbal subjects of LLCT are A1 subjects (see node 1) while, in postverbal position, this percentage is 49.5% (see node 2).

I shall now leave the relative position and return to the linear position variable that was discussed in section 5.2. It is evident that the dependence of the construction type and the subject position can be examined more accurately by word position units than with the binary relative position variable that is used conventionally in word order studies. Figure 6.2. reveals a highly significant dependence ($\chi^2 = 161.08$, $df = 27$, $p < 0.001$) between the construction type and the subject position. It is again ob-vious, however, that the distribution of different subject types that is presented in Figure 6.2. results partly from the uneven distribution of formulaic expressions across the word positions (the same applies to the dependence of animacy and linear subject position in Figure 6.3.). In this respect, the

distribution is certainly more or less accidental, but a statistically significant dependence between the construction type and the linear subject position is detected notwithstanding. Even a slightly wider range would have been statistically significant, but only the excerpt where the subject frequencies are higher than 50, i.e. the same range as in Figure 5.14., is presented. The contingency table and the adjusted standardised residuals of the chi-square test can be seen in Appendix 5.7.

The chi-square test behind Figure 6.2. tests the following hypothesis H_1^5. The corresponding null hypothesis is H_0^5.

H_1^5: The construction type of the subject is related to the linear position of the subject.

H_0^5: The linear position has no significant effect on the construction type.

Figure 6.2. Construction type (%) as a function of the distance from the verb (-8 to +2) in LLCT (χ^2 = 161.08, df = 27, $p < 0.001$).
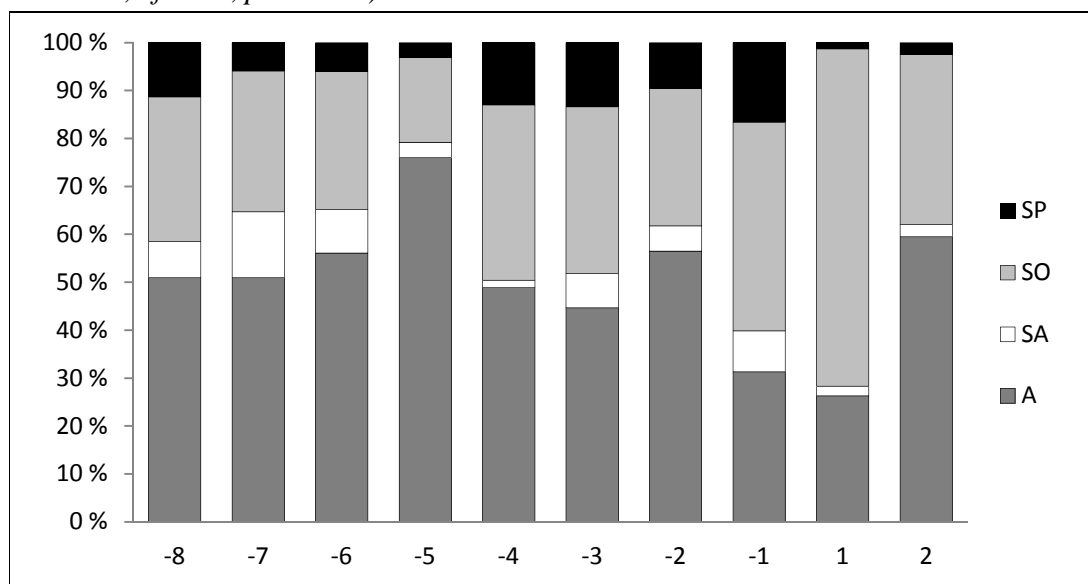
The dependency between linear position and construction type in Figure 6.2. can be perceived in the following trends: the relative share of A subjects seems to decrease when approaching position +1 from the left. Subsequently, the A percentage regains its previous portion at +2 and continues to be relatively high even outside the presented range. In contrast, the S_O percentage seems to increase along with the decrease of the A percentage before +1.

Chapter 4 showed that the construction type and animacy are tightly related. Therefore, it is interesting to examine whether animacy produces a pattern that is similar to the pattern produced by the construction type as tested on the linear subject position in Figure 6.2. This is, indeed, the case, as is seen in Figure 6.3., which is based on the chi-square test applied to hypothesis H_1^6. The corresponding null hypothesis is H_0^6, and the adjusted standardised residuals of the chi-square test can be seen in Appendix 5.8.

H_1^6 : The animacy class of the subject is related to the linear position of the subject.

H_0^6 : The linear position has no significant effect on the animacy class.

Figure 6.3. Animacy class (%) as a function of the distance from the verb (-8 to +2) in LLCT ($\chi^2 =$ 133.53, $df = 18$, $p < 0.001$).
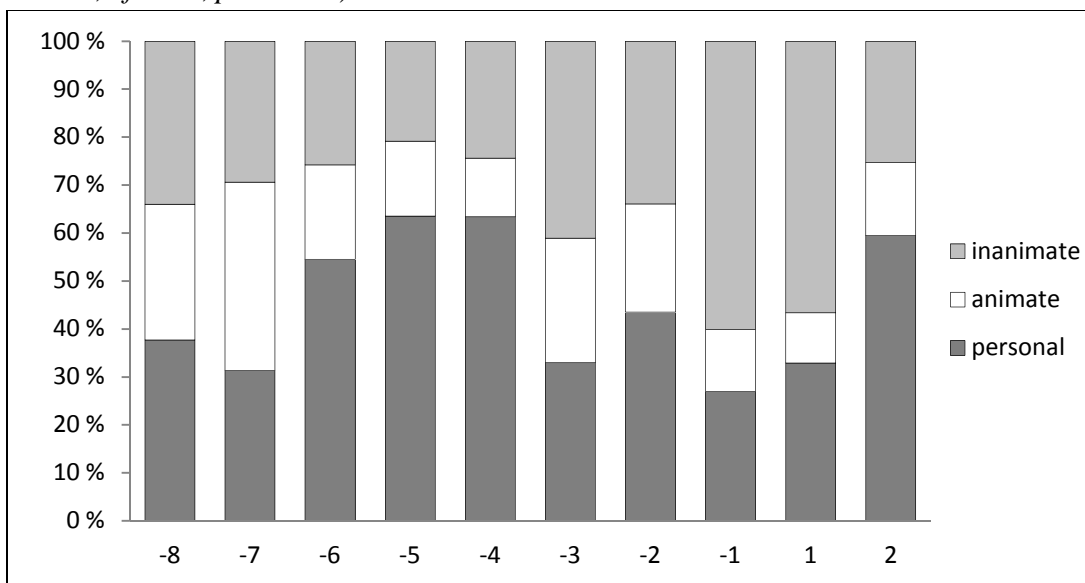
If the A subjects are thought to represent personal name subjects and the S_O subjects inanimate subjects, the trend observed with the construction type resembles rather closely the one observed here with animacy. In other words, animacy is distributed over word positions (at least on the range from -8 to +2) in the same way as the construction type. Regarding linear distance, this means that the semantic properties of the subject NP (animacy class) and those of the verbal construction (construction type) seem to be more or less equivalent. With the subject case selection, the animacy seemed to have a certain supremacy, as was shown in section 4.2.1. and section 4.2.2. From the viewpoint of subject case selection, the here observed trends are, however, rather unexpected. It

might have been expected that the most agentive and most animate subjects would be particularly frequent in the immediate preverbal position where the accusative percentage is as its lowest, as was shown in Figure 5.14. of section 5.3.2.

It also has to be remembered that chapter 4 showed that both construction type and animacy influence subject case selection. Therefore, it would be theoretically justified to ask whether the dependence observed between linear position and construction type in Figure 6.2., as well as the dependence between linear position and animacy class in Figure 6.3., explain the low accusative percentage at positions -2 and -1 (or -8 and -7) in Figure 5.14. (section 5.3.2.). As already mentioned, this does not seem to be the case. On the contrary, the most important position -1 is low in transitive A constructions, which is not compatible with the low accusative percentage (17.8%) of that position (see Figure 5.14.).[326] The high proportion of S_O subjects at position +1 does fit better this assumption (accusative percentage 38.2% in Figure 5.14.), but it is obviously an exception, as will be explained below.

To sum up, the trends in Figure 6.2. form a statistically significant pattern, but this dependence pattern does not match the one that describes the dependence of subject case and position in Figure 5.14. This suggests that the accusative percentage in Figure 5.14. is not completely (or not exclusively) bound to certain construction types (or animacy classes) that, in turn, would be bound to certain word positions. In other words, the construction type, animacy, and the linear distance between the subject and the verb seem, nonetheless, to be at least partly independent of each other. As far as the subject case selection is concerned, it seems that the construction type or animacy cannot be reduced to the word position variable, or vice versa. To conclude, the syntactic factor 'linear position' appears to have an independent (minor) role in subject case selection.

It was said that the high proportion of S_O subjects in position +1 is probably an exception. The high proportion of S_O subjects in position +1 (and -1) derives partly from the fact that most of the occurrences of the verbs *esse* 'to be' and *residere* 'to reside' are located exactly in positions +1 and -1 (54% and 39% of the verbs, respectively). The distribution of *esse* as a function of the linear position can be seen in Appendix 5.9. *Esse*, as a copula, represents apparently the basic linkage between

---

[326] The residual of the A subjects at position -1 is -4.4 (Appendix 5.7.), which suggests that the low accusative percentage is indeed decisive for the statistical significance of the contingency table. The same also applies to the personal name subjects at position -1: their residual is -4.7 (Appendix 5.8.).

item A and item B, hence the smallest possible distance between the copula verb and the subject. Therefore, *esse* is also clearly more frequent in category A1 than in other categories.

By now, it is obvious that several of the independent, i.e. explaining, variables interact. Therefore, it seems that the motivations of subject case selection are not mutually exclusive, but are interrelated. Chapter 4 traced dependences between the accusative subject and animacy and the construction type of the subject. The present chapter has shown that some syntactic factors must be taken into account as well. One possibility is to explain the syntactic and semantic factors as having been conflated in diachrony with each other, as Cennamo (2009) suggests. At this point, Cennamo's statement is again worth quoting:

> The semantic parameters [i.e. animacy and control] conflate, in the course of time, with *syntactic features* and *pragmatic features*. The syntactic features involve the type of clause and the degree of syntactic cohesion between the verb and its argument while the *pragmatic features* involve [--] the topic of the clause gradually being integrated into the predicational nucleus of the clause.[327]

From this viewpoint, it seems logical to interpret the syntactic factors not so much as independent causes of subject case selection than consequences of some higher-level tendency – a tendency that would be the common nominator of both the semantic and syntactic (and pragmatic) explanations. The possibility of this kind of unifying explanation will be discussed below but, first, I want to shed some light on the mutual hierarchy of the variables that I have been using to explain subject case selection. I shall discuss briefly the decision tree that models the dependences between the subject case and certain syntactic and semantic independent variables that can be considered particularly important on the basis of chapter 4 and chapter 5: linear distance, animacy, and construction type. When these three variables are tested on the subject case, the decision tree in Figure 6.4. results. This time only the 3rd declension imparisyllabic subcorpus is examined in order to avoid word classes with ambiguous endings.

---

[327] Cennamo 2009, 327–328. By 'type of clause', Cennamo refers to categories, such as the nominal clause or impersonal construction, but also to more semantic classifications, such as change-of-state clauses or fientive clauses. Pragmatic features are, for example, an extra-syntactic (accusative-form) topic that is little by little integrated into the predicational nucleus. See also Cennamo 2009, 340–341.

Figure 6.4. Decision tree for subject case (dependent variable), animacy, linear distance, and construction type (independent variables).
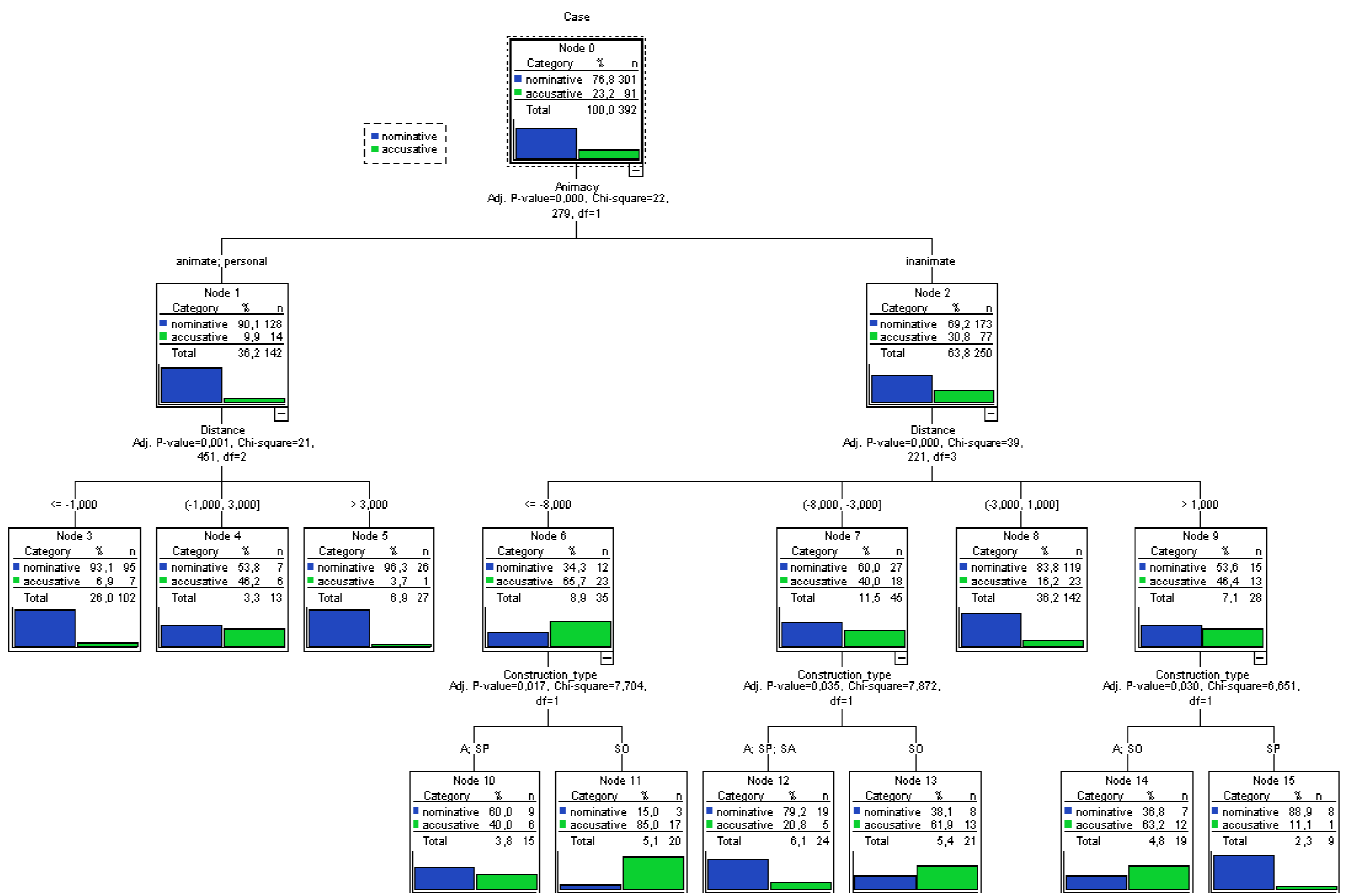
As in Figure 4.4. in section 4.2.2., it is again animacy that effects the first split and contrasts animate and personal name subjects with inanimate subjects. The second branching is, however, something new: linear distance replaces construction type as the second most important variable. The decision tree algorithm classifies the distance variable into statistically significant categories and creates new branches under both animacy nodes ($\chi^2 = 21.45$, $df = 2$, $p = 0.001$ and $\chi^2 = 39.22$, $df = 3$, $p < 0.001$).

The animate/personal name subject branch splits into three statistically significant nodes, and the node with positions -2 to -1 (node 3) has again a very low accusative percentage (6.9%). This time, however, the accusative percentage of node 5 with positions ≥ +4 happens to be even lower (3.7%), but the number is based on one occurrence only (out of 27), whereas node 3 contains 102 occurrences, most of which are at positions -2 to -1. This means that the categories do not behave completely as expected because node 3 with positions -2 to -1 is not the one with the lowest accusative percentage. On the other hand, the special status of positions -2 to -1 is confirmed by the split of the

inanimate subjects (nodes 6 to 9), where the branching is fully predictable: node 8 includes positions -2 and -1 (but also +1) and displays the lowest accusative percentage (16.2%) while all the other nodes show clearly higher percentages (40% to 66%). Thus, the split of node 2 indicates the low accusative percentage of positions -2 to -1, just as was seen earlier in Table 5.6. (and in Appendix 5.5.). The partly unexpected pattern of the animate and personal name subjects remains unexplained for the present.

It is interesting to notice that some of the distance categories of the decision tree (nodes 6, 7, and 9) are further split by the construction type. These categories contain, however, rather few subjects, so the patterns that result cannot be generalised.[328] Node 8, which contains the most occurrences (142) and includes the positions -2 and -1, does not split and the same applies also to the distance nodes 3, 4, and 5 of the animate and personal name branch. In general, the patterns that arise from the distribution of the construction types do not conform to expectations: at node 6, A and S_P align with each other and contrast S_O although S_P and S_O were expected to be aligned. The S_P subjects are, however, very infrequent. The same also happens at node 7, where S_A joins A and S_P. The split at node 9 is the only case where S_P is contrasted with A but, at the same time, also with S_O. As previously mentioned, the frequencies are low and the patterns cannot be generalised to the whole subcorpus data. This again highlights the disadvantages of the decision tree method: the patterns of the lower nodes are only representative of the nodes they split. What is important, however, is the relative order of the splits. In this case, animacy comes first, then distance, and only after that the construction type. This shows that the semantic and syntactic variables are tightly connected in LLCT.

Thus, it can be concluded that although semantics (animacy) is the statistically most significant variable as far as subject case selection is concerned, even syntax (the linear distance between the subject and the verb) heavily affects the subject case. On the basis of the trends of Figure 6.2. and Figure 6.3., both construction type and animacy seem to have a tight relationship with distance. On the other hand, chapter 4 proved that the construction type and animacy are closely related. The statistical methods cannot go further than this, and the conclusions about the causation must be based on theoretical considerations.

---

[328] Note that here, as in all the decision trees, each branching node, i.e. parent node, has to have at least 20 occurrences of subjects and each branch, i.e. child node, at least 7 occurrences (see section 4.2.1.). Therefore, branches can be based on rather few occurrences.

It was stated above that the intransitivity split is semantic by definition and, if syntactic factors play a role in it, they are likely to have become conflated with the semantic properties in the course of time. But it is now time to ask what this 'conflation' really means. There is a huge conceptual difference between whether the syntactic (and pragmatic) factors are *consequences* of the semantic motivations or whether they have just conflated or merged with them in diachrony, so that the syntactic and semantic phenomena seem to be co-patterning in certain contexts. I advocate the conflation interpretation. I suggest that the conflation or merger is perceived as a highly significant dependence between, for example, linear position and construction type in Figure 6.2. and linear position and animacy in Figure 6.3. Nevertheless, when discussing the trends seen in these two figures, it was noticed that the linear subject position does not co-pattern with animacy and the construction type in the critical position -1, where the accusative percentage is at its lowest. There the portion of the agentive A subjects as well as that of the personal name subjects is particularly low contrary to what should be expected were the low accusative percentage directly related to construction type and degree of animacy.[329] In my interpretation, the statistically significant dependences show that the semantic and syntactic factors are likely to have been considerably conflated, but as far as the critical position -1 is concerned, the syntactic linear position variable still has an independent syntactic motivation that cannot be derived from semantic factors, such as animacy or the construction type.

The above conclusion is, perhaps, not surprising considering that it would be, indeed, difficult to explain why the low accusative percentage of the immediate preverbal position would be determined by animacy or construction type. As was seen, Rovai and Cennamo have suggested that semantic control is likely to be the common nominator of subject-inherent properties. In this study, I have shown that this interpretation is also statistically supported. On the other hand, the verb imposes a different kind of 'control', i.e. psycho-syntactic control, on its subject argument by way of agreement and, regardless of the low number of A and personal name subjects, this verb-driven influence is at its highest at the immediate preverbal position where the verb is imminent. Thus, there seem to be two different types of 'control' that affect subject case selection differently as regards the crucial position -1 although in other positions they may be more perfectly conflated, as is suggested by the statistically significant dependences. Of course, it is possible to search for a still higher-level common nominator that would determine both these types of control and could be called loosely the 'cohesion' of the subject/verb structure. This is, however, something that would

---

[329] Moreover, both the A subjects and the personal name subjects display significant residuals at -1 (see Appendix 5.7. and Appendix 5.8.).

need a specifically planned research setting and cannot be realised in this study. Anyway, I am attracted by the idea, albeit without theoretical backing, that the subject case marking may be determined broadly speaking by the overall cohesion of the subject/verb combination – a cohesion which takes into account both the semantic and psycho-syntactic control.

Before ending this section, I shall use a few passages to discuss the formulaicity variable that was introduced in chapter 2. There, the principles of classifying the LLCT sentences into 'free' and 'formulaic' ones were presented. Formulaicity is essentially an extra-linguistic variable, but when it is examined in connection to subject case selection, it may acquire psycholinguistic and even syntactic implications.

Earlier in this chapter, when discussing structural category, I mentioned that recurring formulaic expressions of charter language skew the distributions of several linguistic structures, thus influencing some of the here discussed variables. The idea of the formulaicity degree assessment, as defined in this study, is, however, different: I suppose that the scribes were likely to use more conservative language when writing the formulae than when writing the free or less formulaic parts of charters (see section 1.2.3.). Therefore, it is interesting to find out the possible dependences between the formulaicity degree of the text and the distributions of the linguistic phenomena that are used in it. As for the subject case, this is expressed in hypothesis H_1^7. The corresponding null hypothesis is H_0^7.

H_1^7: The formulaicity degree of the sentence in which the subject occurs is related to the case form of the subject.

H_0^7: The formulaicity degree has no significant effect on the case form of the subject.

Curiously, no statistically significant dependence between formulaicity degree and subject case appears in the basic query set of LLCT or in the subcorpus ($\chi^2 = 1.23$, $df = 1$, $p = 0.268$ and $\chi^2 = 0.39$, $df = 1$, $p = 0.534$, respectively). So, the null hypothesis cannot be rejected. Yet, the formulaicity degree classification does produce meaningful and theory-compatible results with certain other phenomena, such as the increasing use of prepositional phrases with *de* instead of the genitive case in expressing adnominal possession.[330] The percentage of *de* constructions is more than four times

---

[330] E.g. Väänänen 1981, 114.

higher in the free parts (1.28%) than in the formulaic parts (0.29%) of LLCT as proportioned to the total word number.[331] This looks reasonable because the conservative formulaic parts originally did not contain adnominal *de* constructions. The scribes usually remembered rather correctly the traditional genitival possession phrases of the formulaic parts (e.g. *anno regni eius septimo* "on the seventh year of his reign"). With the free parts, however, they did not have models at hand but had to draw on their own language competence, which made recourse to the spoken language of the time. Therefore, relatively more colloquial structures, such as possessive *de* phrases (e.g. *curtem de casa* "courtyard of the house"), infiltrated the free parts than in the formulaic parts.

I have summarised the *de* example here in order to point out how essentially it differs from the subject case endings. A *de* construction consists of three words while a genitival possessive construction consists of two. Hence, the difference between these constructions is far more substantial than that between, say, the endings -*o*,- *u*(*m*), and -*us*. I claim that it is easier to both recognise and memorise the difference between an NP + NP construction and an NP + PP construction than between an NP with an *x* ending and an NP with a *y* ending, especially when these endings belong to abstract syntactic cases, not to semantic cases that convey concrete meanings. Moreover, the classical genitive was a moribund grammatical category at the time of the writing of LLCT charters. It was likely to be a prestige form that the scribes learnt (or not) during their education. The distinction between the syntactic cases, nominative and accusative, on the other hand, hardly allows for this kind of attribution of prestige status.

Although the ancient grammatical tradition knew to distinguish between verb, subject, and object, I doubt whether the scribes were ever exposed to the idea of recognising grammatical entities, such as the subject. Of course, scribes with more intimacy with traditionally written (ecclesiastical) Latin would be expected to have induced from these texts some more or less subconscious rules that they then also applied when writing charters. It is, however, improbable that they would have chosen the subject case form according to some theoretical L2-type knowledge. Therefore, I argue that at least every alignment pattern that arises from the charter texts and differs from the Classical Latin nominative/accusative alignment is likely to reflect the real state of the system. All this is also interestingly related to the question about how the scribes produced the charters: the attestation of a morphosyntactic alignment that cannot have been learnt at school shows that the alignment comes

---

[331] Korkiakangas & Lassila 2013, 70. Note that instead of dividing entire sentences into 'free' and 'formulaic' ones, it would be more exact to analyse shorter phrases because the non-formulaicity slots seldom comprise merely a couple of words and because some formulaicity-sensitive linguistic phenomena might manifest themselves better on a higher granularity level. A more detailed analysis was considered, however, too time-consuming.

from the spoken language, a fact that proves that the scribes did not copy the charters from written models but, in all likelihood, reproduced the formulaic expressions by heart according to their own consideration.

6.2. The accusative as the default case

In this chapter, I shall examine how the central findings of chapter 5 and chapter 6 are linked to the idea of the default case that was originally introduced in section 3.1. My key thesis that I have been applying throughout the preceding chapters is that when the subject/verb cohesion weakens, the subject sometimes ends up in a "wrong" case form. What is essential is that in Late Latin this "wrong" form, which the subjects that withdraw from verbal control slip into, was the accusative (or the accusative-based form), and no longer the nominative, as proved for earlier Latin by Adams (2013).[332]

In section 3.4., I stated that many asyntactic or allegedly isolated occurrences of the accusative subject that have been discussed in linguistic studies, such as anacoluthon or the 'exclamatory' accusative, may be symptoms of a large-scale linguistic change that led to the defaultisation of the accusative: when the Late Latin scribes slipped, they more and more often unconsciously ended up with that case form. From this point of view, it is not essential whether this form really was a genuine accusative with the original *-m* ending or just a morphophonological descendant, such as *filio* or *medietate*.[333] What is important, instead, is that it is a form in contrast with the nominative, such as *filius* and *medietas*. So, it is obvious that any hypothesis concerning the Late Latin default case stands or falls on the successful demonstration of the existence of a nominative/accusative contrast, not necessarily for the entire nominal declension but at least for some parts of it.

In literature, there are allusions to the default or "neutral" status of the Latin accusative or to its shift toward the unmarked case. These intellectually attractive hypotheses have suffered from the scarcity of appropriate empirical evidence. Moreover, it has not always been clear whether they are meant to cover the whole time span of Latinity or only the later phases, as is proposed here. To my mind, Adams (2013) shows rather convincingly that it is not justified to extend the functional unmarkedness of the accusative to the Latin of classical or imperial periods. Consequently, it seems to be wise to apply the unmarkedness hypotheses to later Latin only. Indeed, several scholars with a

---

[332] Adams 2013, 234–256.
[333] La Fauci 2001, 22.

diachronic approach seem to assume that the markedness change was gradual and that the process took place somewhere between the Imperial and the Late Latin periods. They do not, however, necessarily state this explicitly.[334] For changes of this kind, it is obviously impossible to define exact dates.

By default form, I mean here the form that "occurs when there are no obvious criteria for selecting a particular item". The default case is, thus, a case that occurs in settings where it is not clear which case ought to be used.[335] It can be asked what a default case has to do with the subject, whose case is, in theory, directly imposed by the verbal agreement. As explained above, my argument is that where the cohesion of the verbal nucleus is weakened, the case form which appears as the default form is unconsciously applied even to the subject.

Pensado (1986) discusses the markedness change in Latin through a comparison with Old French. Cross-linguistically, the syntactically unmarked case does not usually have a morphological marker. In Latin (and in Indo-European, in general), both the nominative and the accusative, the two syntactic cases, are morphologically marked in the major part of the declension (e.g. Latin *pars* (< *part-s*) nominative, *partem* (< *part-em*) accusative 'part') although the nominative should have no marker in the canonical nominative/accusative alignment. By the time of the bicasual system of Old French, the Latin accusative marking *-m* had disappeared in the *cas régime* and only the nominative marker *-s* remained in the *cas sujet*. Even this nominative morpheme, now marked, was retained only in certain inflexional paradigms and mainly with the nouns that ranked the highest on the animacy hierarchy (see section 4.1.1.). By the late Middle Ages, the accusative-based unmarked form spread finally even to the *cas sujet* but, at that moment, the French word order had already become fixed and assumed all the functions of the case system.[336] All this suggests that a semantically driven

---

[334] Collinge 1978, 623–624; Lehmann 1985, 246–247; Rovai 2012b, 107, 110; Smith 2011, 277–278; Pieroni 1999, 120; Galdi 2013, 77. Instead, Cennamo and La Fauci interpret the accusative as the default case even in early and Classical Latin. This stance is based on the fact that there is evidence of the occurrence of patterns of active alignment even in early Latin (see examples of the accusative as topics and in exclamations and lists in section 3.2.). Cennamo 2001b, 20–21; Cennamo 2009, 308–309, 327; La Fauci 1988, 55; La Fauci 2001, 21–22; Vincent 1982, 88–89. For a diametrically opposed views, see Adams 2013, 254–256; Benucci 2004. It is not possible to examine this complicated question sufficiently in this work. Hence, I refer to Korkiakangas (in press) and the references therein for a more in-depth discussion.
[335] Blake 2001, 199; Smith 2011, 278.
[336] Pensado 1986, 271–274; Schøsler 2001, 169–179; Schøsler 1984, 122; Smith 2011, 281–289; Lehmann 1985, 244–246; Sornicola 2011, 18–31. In some (southern) Romance languages, the process continued and gave birth to the prepositional object. Once the accusative-based form had finally occupied all the subjects, the again strengthened nominative/accusative tendency of the morphosyntactic alignment created a new contrast with an unmarked subject and a marked prepositional object for high-animacy nouns. Even the rise of the Romance definite article, which first appeared with subjects, and the object clitic seem to be a means of accentuating this contrast (Vincent 1997, 163; Ledgeway 2012, 350–351; Sornicola 2011, 36–42; Zamboni 1998a, 130–131).

morphosyntactic realignment took place also in the Latin of Gaul although the case opposition persisted there centuries later than in Italy. As the written Old French material is relatively substantial, the picture sketched by Pensado can be considered accurate. This picture has probably very much in common with the much more obscure picture of Italo-Romance developments.

The syntactic and extra-linguistic phenomena that have been discussed in chapter 5 seem to leave no doubt about the accusative/accusative-based form being the default case form in the Latin of LLCT. Of course, they do not tell anything about the actual time span when the defaultisation of the accusative took place – that it must have happened before the time of LLCT is no surprise. The preverbal position -1 (and -2) is a position where the verbal control and agreement are at their strongest in an SVO language. This is why the nominative form has the highest frequency at that very position. Elsewhere, the verbal control is more weakly felt and, as a consequence, the subject slips more easily into the accusative, i.e. the default case. The unmarkedness of the immediate preverbal position is further underlined by its being favoured by structural category A1 (Figure 5.15.), which turns out to be unmarked in many ways.

Similarly, section 5.1. and section 5.2. showed that the relative position within the subject NP correlates with the case form better when the position is defined within the dependency structure and not as a linear sequence. The further off the orbit of its head the attribute is located, the more easily it occurs in the accusative. What is common to this and the above phenomenon is the following: assuming that a certain syntactic element determines certain morphological features of its dependent, the longer the syntactic distance between the two is, the less successful the outcome of this determination.[337]

6.3. Summary

Chapter 6 concentrated on the interaction of various semantic and syntactic variables that had been discussed in the preceding chapters. It was shown that semantics and syntax are interrelated in a most intricate manner. Linear position with respect to the verb (syntactic variable) seems to depend on the construction type (semantic variable), which is somehow related to the structural category (syntactic and/or extra-linguistic variable).

---

[337] This rather intuitive hypothesis is complicated as far as the subject NP is concerned because the NP head, with which the modifiers agree in case, receives its case through its syntactic/semantic relation to the verb, not through agreement. The two ways of determination are different.

Section 6.1. began by examining some word-order-related phenomena on the sentence level. It was seen that the relative position of the subject, i.e. whether it is preverbal or postverbal, is related to the construction type in which the subject occurs. With the exception of the passive, the LLCT data seem to support the view that, in the Late Latin SV(O) order, S and O are to be understood more broadly as A/S_A and O/S_O, respectively. In other words, actor and undergoer subjects occupy different positions according to the semantically-based alignment. In LLCT, the semantically-based alignment is, thus, not only visible in the case marking, but also in the linearisation of the elements.

After that, the relationship between the construction type, the structural category (A1, A2–4, B1–2), and the relative subject position was discussed. It was shown already in chapter 4 that category A1 can be viewed as the structural category with the least processing cost. Because of the large size (981 occurrences, i.e. 62.3% of the LLCT subjects) and the considerable verb lemma variation of A1, the distracting influence of formulaicity is not likely to skew as much the linguistic data extracted from A1 as those extracted from other structural categories. On the other hand, in section 6.1., it turned out that there is a statistically significant dependence between the construction type and the structural category. Thus, the supposed influence of the structural category on the subject case selection may be partly due to the construction type and not to syntactic complexity. This does not, however, undo the fact that category A1 is likely to be the unmarked category that suits particularly well the needs of corpus linguistic analysis.

Section 6.1. also revealed that formulaic expressions typical of charter language are unevenly distributed across word positions, which is perceived as frequency peaks of certain construction types and animacy classes at certain word positions. It is noteworthy that animacy is distributed over linear word positions essentially in the same way as the construction type. This means that, at least as regards linear distance, the semantic properties of the subject NP (animacy class) and those of the verbal construction (construction type) seem to be more or less equivalent. The distributions of the construction type and animacy degree as a function of linear distance do not, however, correspond to the pattern of the subject case as a function of linear distance as far as the critical position -1 is concerned. This suggests that the construction type and animacy cannot be reduced to the word position variable, or vice versa, although the variables are clearly related.

The decision tree in Figure 6.4. showed that, as concerns the statistical significance order, animacy comes first and distance only after that. This is in concord with the postulated semantic basis of the

intransitivity split. On the other hand, the syntactic position variable may have been conflated with the semantic variables, as suggested by Cennamo (2009). It is probable that the psycho-syntactic motivation of the linear position variable becomes evident precisely at this unmarked, neutral, canonical subject position -1, where the cohesion of the verbal nucleus is at its highest and the retention of the nominative, hence, most successful. This verb-driven 'control' is conceptually different from the control that Cennamo and Rovai have proposed to be the unifying factor of the semantic variables in the semantically-based alignment. I suggested, however, that both these types of control may be perceived as *per se* independent elements that contribute to the general cohesion of the subject/verb combination.

A central premise of the syntactic analysis in chapter 5 and chapter 6 has been the assumption of the accusative or accusative-based form as the default case of Late Latin. The subject is supposed to slip into the default case in syntactically non-prototypical conditions. Indeed, the findings of the previous chapters give substantial support for the view that the markedness change between the nominative and the accusative had been completed by the time of the LLCT charters. Section 6.2. suggested that anacoluthon-based or 'exclamatory' accusative subjects attested in the earlier stages of Latinity are likely to have been precocious symptoms of a large-scale linguistic change that contributed to the defaultisation of the accusative in Late Latin. Of course, it is reasonable to speak of a default case only within a system like that of LLCT, where a nominative/accusative contrast still exists (on the basis of the findings in chapter 4), not necessarily in the entire nominal declension, but at least in parts of it.

To close this chapter, I will briefly summarise the dependences that have been detected between the variables under examination. The dependent/independent variable pairs with statistically significant dependences (either in the entire LLCT or in the 3rd declension subcorpus) are listed in the following (cf. Table 6.1.):

- the case form of the subject and the relative subject position (preverbal/postverbal) (H_1^1),
- the relative subject position (preverbal/postverbal) and the construction type (H_1^2),
- the construction type and the structural category (H_1^3),
- the structural category and the relative subject position (preverbal/postverbal) (H_1^4),
- the construction type and the subject position as a function of the linear distance from the verb (H_1^5),

- the animacy class and the subject position as a function of the linear distance from the verb (H_1^6).

7. Conclusions

In section 1.1., I posed the following research questions: Does the evidence of LLCT reflect a semantically-based morphosyntactic alignment? If so, is the subject case selection dependent on semantic factors or do syntactic factors also play a role? What are the most important semantic and syntactic factors? To what extent do these factors interact with each other? Can transitivity degree be used as a measure of subject case selection? Are there also some extra-linguistic factors at play? Are these extra-linguistic factors based on formulaicity or are they psycho-syntactic by nature? I also promised to provide an overview of the conditions that determine the usability of charter Latin in studying linguistic change and variation in the spoken language.

Applying statistical methods to the annotated corpus data of LLCT was an essential part of the study. The above research questions were transformed into testable hypotheses in order to detect significant dependences between subject case and relevant semantic and syntactic variables. When the statistically significant percentage distributions of the independent variables were counted out of the contingency tables, the following was observed:

1) Accusative subjects prefer low-animacy nouns.
2) Accusative subjects often occur with unaccusative verbs.
3) Accusative subjects occur in clauses with low transitivity degree as defined by Hopper & Thompson's transitivity scale.
4) Accusative subjects seem to occur more often in deep, i.e. complex, coordination structures than in simple, non-coordinated structures.
5) Contiguity of attribute to subject NP head implies a higher retention of the nominative than remoteness from the NP head.
6) Immediate preverbal clausal position of the subject implies a high retention of the nominative.
7) The subject position within a sentence is related to animacy and construction type, but cannot be deduced from them.

The following thematic sections explicate the above findings and discuss their implications. The semantic factors obviously occupy most space. Although all the relevant observations are not quantitative by nature, I shall discuss them in apposite places among the results of the statistical analysis. At the end, one section outlines some future perspectives of Late Latin alignment studies.

Semantically-based alignment

The most important result of this study is its support for the theory that predicts a partial semantically-based morphosyntactic alignment in Late Latin. According to this theory, the nominative/accusative contrast was (re)semanticised in Late Latin so that the nominative came to encode all the agent-like arguments and the accusative all the patient-like arguments. The logical corollary – or premise – of the result is that there was still a functional, albeit reduced nominative/accusative contrast between the nominative and the accusative in the Latin of Tuscany in the 8th and 9th centuries – not in the whole declension but in parts of it. A semantically-aligned case marking system that was essentially alien to Classical Latin, i.e. the language taught to the scribes, cannot be interpreted as a learned relic, a term that is sometimes used to describe the case use of charter Latin. On the other hand, the case distribution pattern of LLCT is not entirely determined by semantics: the linear position of the subject within a sentence and the linear position of an attribute within the subject NP must be taken into account.

While the above dependences 1) and 2) suggest the presence of semantically-based case marking in LLCT, a Romance-type neutral system and a Classical Latin nominative/accusative aligned opposition are attested as well. Most of the extended accusatives of LLCT match the chronological model that Rovai (2005) and Cennamo (2009) outlined on the basis of previous evidence. Accordingly, the low-animacy subjects of LLCT occur more often in the accusative than the agentive high-animacy subjects. Given that animacy is tightly connected with construction type, the unaccusative and passive constructions with S_O subjects occur more often in the accusative than the unergative and transitive constructions with S_A and A subjects, respectively (section 4.2.1. and section 4.2.2.). The systemic case marking split appears, thus, between the low and high animacy and, on the other hand, between the two intransitive arguments S_O and S_A. As suggested by Rovai, the common nominator can be the control that the subject exercises over the verbal process.

It has to be remembered that this study examined only noun and adjective subjects. Pronominal subjects were left out because they have irregular paradigms. Especially personal pronouns are often animate and highly agentive and, thus, usually appear in the nominative. That is to say that pronominal subjects follow the nominative/accusative alignment as suggested by the theory of semantic alignment. Note, however, that the semantics of the subject NP is only the other side of the coin: the evidence of LLCT shows that the verb-motivated construction type does correlate with the subject case selection almost on a par with the subject-inherent feature, i.e. animacy (section 4.2.1.).

As earlier mentioned, the attested case contrast is only likely to have existed in part of the declension. The syncretistic nominative/accusative forms as well as the 1st declension singulars were excluded from the analysis beforehand. It became obvious that the 2nd declension singular *-us/um/o* and the 3rd declension parisyllabic *-is/em/e* have also lost case contrast for the most part – their residues may be learned relics. In order to avoid these neutralised categories and, especially, the frequent 2nd declension singular personal names in *-o*, I sometimes utilised a subcorpus of 3rd declension singular imparisyllabic subjects (section 4.2.2.). As the imparisyllabic nouns retain the morphophonological case contrast, the subcorpus is supposed to provide the most reliable results concerning the case system. The results of the subcorpus are, indeed, also the most theory-compatible.

When analysing corpus data, it is necessary to define what kind of distributional patterns are considered theory-compatible. Quantitative corpus study of Late Latin has not been possible before LLCT. Therefore, the question concerning the expected distribution of accusative subjects has not come up earlier. There are no predictions about the nominative and accusative percentages in different animacy or construction type categories in a corpus that is supposed to feature an ongoing morphosyntactic realignment. I decided to align the size order of the accusative percentage with the advancedness of the realignment at a given time point: the larger the observed accusative percentage, the earlier the extended accusative is supposed to have penetrated that category (section 4.2.1.). Although highly formulaic and conservative written texts obviously do not reflect the developments of the spoken language as closely as would be desired, the findings of LLCT match rather well the hypothesis that the extension of the accusative begins from the low-animacy domains with low-control subjects.

As this study does not pursue a chronological approach, it can only be concluded that, based on a synchronic analysis of the LLCT data, the accusative subject distribution suggests a rather advanced – but not the very last – stage of realignment. Given that the S_A subjects of unergative constructions and even the A subjects of transitive constructions display accusatives, the alignment attested in LLCT cannot be described as 'purely' semantically or active/inactive aligned. Although the main split is obviously semantically motivated, the alignment ought to be seen as a continuum: some subjects (such as personal pronouns and agentive nouns) represent the nominative/accusative alignment, some are clearly active/inactive aligned, some probably ergative/absolutive aligned, and some others seem to reflect an already neutralised case system. Note, however, that some of the accusa-

tive-form A subjects of LLCT appear to be very low in transitivity, which may explain the already materialised extension of accusative into A (section 4.3.2.). A detailed diachronic analysis of the LLCT data will be postponed to a future study. In any case, the current study suggests that the neutralisation of the Latin case system was a prolonged slide across the alignment continuum and took place over centuries.

Transitivity component analysis

In section 4.3., I tested whether Hopper and Thompson's (1980) multifactorial transitivity scale would provide a useful tool for predicting subject case selection in LLCT. It turned out that Hopper and Thompson's transitivity scale does not give more significant dependences or better predictions on the accusative percentage of subjects in LLCT than the animacy degree or the fourfold construction type classification (A, S_A, S_O, S_O passive). The transitivity analysis, however, confirmed that accusative subjects occur typically in clauses with low transitivity degree. The analysis also yielded some important observations: first, the nearly similar accusative percentages of the A and S_A subjects of LLCT seem to be related to their almost equal transitivity degrees (section 4.3.2.). Second, using the fourfold construction type classification as a measure of subject case selection is problematic because the different construction types are located discontinuously on the transitivity scale. Third, the verbal panorama of LLCT is one of a rather low transitivity: even the syntactically transitive clauses rank low on Hopper and Thompson's scale (section 4.3.3.). As a byproduct, the transitivity component analysis provided the first results concerning the overall transitivity degree of Latin legal genre, which appeared to be particularly low (section 4.4.1.). It can be concluded that charter Latin is not an optimal material for studying case alignment because its transitivity degree variation is so narrow.

Syntactic factors

The above points 4) to 6) are related to dependences between the subject case and syntactic variables. Although intransitivity splits originate from semantics by default, three previously unexplored syntactic factors seem to influence the subject case selection in LLCT: depth of coordination structure (section 5.1.), position within subject NP (section 5.2.) and position within clause (section 5.3.2.). It is important to isolate these factors so that their influence is not ascribed erroneously to the semantic factors. The two phenomena that are related to word position allow a rather painless interpretation: the attributes located at the end of attribute chains have slightly higher accusative

shares than the attributes that are closer to the head of the subject NP. Similarly, the immediate pre-verbal clausal position of the subject implies a high retention of the nominative. The first-mentioned variable, i.e. the depth of coordination, is more problematic because it may be determined by formulaicity and construction type, which is a semantic variable.

The syntactic findings need to be related to the well-grounded assumption that the accusative had become the default case in Late Latin, i.e. the case form that occurs in settings where there are no obvious criteria for selecting a particular case. Moreover, the explanation draws on the psycholinguistic model of syntactic complexity and dependency lengths: at the beginning of chapter 5, I adopt the view that mental processing difficulty is essentially memory-related and reflects dependency lengths. This dependency length can be measured in words between the head and the dependent. I assume that the longer the distance between the subject NP head and its attribute or between the subject and the verb is, the more hesitation there is about the required case and the more easily the subject slips into the default case, i.e. the accusative. In contrast, the nominative prevails in the most canonical subject positions, i.e. the immediate preverbal position (in SVO languages) (section 5.2.) and the closest-to-head attribute position section 5.3.2.). This can be seen as evidence of a tendency to maintain the cohesion of the verbal nucleus, i.e. to keep the dependency lengths short, by an immediate realisation of the subject argument before the verbal head.

As has been pointed out, complex structures with coordinated subjects, which are highly frequent in LLCT, are problematic. The depth of these coordination structures was utilised as a measure of syntactic complexity in section 5.1. There appeared, indeed, to be a partial statistically significant dependence between the structural (coordination) category and the case form of the subject: the accusative percentage is higher with certain coordinated subjects than with normal, non-coordinated subjects. The reliability of this dependence was lessened, however, by the fact that section 6.1. revealed a high interaction of structural category and construction type, which may explain the dependence between the structural category and subject case selection. The cause of this interaction is likely to be the particularly high formulaicity of coordinated structures that skews the distributions of construction types in certain coordination categories.

The driving forces of case marking in LLCT

Chapter 6 concentrated on the interaction between the above-discussed semantic and syntactic variables. The above point 7) presents the most important finding: linear subject position within a sen-

tence displays a statistically significant dependence with both animacy and construction type. The subject position cannot be, however, directly deduced from these semantic variables, i.e. animacy and the construction type. This is because the distributions of the construction type and animacy degree as a function of the linear distance do not correspond to the subject case pattern that displays the lowest accusative percentage at the immediate preverbal position (-1) (section 6.1.).

Another central observation of section 6.1. is that when the statistical significance order of the semantic and syntactic variables is surveyed, animacy comes first and distance only after that. This is in concord with the postulated semantic basis of the intransitivity split. On the other hand, the syntactic position may have been conflated with the semantic variables, as suggested by Cennamo (2009). Although the subject position is not deducible from semantic variables, it is probable that the psycho-syntactic coherence required for producing the nominative form is focalised precisely at the unmarked, immediate preverbal subject position. This verb-driven 'control' differs from the control that has been proposed to be the unifying factor of the semantic variables in the semantically-based alignment. Yet, although both these types of control are independent, they may contribute to the general cohesion of the subject/verb combination.

Thus, this study identified another type of force that seems to affect subject case selection in LLCT, i.e. the psycho-syntactic cohesion of the verbal nucleus. This force is conceptually different from the semantic factors of subject case selection, but may still be partly conflated with them owing to systematic co-occurrence. As a consequence, it is necessary to recognise that the subject case selection in texts like LLCT cannot be exhaustively analysed if the effect of linear position is not taken into account.

Charter Latin and linguistic studies

On the whole, the charters appear to provide a privileged overview on the last stages of the early medieval Latin case system. It is true that the innate formulaicity of the charter genre causes repetition that misleads the interpretation easily if the special character of charter Latin is not accounted for. Formulaicity is essentially an extra-linguistic factor, but when it is examined in connection to subject case selection, it acquires psycholinguistic and even syntactic implications. Another challenge for interpretation is the non-standard use of case endings. Nevertheless, the meaningful statistical results of this study prove that when the stumbling blocks of formulaicity and non-standard forms are properly recognised and their influence minimised, charter texts can be utilised to study

the linguistic variation and change that took place in the spoken language: the scribes wrote from memory and let the non-standard forms creep in. The correct treatment of charter texts requires sensitivity for the textual particularity of charter texts and comprehensive knowledge on the evolution of Latin. In sum, it requires a seamless combination of philological and linguistic approaches.

Future perspectives

This study has shown that charter material can be used for corpus-linguistic study after careful philological preprocessing. As a consequence, a variety of interesting linguistic and extra-linguistic questions concerning charter Latin and Late Latin in general can be studied in the LLCT data. In the following, I point out some future research subjects that I find important and that can be studied with charter data.

As I mentioned earlier, the present study ought to be complemented by analysing the diachronic, diastratic, and diatopic variables and their influence on the subject case selection as well as on other linguistic phenomena that may reflect the state of the spoken language. It is also possible to annotate additional charter material that broadens the chronological and/or areal scope of the present research setting. To my mind, the diachronic examination of alignment change in charter Latin is the most urgent task. It would serve best ongoing research on the Latin morphosyntactic alignment, which has been lacking comprehensive evidence concerning the ultimate stages of the case system.

There is also much to do with analysing other alignment-related issues, such as the reorganisation of voice and agreement patterns within the verbal system in Late Latin. The charter data may be of great help in detecting the interactions between the realignment outcomes in case marking and in the verbal system. Another field, in which LLCT can give its contribution to general linguistics, is genre studies: the transitivity of Latin textual genres (charter language, historical narrative, speeches, dialogues, etc.) would be worth examining in a future corpus study.

A general quantitative overview of the non-standard features of charter Latin would also be a desideratum. It could serve as the basis for several future studies. The present work concentrated on the two syntactic cases, but a complete picture of the Late Latin case system requires finding out how the other cases, i.e. mainly the genitive/dative, were used. This kind of examination would obviously include even the prepositional phrases.

Charter material also opens up new perspectives on other, related fields of study. A charter corpus, such as LLCT, can be the basis of a 'sociolinguistic' model that examines how charter texts were produced in their diplomatic and socio-historical context. Particularly, the psycholinguistic conditions of writing from memory would deserve attention: it is a thin line between memory lapse and psycho-syntactic performance preference, and the rather well contextualised charters might possibly provide evidence that helps to understand the fundamentally similar procedure of copying manuscripts. On the other hand, charters can be utilised not only to study real language change, but also to investigate the learned language skills and, thus, the language attitudes of the scribes. For example, the scribes' preference for certain defunctionalised prestige forms sometimes shows as the constant or increasing frequency of those forms used both correctly and hypercorrectly.

In general, using annotated corpus material will contribute to establishing Late Latin linguistics on a new empirical and quantitative foundation.

Bibliography

Adams 1976a = Adams, J.N. 1976. 'A typological approach to Latin word order', in *Indogermanische Forschungen* 81, 70–99.

Adams 1976b = Adams, J.N. 1976. *The Text and Language of a Vulgar Latin Chronicle (Anonymus Valesianus II)*. Institute of Classical studies: London. Bulletin supplement 36.

Adams 2005 = Adams, J.N. 2005. 'The accusative + infinitive and dependent *quod-*/*quia-* clauses', in Kiss, S., Mondin, L. & Salvi, G. (eds.). *Latin et langues romanes: études de linguistique offertes à József Herman à l'occasion de son 80ème anniversaire*, 194–206. Tübingen: Max Niemeyer Verlag.

Adams 2007 = Adams, J.N. 2007. *The Regional Diversification of Latin 200 BC–AD 600*. Cambridge: Cambridge University Press.

Adams 2013 = Adams, J.N. 2013. *Social variation and the Latin language*. Cambridge: Cambridge University Press.

Agresti 2007 = Agresti, A. 2007. *An introduction to categorical data analysis* (2nd ed.). Hoboken & Chichester: Wiley Interscience.

Ahearn 2010 = Ahearn, L.M. 2010. 'Agency and language', in Jaspers, J., Östman, J.-O. & Verschueren, J. (eds.) *Society and Language* Use, 28–48. John Benjamins: Amsterdam & Philadelphia.

Almas & Berti 2013 = Almas, B. & Berti, M. 2013. 'The Linked Fragment: TEI and the Encoding of Text Re-uses of Lost Authors', in *The Linked TEI: Text Encoding in the Web*. TEI Conference and Members Meeting 2013.

Amelotti & Costamagna 1975 = Amelotti, M. & Costamagna, G. 1975. *Alle origini del Notariato italiano*. Studi storici sul Notariato italiano 2. Roma.

Audollent 1967 = Audollent, A. 1967. *Defixionum tabellae quotquot innotuerunt tam in Graecis Orientis, quam in totius Occidentis partibus praeter Atticas in Corpore Inscriptionum Atticarum editas*. Minerva: Frankfurt am Main (Unveränderter Nachdruck, 1904).

Auernheimer 2003 = Auernheimer, B. 2003. *Die Sprachplanung der karolingischen Bildungsreform im Spiegel von Heiligenviten*. K.G. Saur: München & Leipzig.

Austin 1962 = Austin, J.L. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.

Azzara & Gasparri 2005 = Azzara, C. & Gasparri, S. 2005. *Le leggi dei longobardi: storia, memoria e diritto di un popolo germanico*. Viella: Roma.

Bamman & al. 2007a = Bamman, D., Passarotti, M., Crane, G. & Raynaud, S. 2007. 'A Collaborative Model of Treebank Development', in *Proceedings of the Sixth International Workshop on*

Treebanks and Linguistic Theories (TLT2007), Bergen, 1–6. (http://tlt07.uib.no/papers/4.pdf, accessed 8 Apr 2015)

Bamman & al. 2007b = Bamman, D., Passarotti, M., Crane, G. & Raynaud, S. 2007. *Guidelines for the Syntactic Annotation of Latin Treebanks* (v. 1.3). (http://nlp.perseus.tufts.edu/syntax/treebank/ldt/1.5/docs/ guidelines.pdf, accessed 7 Apr 2015)

Bamman & Crane 2008 = Bamman, D. & Crane, G. 2008. 'Building a Dynamic Lexicon from a Digital Library', in *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (JCDL), Pittsburgh, 2008, 11–20. ACM: New York.

Banniard 1992 = Banniard, M. 1992. *Viva voce: communication écrite et communication orale du IVe aux IXe siècle en Occident latin*. Paris: Institut des Études Augustiniennes.

Bartoli Langeli 2006 = Bartoli Langeli, A. 2006. *Notai: scrivere documenti nell'Italia medievale*. Viella: Roma.

Bastardas Parera 1953 = Bastardas Parera, J. 1953. *Particularidades sintácticas del latín Medieval: Cartularios Españoles de los siglos VIII al XI*. Escuela de Filologia: Barcelona & Madrid.

Bauer 2008 = Bauer, B. 2008. 'Nominal Apposition in Vulgar and Late Latin. At the cross-roads of major language changes', in Wright, R. (ed.). *Latin vulgaire – latin tardif* VIII, Oxford, 2006, 42–50. Olms-Weidmann: Hildesheim.

Bauer 2009 = Bauer, B. 'Word order', in Baldi, P. & Cuzzolin, P. (eds.). *New Perspectives on Historical Latin Syntax* 1: Syntax of the Sentence, 214–316. Mouton de Gruyter: Berlin & New York.

Beaman 1984 = Beaman, K. 1984. 'Coordination and subordination revisited: syntactic complexity in spoken and written narrative discourse', in Tannen, D. & Freedle, R. (eds.). *Coherence is spoken and written discourse*, 44–80. Norwood: Ablex.

Beckmann 1963 = Beckmann, G. 1963. *Die Nachfolgekonstruktionen des instrumentalen Ablativs im Spätlatein und im Französischen*. Tübingen: Max Niemeyer.

Benucci 2004 = Benucci, F. 2004. 'Nominativo e accusativo nelle lingue dell'Italia antica diverse dal latino', in *Linguistica e Filologia* 18, 7–60.

Blaise 1975 = Blaise, A. 1975. *Dictionnaire latin-français des auteurs du moyen-age. Lexicon Latinitatis Medii Aevi praesertim ad res ecclesiasticas investigandas pertinens*. Brepols: Turnholti.

Blake 2001 = Blake, B.J. 2001. *Case*. 2nd edition. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Bolkestein 1976 = Bolkestein, M. 1976. 'Acc.I. and *ut*-clauses with *verba dicendi* in Latin', in *Glotta* 54, 263–291.

Bolkestein 1996 = Bolkestein, M. 1996. 'Free but not arbitrary: "emotive" word order in Latin?', in Risselada, R., De Jong, J.R. & Bolkestein, A.M. (eds.). *On Latin. Linguistic and literary studies in honour of Harm Pinkster*, 7–24. J.C. Gieben: Amsterdam.

Bonnet 1890 = Bonnet, M. 1890. *Le latin de Grégoire de Tours*. Georg Olms Verlagsbuchhandlung: Hildesheim (Reprografischer Nachdruck, 1968).

Bortolussi 2013 = Bortolussi, B. 2013. 'L'ordre des constituants en poesie: verbe initial chez Ovide', *hand-out* at the 17th International Colloquium of Latin Linguistics, Rome 2013.

Boyland 2001 = Boyland, J.T. 2001. 'Hypercorrect pronoun case in English? Cognitive processes that account for pronoun usage', in Bybee, J.L. and Hopper, P.J. (eds.). *Frequency and the Emergence of Linguistic Structure*, 383–404. John Benjamins: Amsterdam.

Bresslau 1958 = Bresslau, H. 1958. *Handbuch der Urkundenlehre für Deutschland und Italien*. Band 1. Walter de Gruyter: Berlin.

Bresslau & Klewitz 1958 = Bresslau, H. & Klewitz, H. 1958. *Handbuch der Urkundenlehre für Deutschland und Italien*. Band 2. Walter de Gruyter: Berlin.

Buridant 2000 = Buridant, C. 2000. *Grammaire nouvelle de l'ancien français*. Sedes: Paris.

Bybee & Hopper 2001 = Bybee, J. & Hopper, P. 2001. 'Introduction to frequency and the emergence of linguistic structure', in Bybee, J.L. and Hopper, P.J. (eds.). *Frequency and the Emergence of Linguistic Structure*, 1–26. John Benjamins: Amsterdam.

Capo 1990 = Capo, L. 1990. 'Paolo Diacono e il problema della cultura', in Cammarosano, P. & Gasparri, S. (eds.). *Langobardia*, 169–235. Udine.

Caprioli 1978 = Caprioli, S. 1978. 'Satura Lanx 11. Per Liutprando 91', in *Studi in memoria di Giuliana d'Amelio*. Vol. 1: Studi storico-giuridici, 203–217. Milano: A. Giuffrè.

CDT = *Codice Diplomatico Toscano*. Parte 2, tomo 1. A cura di Filippo Brunetti. Firenze 1833. (http://nlp.perseus.tufts.edu/hopper/text;jsessionid=A2C7007A5C9E00CEE20937A8E469BFC4?doc=Perseus%3atext%3a2010.11.0005, accessed 15 May 2015)

CDL = *Codice diplomatico longobardo* 1–2. A cura di Luigi Schiaparelli. Tipografia del Senato: Roma, 1929–1933. (http://nlp.perseus.tufts.edu/hopper/text;jsessionid=A2C7007A5C9E00CEE20937A8E469BFC4?doc=Perseus%3atext%3a2010.11.0004, accessed 15 May 2015)

CDL 4:2 = *Codice diplomatico longobardo* 4:2. I diplomi dei duchi di Benevento. A cura di L. Schiaparelli, C. Brühl & H. Zielinski. Fonti per la storia d'Italia 65. Istituto storico italiano: Roma, 2003.

CDL 5 = *Codice diplomatico longobardo* 5. Le chartae dei Ducati di Spoleto e di Benevento. A cura di L. Schiaparelli, C. Brühl & H. Zielinski. Fonti per la storia d'Italia 66. Istituto storico italiano: Roma, 1986.

Cennamo 1998 = Cennamo, M. 1998. 'The loss of the voice dimension between Late Latin and early Romance', in Schmid, M.S. (ed.). *Historical Linguistics 1997*, 81–100. Amsterdam: John Benjamins.

Cennamo 1999a = Cennamo, M. 1999. 'Inaccusatività tardo-latina e suoi riflessi in testi italiani antichi centro-meridionali', in *Zeitschrift für romanische Philologie* 115:2, 300–331.

Cennamo 1999b = Cennamo, M. 1999. 'Late Latin pleonastic reflexives and the unaccusative hypothesis', in *Transactions of the Philological Society* 97, 103–150.

Cennamo 2000 = Cennamo, M. 2000. 'Patterns of "active" syntax in Late Latin pleonastic reflexives', in Smith, J.C. & Bentley, D. (eds.). *Historical Linguistics* 1995, 81–145. Amsterdam: John Benjamins.

Cennamo 2001a = Cennamo, M. 2001. 'Classi verbali e cambiamento sintattico: la reinterpretazione passiva del costrutto riflessivo', in Fábián, Z. & Salvi, G. (eds.). *Semantica e Lessicologia Storiche*. Atti del XXXII Congresso Internazionale della Società Linguistica Italiana, Budapest, 1998, 225–242. Roma: Bulzoni.

Cennamo 2001b = Cennamo, M. 2001. 'L'*extended accusative* e le nozioni di voce e relazione grammatica nel latino tardo e medievale', in Viparelli, V. (ed.). *Ricerche linguistiche tra antico e moderno*, 3–27. Napoli: Liguori.

Cennamo 2001c = Cennamo, M. 2001. 'On the reorganization of voice distinctions and grammatical relations in Late Latin', in Moussy, C. (ed.). *De lingua Latina novae quaestiones*. Actes du Xe Colloque International de Linguistique Latine, Paris & Sèvres, 1999, 51–65. Peeters: Leuven.

Cennamo 2001d = Cennamo, M. 2001. 'Perifrasi passive in testi non toscani delle origini', in Fábián, Z. & Salvi, G. (eds.). *Semantica e Lessicologia Storiche*. Atti del XXXII Congresso Internazionale della Società Linguistica Italiana, Budapest, 1998, 105–127. Bulzoni: Roma.

Cennamo 2006 = Cennamo, M. 2006. 'The rise and grammaticalization paths of Latin *fieri* and *facere* as passive auxiliaries', in Abraham, W. & Leisiö, L. (eds.). *Passivization and typology: form and function*, 311–336. Amsterdam: John Benjamins.

Cennamo 2008 = Cennamo, M. 2008. 'The rise and development of analytic perfects in Italo-Romance', in Eythórsson, T. (ed.). *Grammatical Change and Linguistic Theory*, 115–142. Amsterdam: John Benjamins.

Cennamo 2009 = Cennamo, M. 2009. 'Argument structure and alignment variations and changes in Late Latin', in Barðdal, J. & Chelliah, S.L. (eds.). *The role of semantic, pragmatic, and discourse factors in the development of case*, 307–346. John Benjamins: Amsterdam & Philadelphia.

Cennamo 2010 = Cennamo, M. 2010. 'Perfective auxiliaries in the pluperfect in some southern Italian dialects', in D'Alessandro, R., Ledgeway, A. & Roberts, I. (eds.). *Syntactic Variation: The Dialects of Italy*, 210–224. Cambridge: Cambridge University Press.

Cennamo 2011 = Cennamo, M. 2011. 'Impersonal constructions and accusative subjects in Late Latin', in Malchukov, A. & Siewierska, A. (eds.). *Impersonal Constructions: A cross-linguistic perspective*, 169–188. John Benjamins: Amsterdam & Philadelphia.

Cennamo 2014 = Cennamo, M. 2014. 'Passive and impersonal reflexives in the Italian dialects', in Benincà, P., Ledgeway, A. & Vincent, N. (eds.). *Diachrony and Dialects*, 71–95. Oxford: Oxford University Press.

ChLA[1] = *Chartae Latinae Antiquiores*. Facsimile-edition of the Latin Charters Prior to the Ninth Century. Edited by Albert Bruckner, Robert Marichal & al. Urs Graf Verlag: Olten, Dietikon, Zürich, 1954–2001.

ChLA[2] = *Chartae Latinae Antiquiores*. Facsimile-edition of the Latin Charters. 2nd Series: Ninth Century. Edited by Guglielmo Cavallo, Giovanna Nicolaj & al. Urs Graf Verlag: Dietikon, Zürich, 1997–.

Clackson & Horrocks 2007 = Clackson, J. & Horrocks, G. 2007. *The Blackwell History of the Latin Language*. Chichester: Wiley-Blackwell.

Clary 2014 = Clary, T. 2014. 'The unaccusative hypothesis and case selection of cognate complements in Latin', in *Glotta* 90:1, 87–104.

Coleman 1987 = Coleman, R. 1987. 'Vulgar Latin and the Diversity of Christian Latin', in Herman, J. (ed.). *Latin vulgaire – latin tardif* I, 37–52. Tübingen: Niemeyer.

Coleman 1999 = Coleman, R. 1999. 'Vulgarism and normalization in the text of Regula Sancti Benedicti', in Petersmann, H. & Kettemann, R. (eds.). *Latin vulgaire – latin tardif V. Actes du Ve colloque international sur le latin vulgaire et tardif*, 345–356. Heidelberg: Carl Winter.

Collinge 1978 = Collinge, N.E. 1978. 'Restructuring of noun-cases in syntax. Why "anti-" will not do?', in Abraham, W. (ed.). *Valence, Semantic Case and Grammatical Relations*, 617–632. Amsterdam: John Benjamins.

Comrie 1989 = Comrie, B. 1989. *Language Universals and linguistic typology: syntax and morphology*. 2nd ed. Blackwell: Oxford.

Conti 1973 = Conti, P.M. 1973. 'La Tuscia e i suoi ordinamenti territoriali nell'alto medio evo', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, Lucca, 1971, 61–116. CISAM: Spoleto.

Courtois & al. 1952 = Courtois, C., Leschi, L., Perrat, C. & Saumagne, C. 1952. *Tablettes Albertini: actes privés de l'époque vandale (fin du Ve siècle)*. Éditions arts et métiers graphiques: Paris.

Croft 2003 = Croft, W. 2003. *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.

Cuzzolin 2010 = Cuzzolin, P. 2010. 'Evidentialitätsstrategien im Lateinischen: vorläufige Bemerkungen', in Anreiter, P. & Kienpointner, M. (eds.). *Akten des 15. Internationalen Kolloquiums zur Lateinischen Linguistik*, Innsbruck 2009, 247–256. Innsbruck: IBS.

Dahl 2000 = Dahl, Ö. 2000. 'Egophoricity in discourse and syntax', in *Functions of language* 7:1, 37–77.

de Dardel & Wüest 1993 = de Dardel, R. & Wüest, J. 1993. 'Les systèmes casuels de protoroman. Les deux cycles de simplification', in *Vox Romanica* 52, 25–65.

Devoto & Giacomelli 1971 = Devoto, G. & Giacomelli, G. 1971. *I Dialetti delle Regioni d'Italia*. Sansoni Editore: Firenze.

Dik 1968 = Dik, S. 1968. *Coordination: Its Implications for the Theory of General Linguistics*. Amsterdam: North Holland.

Dirr 1912 = Dirr, A. 1912. 'Rutulskij jazyk', in *Sbornik Materialov dlya Opisaniya y Plemen Kavkaza* 42:3, 1–204.

Dixon 1968 = Dixon, R. 1968. *The Dyirbal language of North Queensland*. PhD thesis, University of London.

Dixon 1994 = Dixon, R. 1994. *Ergativity*. Cambridge: Cambridge University Press.

Donohue 2008 = Donohue, M. 2008. 'Semantic alignment systems: what's what, and what's not', in Donohue, M. & Wichmann, S. (eds.). *The Typology of Semantic Alignment*, 24–75. Oxford University Press: Oxford.

Du Bois 1987 = Du Bois, J. 1987. 'The discourse basis of ergativity', in *Language* 63:4, 805–855.

Eddington 2010 = Eddington, D. 2010. 'A comparison of two tools for analyzing linguistic data: logistic regression and decision trees', in *Italian Journal of Linguistics* 22:2, 265–286.

Ellis & Beattie 1986 = Ellis, A. & Beattie, G. 1986. *The psychology of language and communication*. Weidenfeld and Nicolson: London.

Everett 2009 = Everett, C. 2009. 'A reconsideration of the motivations for preferred argument structure', in *Studies in Language* 33:1, 1–24.

Fabrizio 2012 = Fabrizio, C. 2012. 'Infinitives as non-canonical subjects', *hand-out* at the Non-Canonically Case-Marked Subjects within and across Languages and Language Families, Iceland, 2012.

Fabrizio (in press) = Fabrizio, C. (in press). 'The semantically-oriented syntax of subject infinitives in Latin and Homeric Greek', in Schmitt, R., Sadovski, V. & Luschutzky, H. (eds.). *Iranian and*

Indo-European Linguistics and Onomastics. Contributions in memory of M. Mayrhofer. Austrian Academy of Sciences Press: Vienna.

Faraoni 2010 = Faraoni, V. 2010. *L'origine dei plurali italiani in* -e *ed* -i. Doctoral thesis, La Sapienza, Rome.

Faraoni 2014a = Faraoni, V. 2014. 'La formazione del plurale italo-romanzo nella documentazione notarile altomedievale', in Molinelli, P., Cuzzolin, P. & Fedriani, C. (eds.). *Latin Vulgaire – Latin Tardif X. Actes du Xe colloque international sur le latin vulgaire et tardif*, Bergamo, 2012.

Faraoni 2014b = Faraoni, V. 2014. 'Tracce romanze della riorganizzazione morfosintattica del sistema nominale di transizione: il caso dei plurali italiani', *hand-out* at the Latin vulgaire – latin tardif XI, Oviedo, 2014.

Finco 2013 = Finco, F. 2013. 'La flessione nominale tra plurale sigmatico e plurale palatale: considerazioni storiche e confronti tra varietà ladine e friulane', in Arcamone, M.G. & Mastrelli, C.A. (eds.). *Lingua e cultura nelle Alpi: studi in onore di Johannes Kramer*, 341–358. Istituto di Studi per l'Alto Adige: Firenze.

Fiorentino 1994 = Fiorentino, G. 1994. 'Aspetti della morfosintassi nominale nelle carte cavensi del IX secolo', in *Archivio Glottologico Italiano* 79:1, 23–50. Le Monnier: Firenze.

Foley & Van Valin 1984 = Foley, W.A. & Van Valin, R.D. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Formentin 2002 = Formentin, V. 2002. 'Arrigo Castellani, *Grammatica storica della lingua italiana*, I. *Introduzione*, Bologna (Il Mulino) 2000, xxxviii–618 p. (Collezione di testi e di studi – Linguistica e critica letteraria)', in *Vox Romanica* 61, 295–302.

Formentin 2004 = Formentin, V. 2004. 'Un esercizio ricostruttivo: veneziano antico fondi 'fondo', ladi 'lato', peti 'petto'', in Drusi, R., Perocco, D. & Vescovo, P. (eds.). *"Le sorte dele parole": testi veneti dalle origini all'Ottocento*. Atti dell'Incontro di studio, Venezia, 2002. Esedra editrice: Padova.

Formentin 2012 = Formentin, V. 2012. 'Un nuovo testo per la storia del romanesco medievale', in Loporcaro, M., Faraoni, V. & Di Pretoro, P. (eds.). *Vicende storiche della lingua di Roma*. Edizioni dell'Orso: Alessandria.

Gaeng 1990 = Gaeng, P. 1990. 'La flexion nominale à l'époque du latin tardif: essai de reconstruction', in Calboli, G. (ed.). *Latin vulgaire – latin tardif II. Actes du IIème Colloque international sur le latin vulgaire et tardif*, 111–128. Tübingen: Niemeyer.

Galdi 2012 = Galdi, G. 2012. 'Again on as-nominatives: A New Approach to the Problem', in Leiwo, M., Halla-aho, H. & Vierros, M. (eds.). *Variation and change in Greek and Latin*, 139–152. Helsinki: Suomen Ateenan-Instituutin säätiö.

Galdi 2013 = Galdi, G. 2013. *Syntaktische Untersuchungen zu Jordanes. Beiträge zu den* Romana. Hildesheim, Zürich & New York: Georg Olms Verlag.

García-Miguel & Vázquez Rozas 2009 = García-Miguel, J. & Vázquez Rozas, V. 2009. 'Transitivity, subjectivity and discourse: A Corpus Analysis of Spanish Argument Structure', *presentation* at the Fillmore Fest, Berkeley, 2009. (http://gramatica.usc.es/~vvazq/pdf_pres/Garcia_Vazquez_FillmoreFest_2009.pdf, accessed 5 May 2015)

Gerola 1949–1950 = Gerola, B. 1949–1950. 'Aspetti della sintassi del nominativo e dell'accusativo nel tardo latino', in *Atti dell'istituto veneto di scienze, lettere ed arti* 108, 207–236.

Gianollo 2005 = Gianollo, C. 2005. 'Middle Voice in Latin and the Phenomenon of Split Intransitivity', in Calboli, G. & al. (eds.) *Latina lingua!* Proceedings of the Twelfth International Colloquium on Latin Linguistics (ICLL), Bologna, 2003, 97–110. Herder: Roma.

Gordon & Caramazza 1985 = Gordon, B. & Caramazza, A. 1985. 'Lexical access and frequency sensitivity: frequency saturation and open/closed class equivalence', in *Cognition* 21, 95–115.

Grant 1996 = Grant, M. 1996. *Anthimus. De observatione ciborum. On the Observance of Foods*. Totnes: Devon.

Greco 2012. = Greco, P. 2012. *La complementazione frasale nelle cronache latine dell'Italia centro-meridionale* (secoli X–XII). Napoli: Liguori.

Griffin & Crew 2012 = Griffin, Z.M. & Crew, C.M. 2012. 'Research in language production', in Spivey, M., McRae, K. & Joanisse, M. (eds.). *The Cambridge Handbook of Psycholinguistics*. Cambridge: Cambridge University Press.

Guyotjeannin & al. 1993 = Guyotjeannin, O., Pycke, J. & Tock, B.-M. 1993. *Diplomatique médiévale*. L'atelier du médiéviste, 2. Brepols: Paris.

Haig & Schnell (in preparation) = Haig & Schnell (in preparation). 'The discourse basis of ergativity revisited', *manuscript* currently under review (February 2015).

Hajič & al. 1999 = Hajič, J., Panevová, J., Buráňová, E., Urešová, Z. & Bémová, A. 1999. *Annotations at Analytical Level: instructions for annotators*. Institute of Formal and Applied Linguistics: Prague. (http://ufal.mff.cuni.cz/pdt/Corpora/PDT_1.0/References/aman_en.pdf, accessed 15 May 2015)

Hajič 1998 = Hajič, J. 1998. 'Building a syntactically annotated corpus: The Prague Dependency Treebank', in Hajičová, E. (ed.). *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, 12–19. Charles University Press: Prague.

Halla-aho 2009 = Halla-aho, H. 2009. *The Non-literary Latin Letters. A study of their syntax and pragmatics*. Societas Scientiarum Fennica: Tammisaari.

Harris & Campbell 1995 = Harris, A.C. & Campbell, L. 1995. *Historical Syntax in Cross-Linguistic Perspective*. Cambridge: Cambridge University Press.

Havers 1928 = Havers, W. 1928. 'Zur Syntax des Nominativs', in *Glotta* 16, 94–127.

Havers 1931 = Havers, W. 1931. *Handbuch der erklärenden Syntax: ein Versuch zur Erforschung der Bedingungen und Triebkräfte in Syntax und Stilistik*. Carl Winter: Heidelberg.

Hawkins 1983 = Hawkins, J.A. 1983. *Word Order Universals*. Academic Press: New York.

Hawkins 2004 = Hawkins, J.A. 2004. *Efficiency and Complexity in Grammars*. Oxford University Press: Oxford.

Herman 1987 = Herman, J. 1987. 'La disparition de *-s* et la morphologie dialectale du latin parlé', in Herman, J. (ed.). *Latin vulgaire – latin tardif* I. Actes du Ier Colloque international sur le latin vulgaire et tardif, Pécs, 1985, 97–108. Tübingen: Niemeyer.

Herman 1990 = Herman, J. 1990. 'Sur un exemple de la langue parlée à Rome au VIe siècle', in Kiss, S. (ed.). *Du latin aux langues romanes* 2: nouvelles études de linguistique historique, 158–168. Tübingen: Max Niemeyer Verlag.

Herman 1997 = Herman, J. 1997. 'À propos du débat sur le pluriel des noms italiens (et roumains): à la recherche d'une conclusion', in Holtus, G., Kramer, J. & Schweickard, W. (eds.). *Italica et romanica. Festschrift für Max Pfister zum 65. Geburtstag*. Bd. 2, 19–30. Tübingen: Niemeyer.

Herman 1998 = Herman, J. 1998. 'La chronologie de la transition: un essai', in Herman, J. (ed.). *La transizione dal latino alle lingue romanze: atti della Tavola Rotonda di Linguistica Storica*, Venezia, 1996, 5–26. Tübingen: Niemeyer.

Herman 2002 = Herman, J. 2002. 'La disparition du passif synthétique latin: nouvel essai sur l'écrit et le parlé en latin mérovingien', in *Estudios Romanicos* 24, 31–46.

Hopper & Thompson 1980 = Hopper, P.J. & Thompson, S.A. 1980. 'Transitivity in Grammar and Discourse', in *Language* 56:2, 251–299.

Hundt 2008 = Hundt, M. 2008. 'Text corpora', in Lüdeling, A. & Kytö, M. (eds.) *Corpus Linguistics: An International Handbook*, 168–187. Mouton de Gruyter: Berlin.

IBM = *IBM SPSS Decision Trees* 20. IBM Corporation 1989, 2011.

Itkonen 1974 = Itkonen, T. 1974. 'Ergatiivisuutta suomessa', in *Virittäjä* 78:4, 379–398.

Jaeger & Tily 2010 = Jaeger T. F. & Tily, H. 'On language 'utility': processing complexity and communicative efficiency', in *WIREs Cogn Sci* 2011, 2: 323–335. (doi: 10.1002/wcs.126)

Kintsch 1977 = Kintsch, W. 1977. *Memory and Cognition*. Robert E. Krieger Publishing Company: Malabar.

Kittilä 2002 = Kittilä, S. 2002. *Transitivity: toward a comprehensive typology*. Åbo: Åbo Akademiska Tryckeri.

Korkiakangas 2010a = Korkiakangas, T. 2010. *Neutri Ravennan papyruksissa: määrällinen ja laadullinen analyysi neutrikategorian tilasta ja muutoksesta myöhäislatinalaisessa tekstikorpuksessa* [*Neuter in the Ravenna Papyri: A quantitative and qualitative analysis of the state and change of the neuter category in a Late Latin text corpus*]. Unpublished Master's thesis, Romance and Classical Languages, University of Jyväskylä. (http://urn.fi/URN:NBN:fi:jyu-201003161324, accessed 15 May 2015)

Korkiakangas 2010b = Korkiakangas, T. 2010. 'Neutro plurale e femminile singolare: il fattore grafico nell'interpretazione delle peculiarità della declinazione tardolatina', *Arctos* 44 (Acta Philologica Fennica), 133–145.

Korkiakangas (in press) = Korkiakangas, T. (in press). "Morphosyntactic realignment and markedness change in Late Latin: Evidence from charter texts", in *Pallas* (Presses universitaires du Mirail) 102/103 (2016/2017).

Korkiakangas & Lassila 2013 = Korkiakangas, T. & Lassila, M. 2013. 'Abbreviations, fragmentary words, formulaic language: treebanking medieval charter material', in Mambrini, F. & al. (eds.). *Proceedings of The Third Workshop on Annotation of Corpora for Research in the Humanities* (ACRH-3), Sofia, 2013, 61–72. Bulgarian Academy of Sciences.

Korkiakangas & Passarotti 2011 = Korkiakangas, T. & Passarotti, M. 2011. 'Challenges in Annotating Medieval Latin Charters', in *Journal of Language Technology and Computational Linguistics* (JLCL) 26:2, 103–114. Proceedings of the ACRH Workshop, Heidelberg, 2012.

Kurze 1974 = Kurze, W. 1974. *Codex Diplomaticus Amiatinus*. Erster Band: Von den Anfängen bis zum Ende der Nationalkönigsherrschaft (736–951). Max Niemeyer Verlag: Tübingen.

Kälviäinen 2013 = Kälviäinen, N. 2013. 'Levels of Style in Byzantine Greek and the Role of Syntactic Complexity. A Quantitative Analysis of the Sentence Structure of Three Early Byzantine Hagiographic Texts', in *Arctos* 47 (Acta Philologica Fennica), 173–198.

La Fauci 1988 = La Fauci, N. 1988. *Oggetti e soggetti nella formazione della morfosintassi romanza*. Nuova collana di linguistica, 7. Pisa: Giardini.

La Fauci 1997 = Fauci, N. 1997. *Per una teoria grammaticale del mutamento morfosintattico*. Dal latino verso il romanzo. Pisa: ETS.

La Fauci 2001 = La Fauci, N. 2001. 'Quel pasticciaccio brutto della declinazione scomparsa?', in *Vox Romanica* 60, 15–24.

Lakoff 1977 = Lakoff, G. 1977. 'Linguistic gestalts', in Woodford, B., Fox, S.E. & Philosoph, S. (eds.). *Proceedings of the Thirteenth Regional Meeting of the Chicago Linguistic Society*, 236–287. Chicago: University of Chicago.

Larson 1988 = Larson, P. 1988. *Gli elementi volgari nelle carte del Codice diplomatico longobardo*. Tesi di laurea (inedita), Università degli Studi di Firenze.

Larson 2000 = Larson, P. 2000. 'Tra linguistica e fonti diplomatiche: quello che le carte dicono e non dicono', in Herman, J. & Marinetti, A. (eds.). *La preistoria dell'italiano*. Atti della Tavola Rotonda di Linguistica Storica. Università Ca' Foscari di Venezia, 1999, 151–166. Niemeyer: Tübingen.

Larson 2003 = Larson, P. 2003. 'Il volgare del Mille: fonti per la conoscenza dell'italiano preletterario', in Poggi Salani, T. & Maraschio, N. (eds.). *Italia linguistica anno mille, Italia linguistica anno duemila*. Atti del XXXIV congresso internazionale di studi della Società di linguistica italiana, Firenze, 2000, 129–137. Bulzoni: Roma.

Larson 2012 = Larson, P. 2012. 'Le carte alto-medievali come fonte di lingua: qualche esperienza personale', in Sornicola, R. & Greco, P. (eds.). *La lingua dei documenti notarili alto-medievali dell'Italia meridionale: bilancio degli studi e prospettive di ricerca*. Tavolario edizioni: Napoli.

Lausberg 1962 = Lausberg, H. 1962. Romanische Sprachwissenschaft III: Formenlehre. Erster Teil. Walter de Gruyter: Berlin.

Lazard 1991 = Lazard, G. 1991. 'Researches on actancy', in *Actances* 6, 3–58.

Lazard 1993 = Lazard, S. 1993. 'Indices de la langue parlée à Ravenne au V^{ie} siècle à travers le témoignage des chartes', in *Actes du XXe Congrès international de linguistique et philologie romanes*. Vol. 3, 391–402.

Lazard 2007 = Lazard, S. 2007. 'La crise de la scripta au Xe siècle: Ce que nous apprend le plaid de Lusignano de 915', in Trotter, D. & al. (eds.). *Actes du XXIVe Congrès International de Linguistique et Philologie Romanes*, Aberystwyth, 2004. Vol. 2, 511–526. Niemeyer: Tübingen.

Lazzeroni 2002 = Lazzeroni, R. 2002. 'Ruoli tematici e genere grammaticale: un aspetto della morfosintassi indoeuropea?', in *Archivio glottologico italiano* 87, 1–17.

Ledgeway 2011 = Ledgeway, A. 2011. 'Syntactic and morphosyntactic typology and change', in Ledgeway, A., Maiden, M. & Smith, J.C. (eds.). *The Cambridge history of the Romance languages*, 382–471. Cambridge: Cambridge University Press.

Ledgeway 2012 = Ledgeway, A. 2012. *From Latin to Romance*. Morphosyntactic typology and change. Oxford & New York: Oxford University Press.

Lehmann 1985 = Lehmann, C. 1985. 'Ergative and active traits in Latin', in Plank, F. (ed.) *Trends in linguistics*, 243–255. Mouton de Gruyter: Berlin.

Lehmann 1991 = Lehmann, C. 1991. 'Predicate classes and participation', in Seiler, H. & Premper, W. (eds.). *Partizipation: das sprachliche Erfassen von Sachverhalten*, 183–239. Tübingen: Narr.

Lehto 2013 = Lehto, A. 2013. 'Complexity and genre conventions: Text structure and coordination in Early Modern English proclamations', in Jucker, A. H. & al. (eds.). *Meaning in the History of English: Words and texts in context*, 233–256. Amsterdam: John Benjamins.

Lewis 2000 = Lewis, R.J. 2000. 'An Introduction to Classification and Regression Tree (CART) Analysis', *paper* presented at the Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California. (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.4103&rep=rep1&type=pdf, accessed 31 Mar 2015).

Loebell & Bock 2003 = Loebell, H. & Bock, K. 2003. 'Structural priming across languages', in *Linguistics* 41:5, 791–824.

Löfstedt 1961 = Löfstedt, B. 1961. *Studien über die Sprache der langobardischen Gesetze*. Almqvist & Wiksell: Stockholm, Göteborg, Uppsala.

Magidson & Vermunt 2005 = Magidson, J. & Vermunt, J. 2005. 'An extension of the CHAID tree-based segmentation algorithm to multiple dependent variables', in Weihs, C. & Gaul, W. (eds.). *Classification: The Ubiquitous Challenge*. Heidelberg: Springer, 240–247. (http://statisticalinnovations.com/products/8pagearticle.pdf, accessed 31 Mar 2015)

Maiden & Parry 1997 = Maiden, M. & Parry, M.M. (eds.). 1997. *The Dialects of Italy*. Routledge: London.

Maiden 1996 = Maiden, M. 1996. 'On the Romance inflectional endings -i and -e', in *Romance Philology* 50, 147–182.

Maiden 1998 = Maiden, M. 1998. *Storia linguistica dell'italiano*. Traduzione di Pietro Maturi. Bologna: Mulino.

Maiden 2000 = Maiden, M. 2000. 'Il sistema desinenziale del sostantivo nell'italo-romanzo preletterario. Ricostruzione parziale a partire dai dialetti moderni (il significato storico di plurali del tipo amici)', in Herman, J. & Marinetti, A. (eds.). *La preistoria dell'italiano: atti della Tavola Rotonda di Linguistica Storica*, Venezia, 1999, 167–179. Tübingen: Niemeyer, 2000.

Malchukov 2006 = Malchukov, A. 2006. 'Transitivity parameters and transitivity alternations', in Kulikov, L., Malchukov, A. & de Swart, P. (eds.). *Case, Valency, and Transitivity*, 329–357. Amsterdam: John Benjamins.

MED = Memorie e documenti per servire all'istoria del Ducato di Lucca. Parte 5, tomo 2. A cura di Domenico Barsocchini. Lucca 1837. ([http://nlp.perseus.tufts.edu/hopper/text;jsessionid=A2C700 7A5C9E00CEE20937A8E469BFC4?doc=Perseus%3atext%3a2010.11.0006](http://nlp.perseus.tufts.edu/hopper/text;jsessionid=A2C7007A5C9E00CEE20937A8E469BFC4?doc=Perseus%3atext%3a2010.11.0006), accessed 15 May 2015)

Mel′čuk 1988 = Mel′čuk, I. 1988. *Dependency syntax: theory and practice*. State University Press of New York: Albany.

Mithun 1991 = Mithun, M. 1991. 'Active/agentive case marking and its motivations', in *Language* 67, 510–546.

Moravcsik 1978 = Moravcsik, E.A. 1978. 'On the distribution of ergative and accusative patterns', in *Lingua* 45, 233–279.

Munro 1982 = Munro, P. 1982. 'On the transitivity of "Say" verbs', in Hopper, P. & Thompson, S. A. (eds.). *Studies in Transitivity*, 301–318. New York: Academic Press.

Muratori 1751 = Muratori, L. 1751. *Dissertazioni sopra le Antichità italiane* 2. Giambatista Pasquali: Milano.

Naess 2007 = Naess, Å. 2007. *Prototypical transitivity*. John Benjamins: Amsterdam & Philadelphia.

Nivre 2008 = Nivre, J. 2008. 'Treebanks', in Lüdeling, A. & Kytö, M. (eds.). *Corpus Linguistics: An International Handbook*, 225–241. Mouton de Gruyter: Berlin.

Orlandini 1996 = Orlandini, A. 1996. '1, 2, 3 videor: analyse d'un prédicat polysémique', in Bammesberger, A. & Heberlein, F. (eds.). *Akten des VIII. internationalen Kolloquiums zur lateinischen Linguistik*. C. Winter: Heidelberg.

Pei 1932 = Pei, M.A. 1932. *The language of the eight*[*h*]*-century texts in Northern France*. New York: Columbia University Press.

Pei 1937 = Pei, M.A. 1937. 'Accusative or oblique? A synthesis of the theories concerning the origin of the oblique case of the Old French and the single-case system of other Romance languages', in *Romanic Review* 28, 241–267.

Penny 1980 = Penny, R. 1980. 'Do Romance nouns descend from the Latin accusative?', in *Romance Philology* 33, 501-509.

Penny 2002 = Penny, R. 2002. *A History of the Spanish Language*. Cambridge University Press: Cambridge.

Pensado 1986 = Pensado, C. 1986. 'Inversion de marquage et perte du système casuel en ancien français', in *Zeitschrift für romanische Philologie* 102, 271–296.

Perlmutter 1978 = Perlmutter, D. 1978. 'Impersonal Passives and the Unaccusative Hypothesis', in *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, 157–189. University of Carolina: Berkeley.

Petrucci & Romeo 1992 = Petrucci, A. & Romeo, C. 1992. *Scriptores in urbibus*: *alfabetismo e cultura scritta nell'Italia altomedievale*. Il Mulino: Bologna.

Philippart de Foy 2012 = Philippart de Foy, C. 2012. 'Lemmatiser un corpus de textes hagiographiques: enjeux et modalités pratiques', in Biville, F. (ed.). *Latin vulgaire – latin tardif* IX. Actes du IXe colloque international sur le latin vulgaire et tardif (LVLT), Lyon, 481–490. Publications de la Maison de l'Orient: Lyon.

Pieroni 1999 = Pieroni, S. 1999. 'Non-promotional objects in Late Latin', *Verbum* 21, 117–129.

Pinkster 1990 = Pinkster, H. 1990. *Latin Syntax and Semantics*. Routledge: London.

Pinkster 1991 = Pinkster, H. 1991. 'Evidence for SVO in Latin?', in Wright, R. (ed.) *Latin and the Early Romance Languages in the Middle Ages*, 69–82. London: Routledge.

PItal = *The Ravenna papyri*. See Tjäder 1955 and Tjäder 1982.

Plank 1979 = Plank, F. 1979. 'The Functional Basis of Case Systems and Declension Classes: From Latin to Old French', in *Linguistics* 17, 611–640.

Plank 1985 = Plank, F. 1985. 'The extended accusative/restricted nominative in perspective', in Plank, F. (ed.). *Relational typology*, 269–310. Berlin & New York: Mouton de Gruyter.

Plank 1995 = Plank, F. 1995. 'Syntactic change: Ergativity', in Jacobs, J. & al. (eds.). *Syntax: An International Handbook of Contemporary Research* 2, 1184–1199. Berlin: Walter de Gruyter.

Poccetti & al. = Poccetti, P., Poli, D. & Santini C. 1999. *Una storia della lingua latina*: *formazione, usi, comunicazione*. 7^a ristampa 2008. Roma: Carocci editore.

Politzer & Politzer 1953 = Politzer, F.N. & Politzer, R.L. 1953. *Romance Trends in 7th and 8th Century Latin Documents*. The University of North Carolina Press: Chapel Hill.

Pratesi 1979 = Pratesi, A. 1979. *Genesi e forme del documento medievale*. Jouvence: Roma.

Priiki 2014 = Priiki, K. 2014. 'Kaakkois-Satakunnan henkilöviitteiset *hän*, *se*, *tää* ja *toi* subjekteina', in *Sananjalka* 56, 86–107.

Priiki (in press) = Priiki, K. (in press). 'Kuka puhuu ja kenestä? Pronominien *hän*, *se*, *tää* ja *toi* henkilöreferentit'.

Proskauer 1910 = Proskauer, C. 1910. *Das auslautende -s auf den lateinischen Inschriften*. Trübner: Strassburg.

Raumolin-Brunberg 1991 = Raumolin-Brunberg, H. 1991. *The Noun Phrase in Early Sixteenth-Century English*. Mémoires de la Société Néophilologique de Helsinki 50. Helsinki: Société Néophilologique.

Rio 2009 = Rio, A. 2009. *Legal Practice and the Written Word in the Early Middle Ages: Frankish Formulae, c. 500–1000*. Cambridge Studies in Medieval Life and Thought. Fourth Series 75.

Rivas 2000 = Rivas, J. 2000. *Ergativity and Transitive Gradients in the Accusative and Infinitive Construction*. A cross-linguistic typological approach. Universidade de Santiago de Compostela: Santiago de Compostela.

Rovai 2012a = Rovai, F. 2012. 'Between feminine singular and neuter plural: re-analysis patterns', in *Transactions of the Philological Society* 110:1, 94–121.

Rovai 2005 = Rovai, F. 2005. 'L'estensione dell'accusativo in latino tardo e medievale', in *Archivio glottologico italiano* 90, 54–89.

Rovai 2007 = Rovai, F. 2007. 'Manifestazioni di sub-sistemi tipologici attivi in latino', in *Archivio Glottologico Italiano* 92, 51–64.

Rovai 2010 = Rovai, F. 2010. 'Active traits in Latin: evidence from literary and epigraphic texts', in Anreiter, P. & Kienpointner, M. (eds.). *Akten des 15. Internationalen Kolloquiums zur Lateinischen Linguistik*, Innsbruck 2009, 317–329. Innsbruck: IBS.

Rovai 2012 = Rovai, F. 2012. *Sistemi di codifica argomentale. Tipologia ed evoluzione*. Pisa: Pacini Editore.

Sabatini 1963–1964 = Sabatini, F. 1963–1964. 'Riflessi linguistici della dominazione longobarda nell'Italia mediana e meridionale', in *Atti e memorie dell'Accademia toscana di scienze e lettere "La colombaria"* 28, 125–249. Leo S. Olschki: Firenze.

Sabatini 1965 = Sabatini, F. 1965. 'Esigenze di realismo e dislocazione morfologica in testi preromanzi', in *Rivista di Cultura Classica e Medievale* 7, 972–998.

Sabatini 1968 = Sabatini, F. 1968. 'Dalla "scripta latina rustica" alle "scriptae" romanze', in *Studi medievali* 9, 320–358.

Sala 2005 = Sala, M. 2005. *From Latin to Romanian: The Historical Development of Romanian in a Comparative Romance Context*. Romance Monographs: Mississippi.

Sanga & Baggio 1995 = Sanga, G. & Baggio, S. 1995. 'Sul volgare in età longobarda', in *Italia settentrionale: crocevia di idiomi romanzi*. A cura di E. Banfi & al., 247–260. Niemeyer: Tübingen.

Sanga 1995 = Sanga, G. 1995. 'Italienische Koine – La koiné italiana', in Holtus, G., Metzeltin, M. & Schmitt, C. (eds.). *Lexikon der Romanistischen Linguistik* II:2, 81–98. Tübingen: Max Niemeyer.

Schiaparelli 1929 = Schiaparelli, L. 1929. 'Avvertenza', in *Codice Diplomatico Longobardo* (CDL) 1, vii–ix. Fonti per la Storia d'Italia, 62. Istituto storico italiano: Roma.

Schiaparelli 1933 = Schiaparelli, L. 1933. 'Note diplomatiche sulle carte longobarde, II: Tracce di antichi formulari nelle carte longobarde', in *Archivio storico italiano* 19, 3–34.

Schmidt-Wiegand 2001 = Schmidt-Wiegand, R. 2001. 'Leges Alamannorum', in *Reallexikon der germanischen Altertumskunde* 18, 201–202. Berlin: de Gruyter.

Schøsler 1984 = Schøsler, L. 1984. *La déclinaison bicasuelle de l'ancien français, son rôle dans la syntaxe de la phrase, les causes de sa disparition*. Études Romanes de l'Université d'Odense 19. Odense: Odense university Press.

Schøsler 2001 = Schøsler, L. 2001. 'From Latin to Modern French: Actualization and Markedness', in Andersen, H. (ed.) *Actualization*: *linguistic change in progress*. Amsterdam: John Benjamins. Amsterdam studies in the theory and history of linguistic science. Ser. 4. Current issues in linguistic theory 219.

Seidl 1995 = Seidl, C. 1995. 'Le système acasuel des protoromans ibérique et sarde: dogmes et faits', in *Vox Romanica* 54, 41–73.

Sgall & al. 1986. = Sgall, P., Hajičová, E. & Panevová, J. 1986. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Reidel Publishing Company: Dordrecht / Academia: Prague.

Siewierska 1988 = Siewierska, A. 1988. *Word Order Rules*. Croom Helm: London, New York & Sydney.

Silverstein 1976 = Silverstein, M. 1976. 'Hierarchy of features and ergativity', in Muysken, P. & van Riemsdijk, H. (eds.). *Features and Projections*, 163–232. Dordrecht: Foris. [1986]

Singer 1990 = Singer, M. 1990. *Psychology of language: an introduction to sentence and discourse processes*. Hillsdale: Erlbaum.

Smith 2011 = Smith, J.C. 2011. 'Change and continuity in form-function relationships', in Maiden, M., Smith, J.C. & Ledgeway, A. (eds.) *The Cambridge history of the Romance languages*. Vol. 1: Structures, 268–317. Cambridge: Cambridge University Press.

Sorace 2000 = Sorace, A. 2000. 'Gradients in auxiliary selection with intransitive verbs', in *Language* 76, 859–890.

Sornicola 1995 = Sornicola, R. 1995. 'Theticity, VS Order and the Interplay of Syntax, Semantics and Pragmatics', *Sprachtypologie und Universalienforschung* 48, 1–2: 72–83.

Sornicola 2008a = Sornicola, R. 2008. 'Nominal Inflection and Grammatical Relations in Tenth-Century Legal Documents from the South of Italy (Codex Diplomaticus Amalfitanus)', in Wright, R. (ed.). *Latin vulgaire – latin tardif* VIII, Oxford, 2006, 510–520. Olms-Weidmann: Hildesheim.

Sornicola 2010 = Sornicola, R. 2010. 'Syntactic conditioning of case marking loss: a long term factor between Latin and Romance?', in van Acker, M., van Deyck, M. & van Uytfanghe, M.

(eds.). *Latin écrit – roman oral?: de la dichotomisation à la continuité*. Corpus Christianorum 5, Brepols: Turnhout.

Sornicola 2011 = Sornicola, R. 2011. 'Romance Linguistics and Historical Linguistics: Reflections on Synchrony and Diachrony', in Maiden, M., Smith, J.C. & Ledgeway, A. (eds.) *The Cambridge history of the Romance languages*. Vol. 1: Structures, 1–49. Cambridge: Cambridge University Press.

Sornicola 2012 = Sornicola, R. 2012. 'Bilinguismo e diglossia dei territori bizantini e longobardi del Mezzogiorno. Le testimonianze dei documenti del IX e X secolo', in *Quaderni dell'Accademia Pontaniana* 59, 1–102. Accademia Pontaniana: Napoli.

Sornicola 2013 = Sornicola, R. 2013. 'Variazione strutturale e stilistica nel tempo e cambiamento linguistico: alcune riflessioni sul Cartulario del Chronicon Sanctae Sophiae', in Boutier, M.-G., Hadermann, P. & van Acker, M. (éd.) *La variation et le changement en langue (langues romanes)*. Mémoires de la Societé Néophilologique de Helsinki 37, 21–46.

Spevak 2010 = Spevak, O. 2010. *Constituent Order in Classical Latin Prose*. Studies in Language Companion Series 117. Amsterdam: John Benjamins.

Spevak 2014 = Spevak, O. 2014. *The Noun Phrase in Classical Latin Prose*. Amsterdam Studies in Classical Philology 21. Leiden: Brill.

Štěpánek & Pajas 2010 = Štěpánek, J. & Pajas, P. 2010. 'Querying Diverse Treebanks in a Uniform Way', in *Proceedings of the 7th International Conference on Language Resources and Evaluation* (LREC 2010), 1828–1835. European Language Resources Association: Valletta.

Szmrecsányi 2004 = Szmrecsányi, B. 2004. 'On operationalizing syntactic complexity', in Purnelle, G., Fairon, C. & Dister, A. (eds.). *Le Poids des mots. 7th International Conference on Textual Data Statistical Analysis*, vol. II, 1032–1039. Louvain-la-Neuve: Presses universitaires de Louvain.

Taavitsainen, I. 2001. 'Changing conventions of writing: The dynamics of genres, text types, and text traditions', in *European Journal of English Studies* 5:2, 139–150.

Talmy 1996. = Talmy, L. 1996. 'Fictive motion in language and "ception"', in Bloom, P., Peterson, M.A., Nadel, L. & Garrett, M.F. (eds.). *Language and space*, 211–276. Cambridge, MA: MIT Press.

Taylor 1924 = Taylor, P. 1924. *The Latinity of the* Liber historiae francorum: *a phonological, morphological and syntactical study*. New York:Carranza & Co.

Tekavčić 1972 = Tekavčić, P. 1972. *Grammatica storica dell'italiano.* Volume II: Morfosintassi. Bologna: Mulino.

Tekavčić 1975 = Tekavčić, P. 1975. 'Agli albori dell'italiano', in *Linguistica* 15, 209–239.

Thompson & Hopper 2001 = Thompson, S.A. & Hopper, P.J. 2001. 'Transitivity, clause structure, and argument structure: evidence from conversation', in Bybee, J. & Hopper, P. (eds.) *Frequency and the Emergence of Linguistic Structure*. John Benjamins: Amsterdam/Philadelphia.

Thompson 2002 = Thompson, S.A. 2002. '"Object complements" and conversation: Towards a realistic account', in *Studies in Language* 26:1, 125—164.

Timberlake 1975 = Timberlake, A. 1975. 'Hierarchies in the genitive of negation', *Slavic and East European Journal* 19, 123–138.

Timberlake 1977 = Timberlake, A. 1977. 'Reanalysis and actualization in syntactic change', in Li, C.N. (ed.). *Mechanisms of syntactic change*, 141–177. Austin: University of Texas Press.

Tjäder 1955 = Tjäder, J.-O. 1955. *Die nichtliterarischen Lateinischen Papyri Italiens aus der Zeit 445–700*. Vol. 1: Papyri 1–28. Acta Instituti Romani Regni Sueciae, Series in 4˚, XIX:1. C.W.K. Gleerup: Lund.

Tjäder 1982 = Tjäder, J.-O. 1982. *Die nichtliterarischen Lateinischen Papyri Italiens aus der Zeit 445–700*. Vol. 2: Papyri 29–59. Acta Instituti Romani Regni Sueciae, Series in 4˚, XIX:2. Paul Åströms Förlag: Lund.

TLL = *Thesaurus Linguae Latinae*. B.G. Teubner: Lipsiae. 1900–.

Torrego 2009 = Torrego, M. 2009. 'Coordination', in Baldi, P. & Cuzzolin, P. (eds.). *New Perspectives on Historical Latin Syntax* 1: Syntax of the Sentence, 443–487. Mouton de Gruyter: Berlin & New York.

Traugott 1989 = Traugott, E. 1989. 'On the rise of epistemic meanings in English: An example of subjectification in semantic change', in *Language* 65:1, 31–55.

Uddholm 1953 = Uddholm, A. 1953. *Formulae Marculfi*. Études sur la langue et le style. Almqvist & Wiksells Boktryckeri AB: Uppsala.

Uhlenbeck 1901–1902 = Uhlenbeck, C.C. 1901–1902. 'Agens und Patiens im Kasussystem der indogermanischen Sprachen', *Indogermanische Forschungen* 12, 170–172; 13, 101–172.

Wackernagel 1926 = Wackernagel, J. 1926. *Vorlesungen über Syntax mit besonderer Berücksichtigung von Griechisch, Lateinisch und Deutsch*. Reihe 1. Basel: Birkhäuser.

Valentini 2012 = Valentini, C. 2012. 'La questione dell'accusativo esteso in latino volgare', unpublished Master's thesis (Tesi in Glottologia – Linguistica storica e tipologica). University of Florence.

van Reenen & Schøsler 1997 = van Reenen, P. & Schøsler, L. 1997. 'La déclinaison en ancien et en moyen français, deux tendances contraires', in Combettes, B. & Monsonégo, S. (eds.). *Le moyen français, philologie et linguistique. Approches du texte et du discours*. Actes du VIIIe Colloque international sur le moyen français, Nancy, 1994, 595–612. Paris: Didier.

van Reenen & Schøsler 2000 = van Reenen, P. & Schøsler, L. 2000. 'Declension in Old and Middle French', in Smith, J.C. & Bentley, D. (eds.). *Historical linguistics 1995: selected papers from the 12th International Conference on Historical Linguistics*, *Manchester*, *1995*. Volume 1: general issues and non-Germanic languages, 327–344. Amsterdam: John Benjamins

Van Valin 1990 = Van Valin, R. 1990. 'Semantic parameters of split intransitivity', in *Language* 66, 221–260.

Van Valin 2001 = Van Valin, R. 2001. *An introduction to syntax*. Cambridge: Cambridge University Press.

Van Valin 2005 = Van Valin, R. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge University Press: Cambridge.

Vázquez & García-Miguel 2006 = Vázquez Rozas, V. & García-Miguel, J.M. 'Transitividad, subjetividad y frecuencia de uso en español', in *Actes del VII Congrés de Lingüística General*, Barcelona, 2006. Publication in CD-ROM.

Vendler 1957 = Vendler, Z. 1957. 'Verbs and Times', in *The Philosophical Review* 66:2, 143–160.

Wichmann 2008 = Wichmann, S. 2008. 'The study of semantic alignment: retrospect and state of the art', in Donohue, M. & Wichmann, S. (eds.). *The Typology of Semantic Alignment*, 3–23. Oxford University Press: Oxford.

Vielliard 1927 = Vielliard, J. 1927. *Le latin des diplômes royaux et chartes privées de l'époque mérovingienne*. Librairie Ancienne Honoré Champion: Paris.

Vincent 1982 = Vincent, N. 1982. 'The development of the auxiliaries *habere* and *esse* in Romance', in Vincent, N. & Harris, M. (eds.) *Studies in the Romance Verb*. London: Croom Helm.

Vincent 1997 = Vincent, N. 1997. 'The emergence of the D-system in Romance', in van Kemenade A. & Vincent, N. (eds.) *Parameters of Morphosyntactic Change*, 149–169. Cambridge: Cambridge University Press.

Winter 1971 = Winter, W. 1971. 'Formal Frequency and Linguistic Change: some preliminary comments', *Folia Linguistica* 5, 55–61.

Väänänen 1965 = Väänänen, V. 1965. *Étude sur le texte et la langue des Tablettes Albertini*. Helsinki: Suomalainen Tiedeakatemia.

Väänänen 1966 = Väänänen, V. 1966. *Le latin vulgaire des inscriptions pompéiennes*. Akademie-Verlag: Berlin.

Väänänen 1981 = Väänänen, V. 1981. *Introduction au latin vulgaire*. Bibliothèque française et romane. Série A:6. Éditions Klincksieck: Paris.

Väänänen 1987 = Väänänen, V. 1987. *Le journal-épître d'Égérie (Itinerarium Egeriae): étude linguistique*. Helsinki: Suomalainen Tiedeakatemia.

Zamboni 1998a = Zamboni, A. 1998. 'Cambiamento di lingua o cambiameno di sistema? Per un bilancio cronologico della transizione', in Herman, J. (ed.). *La transizione dal latino alle lingue romanze: atti della Tavola Rotonda di Linguistica Storica*, Venezia, 1996, 99–127. Tübingen: Niemeyer.

Zamboni 1998b = Zamboni, Alberto. 1998. 'Dal latino tardo agli albori romanzi: dinamiche linguistiche della transizione', in *Morfologie sociali e culturali in Europa fra tarda antichità e alto medioevo*: *Settimane di Studio del Centro Italiano di Studi sull'Alto Medioevo* 45, 619–702. CISAM: Spoleto.

Zamboni 2000a = Zamboni, A. 2000. *Alle origini dell'italiano: dinamiche e tipologie della transizione dal latino*. Carocci: Roma.

Zamboni 2000b = Zamboni, A. 2000. 'L'emergere dell'italiano: per un bilancio aggiornato', in *La preistoria dell'italiano*. Atti della Tavola Rotonda di Linguistica Storica. Università Ca' Foscari di Venezia, 1999, 231–260. Niemeyer: Tübingen.

Zamboni 2002 = Zamboni, A. 2002. 'I dialetti e le loro origini', in Cortelazzo, M. & al. (eds.). *I dialetti italiani. Storia, Struttura*, 3–25. UTET: Torino.

Zielinski 1972 = Zielinski, H. 1972. *Studien zu den spoletinischen "Privaturkunden" des 8. Jahrhunderts und ihrer überlieferung im Regestum farfense*. Niemeyer: Tübingen.

Zilliacus 1941 = Zilliacus, Henrik. 1941. 'Griechische Papyrusurkunden des VII. Jahrhunderts n. Chr.', *Eranos* 38, 79–107.

Appendix 1.1. MED 395 (ChLA 74:10). *Charta venditionis*. Lucca, AD 815. Archivio Storico Diocesano di Lucca. Written by Rumualdus *clericus notarius*. Austrifonsus, the archdeacon, sells to Iltroda or Eltroda, the nun, a vineyard in a place called Insula, near Lunata (near modern Luni), for fifty *solidi*.

{1} + In ~~nom(ine)~~ Patris et Filii et Sp(iritu)s S(an)c(t)i {2} regnante ~~d(om)n(o)~~ n(ost)ro Hlodowich(us) serenissim(us) august(us), a D(e)o coronat(us) magn(us) et pacificus imperator anno secundo et ~~d(om)n(o)~~ n(ost)ro Bernard(us) rex Langobardor(um) in D(e)i ~~nom(ine)~~, postquam in Etalia reversus est, anno regni ei(us) tertio, sexto decimo ~~kal(endas)~~ magias, ~~ind(ictione)~~ octava. {3} <u>Constat me Austrifonsus arcidiac(o)n(us) filio ~~b(one) m(emorie)~~ Teutprandi presenti die p(er) hanc cartul(am) vendere et tradere videor tibi Iltruda Dei ancilla filia ~~b(one) m(emorie)~~ Argimi,</u> {4} <u>id est</u> **una petia de vinea mea, quam abeo in Insola prope Lunata, qui mihi p(er) duas cartulas obvine a Titulo; una p(er) viganationis, alia p(er) venditionis. Et in ipsa viganationis dedi ego ei terra, quem ego abui de iura parentor(um) meor(um).** {5} Ipsa vinea una cum arborib(us) suis, **sicut in ipse cartule legitur et p(er) s(upra)s(crip)te cartole est circumdata**, tibi ea vendo et trado in ~~integr(um)~~ **una cum p(re)d(ic)te ambe** ~~cartul(e)~~. {6} Et recepi a te pretium pro s(upra)s(crip)ta vinea et ~~cartul(e)~~ argentum ~~sol(idos)~~ quinquaginta in prefinito. {7} Unde repromitto ego ~~q(ui) s(upra)~~ Austrifonsus ~~arcidiac(o)n(us)~~ una cum meis heredib(us) tibi Iltruda D(e)i ancilla, ut si nos tibi ipsa iam dicta vinea, quas tibi supra venundavi aut prefate cartule vob(is) intentionaverim(us) aut subtragi quesierim(us) p(er) quolibet ingenium nos ~~[vel ille]~~ h[om]o, cui nos ea(m) dedissem(us) aut dederim(us), spondim(us) nos tibi comp(onere) s(upra)s(crip)ta vinea [.....] ~~cartul(as)~~ in duplo in ferquidem loco sub extimatione, quales tunc fuerit. {8} Nam ab alio homine nos tibi exinde autores nec defensatores tibi esse non debeam(us), set tu p(er) temedipsa cum procuratore tuo s(upra)s(crip)ta vinea defendere debeatis tam cum ista ~~cartul(a)~~ quam et cum ille alie s(upra)s(crip)te cartule, qualiter meli(us) potueritis. {9} Et Rumualdum ~~cl(ericum) not(arium)~~ scribere rogavi. {10} Actum Luca.
{11} + <u>Ego Ostrifusus arcidiac(o)n(us) in ac cartula a me facta manu mea subs(cripsi).</u>
+ <u>Ego Atrip(er)tus rogatus ec.</u>
~~Sign(um)~~ + m(anu)s Asp(e)rti filio Aliprandi ~~mon(e)t(arii)~~ testis.
+ <u>Ego Cristianus rogatus ec.</u>
~~Sign(um)~~ + m(anu)s Iohanni filio Ghisi testis.
{12} + <u>Ego Rumuald(us)</u> ~~cl(ericus) not(arius)~~ <u>post tradita complevi et dedi.</u>

- The expanded abbreviations are in round brackets () while those expanded abbreviations that are not included in the morphological analysis are crossed out.
- The damaged words and letters are in square brackets [] while those damaged words that are not included in the morphological analysis are crossed out.

Diplomatic segmentation of the above charter:

{1} *invocatio* (formulaic); {2} *datatio* (formulaic); {3} *notificatio* (formulaic); {4–6} *dispositio*: {4} *dispositio* proper (non-formulaic); {5} affirmation clause (partly non-formulaic); {6} *pretium* clause (formulaic); {7–8} further affirmation clauses: {7} *sanctio* (formulaic); {8} *defensio* clause (formulaic); {9} *rogatio* (formulaic); {10} *actum* clause (formulaic); {11} *subscriptio* (formulaic); {12} *completio* (formulaic)

- The **bold** text indicates the non-formulaic sequences (not separate words) while the <u>underlined</u> text inside the *dispositio* shows the sentences that I annotate as non-formulaic. The <u>underlined</u> subscriptions are autographs.

English translation:

{1} +[338] In the name of the Father and the Son and the Holy Spirit {2} in the second year of the reign of our Lord Hlodowichus, most distinguished *augustus*, great and pacific emperor crowned by God, and in the third year of his reign, after his return to Italy, of our Lord Bernardus, in the name of God king of the Lombards; on the sixteenth day after the Kalends of May, under the eighth indiction. {3} It is manifest that I, Austrifonsus, archdeacon, son of the late Teutprandus, in the present day by this charter sell and hand over to you, Iltruda, servant of God, daughter of the late Argimus, {4} namely one part of my vineyard that I have at Insola near Lunata and that came to me from Titulus by two charters: the one was an exchange, the other was a purchase. And in the exchange, I gave him the plot I had from the heritage of my parents. {5} The vineyard with its trees, as it says in those charters and as it is surrounded by the above-mentioned charters,[339] it I sell and hand over to you with the both above-mentioned charters. {6} And I received from you, as the price of the above-mentioned vineyard and charters, fifty silver *solidi*, as was agreed. {7} Therefore, I, the above-mentioned Austrifonsus, archdeacon, together with my heirs – we or that man to whom we give or are going to give it[340] – promise to you, Iltruda, servant of God, that if we, by whichever artifice, aspire to or try to dispossess of the above-mentioned vineyard, which I sold you, and the above-mentioned charters, then we guarantee to compensate to you twice the price of the above-mentioned vineyard [...and the...] charters in the same place under estimation of their actual value. {8} Instead, we need not be your guarantors or legal defenders against other man, but you, with your procurator, must yourself defend the above-mentioned vineyard how you best can both with the help of this charter and of those other above-mentioned charters. {9} And I asked Rumualdus, clerk and notary, to write.[341] {10} Completed at Lucca.
{11} + I, Ostrifusus[342], archdeacon, subscribed this charter made by me in my own hand.
+ I, Atripertus, who was asked etc.[343]
Sign + of the hand of Aspertus, son of Aliprandus, banker, witness.
+ I, Cristianus, who was asked, etc.
Sign + of the hand of Iohannis, son of Ghisus, witness.
{12} + I, Rumualdus, clerk and notary, completed and gave after the tradition.[344]

- Other charter types may contain additional diplomatic elements, such as arengas (donations, wills), salutation formulae (mainly donations) and further *sanctio* clauses (donations, sales contracts, exchanges). The *notitiae iudicati* and *breves* contain elements of their own, e.g. the opening clause *resedente me*.

---

[338] Sign of the Holy Cross (*signum sanctae crucis*)
[339] The scribe has confused the sentence concerning the two charters with a common formula *qualiter circumdata est per designatas locas*.
[340] should be 'them'
[341] this charter
[342] Austrifonsus (the scribe uses a more Latinate variant of the name)
[343] These formulaic subscriptions of this type have been truncated in Barsocchini's edition. The complete wording is: [--] who was asked by Ostrifusus, archdeacon, subscribed as a witness.
[344] "completed the transaction and gave this instrument to the pertinent parties"; *traditio* is an old Roman chancery formality.

Appendix 2.1. The 20 most frequent word forms of LLCT (with ‰ of the total 198,696).

	Word forms				Lemmas		
range	word	N	‰	range	lemma	N	‰
1	*et*	10,346	52.1	1	*et*	10,346	52.1
2	*in*	7,081	35.6	2	*in*	7,081	35.6
3	*ego*	3,252	16.4	3	*ego*	5,133	25.8
4	*de*	2,306	11.6	4	*qui*	3,791	19.1
5	*qui*	2,028	10.2	5	*sum*	3,641	18.3
6	*manus*	1,626	8.2	6	*ipse*	3,299	16.6
7	*ipsa*	1,418	7.1	7	*meus*	2,835	14.3
8	*signum*	1,384	7.0	8	*tu*	2,769	13.9
9	*est*	1,381	7.0	9	*de*	2,306	11.6
10	*vel*	1,371	6.9	10	*is*	2,203	11.1
11	*sancti*	1,243	6.3	11	*ad*	2,176	11.0
12	*me*	1,236	6.2	12	*omnis*	2,082	10.5
13	*filio*	1,235	6.2	13	*hic*	1,979	10.0
14	*ut*	1,205	6.1	14	*vel*	1,931	9.7
15	*a*	1,171	5.9	15	*ecclesia*	1,918	9.7
16	*tibi*	1,102	5.5	16	*manus*	1,912	9.6
17	*ad*	246	1.2	17	*filius*	1,898	9.6
18	*per*	205	1.0	18	*rogo*	1,889	9.5
19	*cum*	132	0.7	19	*ab*	1,826	9.2
20	*nos*	124	0.6	20	*per*	205	1.0
up to 20		40,092	201.8	up to 20		61,220	308.1

Appendix 2.2. Syntactic tree of sentence LLCT 2936132 (CDL 166) drawn by TrEd Tree Editor.
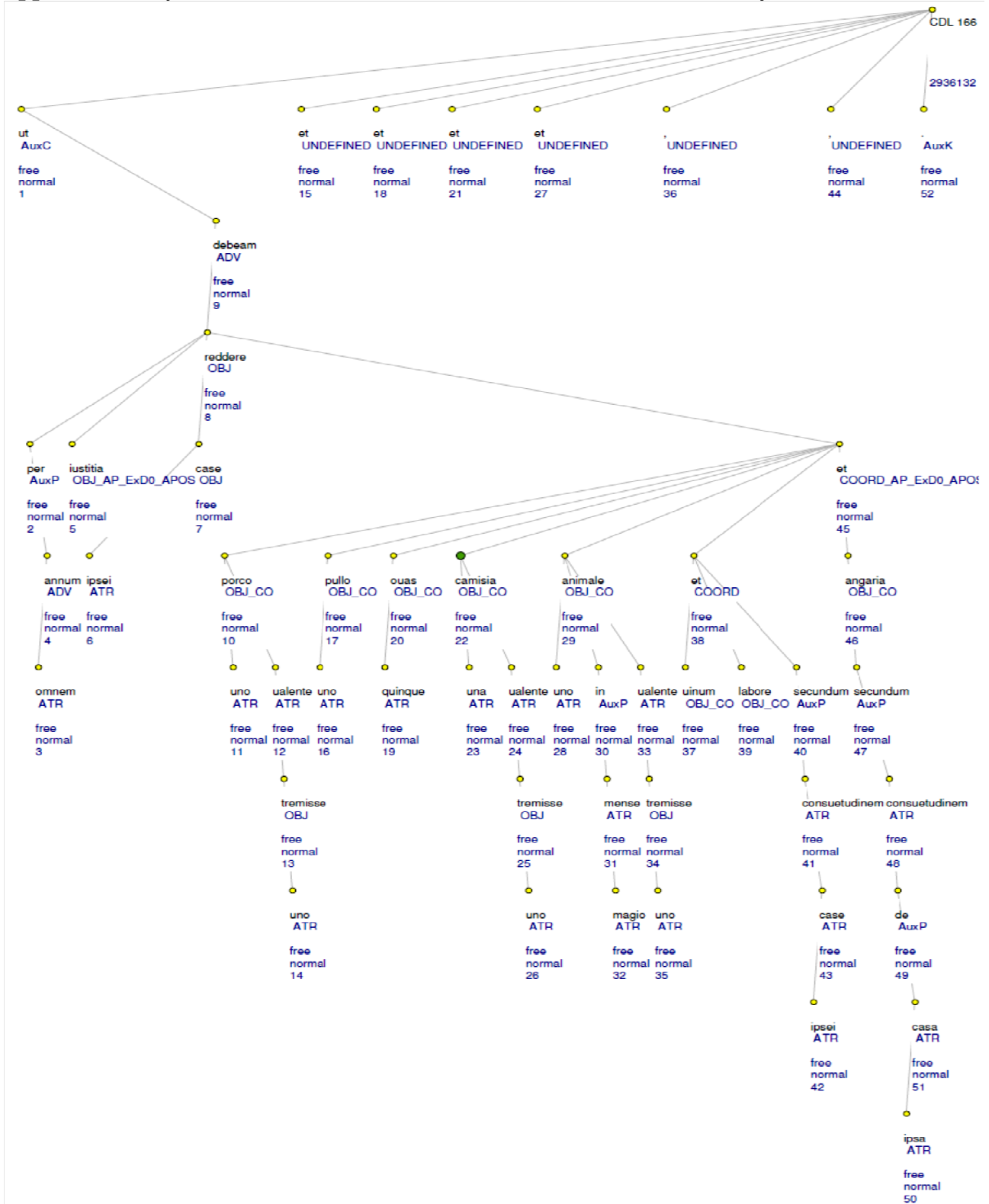
CDL 166 (AD 762) *ut per omnem annum iustitia ipsei case reddere debeam porco uno ualente tremisse uno et uno pullo et quinque ouas et camisia una ualente tremisse uno et uno animale in mense magio ualente tremisse uno, uinum et labore secundum consuetudinem ipsei case et angaria secundum consuetudinem de ipsa casa*

"so that I settle every year as the rent of that house one pig worth a *tremissis* and one hen and five eggs and one shirt worth a *tremissis* and in May one sheep [?] worth a *tremissis*, vine and corn according to the convention of the house and the *corvée* according to the convention of the house"

Appendix 4.1. Dependence between the case and animacy of the subject (singular and plural apart).

Case		Singular				Plural		
		Animacy			Total	Animacy		Total
		inanimate	animate	personal		inanimate	animate	
nominative	N	289	171	402	862	56	114	170
	%	70.3%	84.2%	56.2%	64.9%	64.4%	72.6%	69.7%
	residual	2.8	6.3	-7.1		-1.3	1.3	
accusative	N	122	32	313	467	31	43	74
	%	**29.7%**	**16%**	**43.8%**	35.1%	**36%**	**27%**	30%
	residual	-2.8	-6.3	7.1		1.3	-1.3	
Total	N	411	203	715	1,329	87	157	244
Chi-square		$\chi^2 = 62.21, df = 2, p < 0.001$				$\chi^2 = 1.80, df = 1, p = 0.18$		

Appendix 4.2. Dependence between the case of the subject and the construction type (singular and plural apart).

Case		Singular					Plural				
		Construction type				Total	Construction type				Total
		A	S_A	S_O	S_P		A	S_A	S_O	S_P	
nom.	N	431	47	323	61	862	70	18	61	21	170
	%	66.7%	75%	60.6%	70%	64.9%	76%	75%	61%	75%	69.7%
	residual	1.4	1.7	-2.7	1.1		1.7	0.6	-2.5	0.7	
acc.	N	215	16	210	26	467	22	6	39	7	74
	%	**33.3%**	**25%**	**39.4%**	**30%**	35.1%	**24%**	**c.25%**	**39%**	**c.25%**	30%
	residual	-1.4	-1.7	2.7	-1.1		-1.7	-0.6	2.5	-0.7	
Total	N	646	63	533	87	1,329	92	24	100	28	244
Chi-square		$\chi^2 = 8.90, df = 3, p = 0.031$					$\chi^2 = 6.05, df = 3, p = 0.109$				

$S_P = S_O$ passive

This appendix presents first the following summary table of the transitivity values of each examined subclass. After the summary table, the analyses of the 471 examined clauses are given.

clause type	animacy	sample size	case	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ %
Singular														
unacc.	inanim.	25/182	nom.	0	0	12	4	0	100	20	0	0	0	14
		25/86	acc.	0	0	0	0	0	100	28	0	0	0	13
	animate	17/17	nom.	0	0	18	18	0	82	24	0	0	0	14
		2/2	acc.	-	-	-	-	-	-	-	-	-	-	-
	personal	25/115	nom.	0	0	4	4	0	100	84	0	0	0	19
		25/111	acc.	0	0	0	0	0	100	76	0	0	0	18
unerg.	animate	24/24	nom.	0	100	75	0	100	71	21	71	0	0	44
		2/2	acc.	-	-	-	-	-	-	-	-	-	-	-
	personal	15/15	nom.	0	100	80	0	100	100	93	100	0	0	57
		12/12	acc.	0	100	75	0	100	100	75	75	0	0	53
trans.	inanim.	25/49	nom.	60	68	64	8	68	88	64	68	8	52	55
		17/17	acc.	65	53	29	0	35	71	29	24	0	47	35
	animate	25/98	nom.	80	68	52	0	76	88	28	48	8	44	49
		19/19	acc.	95	63	47	0	58	89	26	32	11	47	47
	personal	25/219	nom.	72	92	84	0	92	100	72	92	8	44	66
		25/161	acc.	80	60	56	0	60	100	84	60	4	60	56
Plural														
unacc.	inanim.	25/25	nom.	0	0	0	0	0	100	16	0	0	0	12
		20/20	acc.	0	0	10	0	0	100	25	0	0	0	14
	animate	25/39	nom.	0	0	4	4	8	88	56	0	0	0	16
		18/18	acc.	0	0	0	0	0	100	28	0	0	0	13
unerg.	inanim.	6/4	nom.	-	-	-	-	-	-	-	-	-	-	-
		5/2	acc.	-	-	-	-	-	-	-	-	-	-	-
	animate	15/15	nom.	0	100	60	0	100	87	13	80	0	0	44
		4/3	acc.	-	-	-	-	-	-	-	-	-	-	-
trans.	inanim.	5/5	nom.	-	-	-	-	-	-	-	-	-	-	-
		2/2	acc.	-	-	-	-	-	-	-	-	-	-	-
	animate	25/52	nom.	88	60	48	4	60	84	36	52	12	28	47
		19/19	acc.	95	47	74	5.3	47	89	5.3	37	5.3	26	43

Clauses with singular subjects

A. Unaccusative clauses

A1. Nominative-form inanimate singular subjects of unaccusative clauses (25 in 182)

A1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
1.	advenio	arrive	die	advenerit		0	0	1	0	0	1	0	0	0	0	2
2.	debeo	have to (be)	pars	debeat	esse	0	0	0	0	0	1	0	0	0	0	1
3.	illuceo	dawn	dies	inluxerit		0	0	1	0	0	1	0	0	0	0	2
4.	occurro	occur/arrive	finis	hoccurra		0	0	1	1	0	1	0	0	0	0	3
5.	permaneo	remain	manus	permaneat		0	0	0	0	0	1	0	0	0	0	1
6.	permaneo	remain	donatio	permaneat		0	0	0	0	0	1	0	0	0	0	1
7.	remaneo	remain	res	remansere		0	0	0	0	0	1	0	0	0	0	1
8.	sum	be	autilitas	fuerit		0	0	0	0	0	1	0	0	0	0	1
9.	sum	be	autilitas	fuerit		0	0	0	0	0	1	0	0	0	0	1
10.	sum	be	consuetudo	fuit		0	0	0	0	0	1	1	0	0	0	2
11.	sum	be	demandatio	fuet		0	0	0	0	0	1	1	0	0	0	2
12.	sum	be	domus	est		0	0	0	0	0	1	1	0	0	0	2
13.	sum	be	domus	est		0	0	0	0	0	1	1	0	0	0	2
14.	sum	be	medietas	sit		0	0	0	0	0	1	0	0	0	0	1
15.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
16.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
17.	sum	be	uirtus	fuerit		0	0	0	0	0	1	0	0	0	0	1
18.	sum	be	utilitas	fuere		0	0	0	0	0	1	0	0	0	0	1
19.	sum	be	utilitas	fuerit		0	0	0	0	0	1	0	0	0	0	1
20.	sum	be	utilitas	fuerit		0	0	0	0	0	1	0	0	0	0	1
21.	sum	be	volumptas	fuere		0	0	0	0	0	1	0	0	0	0	1
22.	sum	be	voluntas	fuerit		0	0	0	0	0	1	0	0	0	0	1
23.	sum	be	res	sint		0	0	0	0	0	1	0	0	0	0	1
24.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
25.	sum	be	veritas	est		0	0	0	0	0	1	1	0	0	0	2
					Σ	0	0	3	1	0	25	5	0	0	0	34
					%	0	0	12	4	0	100	20	0	0	0	14

1. CDT 46 cum <u>die</u> accessionis nostre advenerit
2. CDL 114 una <u>pars</u> de ipse duo portionis ... debeat esse in senodocio
3. CDL 175 dum cunctis <u>inluxerit</u> dies ille tremendus
4. CDT 12 qualiter mihi <u>finis</u> mortis hoccurra [= occurrat] non iscio
5. CDL 99 et hanc <u>manus</u> in suo rouore permaneat
6. MED 177 omni in tempore in vos ipse mea <u>donatio</u> firme et istavile permaneat
7. CDL 214 et si ... <u>res</u> mea a me iterum iniudicata uel non data remansere
8. MED 351 ad quem vobis inibi <u>autilitas</u> fuerit
9. MED 758 ad que vobis <u>autilitas</u> fuerit per singulos annos
10. MED 418 quale <u>consuetudo</u> fuit reddere de suprascripta angaria

11. CDL 113 et qualiter ipsius <u>demandatio</u> fuet conpliui et dedi

12. MED 213 ecclesiae beatissimi sancti Martini ubi est <u>domus</u> episcoporum Lucense

13. MED 397 ecclesie sancti Martini ubi est <u>domus</u> episcoporum

14. MED 797 suprascripta <u>medietas</u> de vinea in mea sit potestatem diebus vite mee abendi

15. CDL 145 omnia suprascripta <u>res</u> in mea sit potestatem usufructando

16. MED 178 omnia suprascripta <u>res</u>, dum advixero, in mea sit potestatem regendi

17. CDL 256 et eos gubernare debeam, ut <u>uirtus</u> mea fuerit

18. CDL 261 et si eorum <u>utilitas</u> fuere

19. MED 468 at que tibi <u>utilitas</u> fuerit

20. MED 763 ad que <u>utilitas</u> fuerit ad ipsa plebe

21. CDT 15 et si eius fuere <u>volumptas</u> recipiendum da ipsus abbas

22. MED 788 si vestra fuerit <u>voluntas</u> exinde causas agendi

23. MED 240 omnis suprascripta <u>res</u> et monasteria cum rebus suis in mea sint potestatem regendum

24. MED 787 suprascripta casa et <u>res</u>, dum vita mea fuerit, in meam sit potestatem abendi

25. MED 244 sic manifestavi et sic est <u>veritas</u> et manu mea subscripsi

A2. Accusative-form inanimate singular subjects of unaccusative clauses (25 in 86)

A2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
26.	accresco	increase	mercidem	adcriscat		0	0	0	0	0	1	0	0	0	0	1
27.	contingo	fall to	portione	continxit		0	0	0	0	0	1	1	0	0	0	2
28.	debeo	have to (remain)	offerationem	diueam	permanire	0	0	0	0	0	1	0	0	0	0	1
29.	percurro	be customary	quale	percurrit		0	0	0	0	0	1	1	0	0	0	2
30.	permaneo	remain	donationem	permaneat		0	0	0	0	0	1	0	0	0	0	1
31.	permaneo	remain	portionem	permaneat		0	0	0	0	0	1	0	0	0	0	1
32.	sum	be	binditiones	sit		0	0	0	0	0	1	0	0	0	0	1
33.	sum	be	domo	est		0	0	0	0	0	1	1	0	0	0	2
34.	sum	be	domo	est		0	0	0	0	0	1	1	0	0	0	2
35.	sum	be	fini	est		0	0	0	0	0	1	1	0	0	0	2
36.	sum	be	medietatem	sit		0	0	0	0	0	1	0	0	0	0	1
37.	sum	be	partem	sint		0	0	0	0	0	1	0	0	0	0	1
38.	sum	be	portione	sit		0	0	0	0	0	1	0	0	0	0	1
39.	sum	be	portionem	sit		0	0	0	0	0	1	0	0	0	0	1
40.	sum	be	portionem	fuit		0	0	0	0	0	1	1	0	0	0	2
41.	sum	be	quales	fueret		0	0	0	0	0	1	0	0	0	0	1
42.	sum	be	ueritate	fuisset		0	0	0	0	0	1	0	0	0	0	1
43.	sum	be	usumfructu	sit		0	0	0	0	0	1	0	0	0	0	1
44.	sum	be	curte	sint		0	0	0	0	0	1	0	0	0	0	1
45.	sum	be	fine	sunt		0	0	0	0	0	1	1	0	0	0	2
46.	sum	be	ortum	fuerit		0	0	0	0	0	1	0	0	0	0	1
47.	sum	be	potestatem	sit		0	0	0	0	0	1	0	0	0	0	1
48.	sum	be	sepe	sint		0	0	0	0	0	1	0	0	0	0	1
49.	sum	be	medietate	sit		0	0	0	0	0	1	0	0	0	0	1
50.	sum	be	voluntate	fuerit		0	0	0	0	0	1	0	0	0	0	1
				Σ		0	0	0	0	0	25	7	0	0	0	32
				%		0	0	0	0	0	100	28	0	0	0	13

26. CDL 31 ut de paruitatis rebus nostris nouis <u>mercidem</u> adcriscat
27. CDL 287 simul et quale <u>portione</u> ipsius [= ipsi] filio meo continxit de monasterio sancte Cristine
28. CDL 211 ipse mea <u>offerationem</u> ferma et istabile diueam permanire
29. MED 418 una anfora de vino bono sine aqua ad tale mensura <u>quale</u> in suprascripto loco percurrit
30. CDL 58 ut ipsa donationem nostra firma et stauilita <u>permaneat</u>
31. MED 238 ut ab hodierna die prefata <u>portionem</u> ... permaneat in potestatem sepedicte Dei ecclesie
32. CDT 65 suprascripta <u>binditiones</u> mea ... in tua qui supra domno Ermari abbati ... sit potestate
33. CDL 132 eclesie sancti Martini ubi est <u>domo</u> episcopii
34. CDL 200 ecclesiae beati sancti Martini ubi est <u>domo</u> episcoporum
35. MED 200 et est in terre <u>fini</u> ab uno latere terra Saxuli
36. CDL 281 reliquam uero <u>medietatem</u> rem meam [= rei meae] sit in potestate coniugi meae Teusprandae
37. CDL APPENDICE <u>partem</u> ipsorum cui exinde postea ipse Deusdona presbiter cartulam emisit sint exinde contemti et remoti
38. MED 308 omnia iam dicta mea <u>portione</u> ... in prenominata Dei ecclesia et de eius rectoribus sit potestatem habendi
39. CDL 239 nam meam <u>portionem</u>, ut dixi, in integrum sit in potestate suprascripte ecclesiae in prefinito
40. MED 172 in loco Vicinia quantum ibidem mea fuit <u>portionem</u> in omnibus
41. CDL 80 sum [= sub] stimationem [= aestimatione] <u>quales</u> tunc fueret
42. CDL APPENDICE ut nos certam poterimus iungere ueritatem ... si fuisset <u>ueritate</u>
43. CDL 27 sic tamen ut, dum die uite me fueret, in mea sit potestatem <u>usumfructu</u>
44. MED 576 predicta <u>curte</u> seo casis et omnibus rebus ad eas pertinentibus in mea ... sint potestatem habendi
45. CDL 111 et da finis sunt <u>fine</u> sancti Richuli et fine de Tricchase et fine de Aque Albule
46. CDL 237 et si plus fuerit fundamentum ipsius Sprincae et <u>ortum</u>
47. CDL 100 ut post meo obitum in gremio ipsius ecclesie et monasthERIO sit <u>potestatem</u>
48. MED 789 suprascripta ecclesia sancti Nazari ... sive <u>sepe</u> [= saepem] ... in mea ... sint potestatem abendum
49. MED 274 omnia <u>medietate</u> rebus meis ... sit in potestate de iam dicta Dei ecclesia vel de recture eius
50. MED 277 et si Dei omnipotenti fuerit <u>voluntate</u> et de hoc seculo migratus fuero

A3. Nominative-form animate singular subjects of unaccusative clauses (all 17)

A3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
51.	advivo	live on	soror	advixerimus		0	0	0	0	0	1	0	0	0	0	1
52.	habito	dwell	homo	avitavero		0	0	0	0	0	1	0	0	0	0	1
53.	mereo	deserve (to live on)	unus	merueremus	aduiuere	0	0	0	0	0	1	0	0	0	0	1
54.	mereo	deserve (to live on)	unus	meruerimus	aduiuere	0	0	0	0	0	1	0	0	0	0	1
55.	migro	migrate (= die)	abbas	migraueret		0	0	1	1	0	1	0	0	0	0	3
56.	migro	migrate (= die)	abbas	migraueret		0	0	1	1	0	1	0	0	0	0	3
57.	praesumo	dare (to dwell)	sacerdos	presumat	auitare	0	0	0	0	0	0	0	0	0	0	0
58.	praesumo	dare (to dwell)	sacerdus	presummat	habitare	0	0	0	0	0	0	0	0	0	0	0
59.	recedo	recede (= die)	unus	recesseret		0	0	1	1	0	1	0	0	0	0	3
60.	resideo	reside	genitor	residet		0	0	0	0	0	1	1	0	0	0	2
61.	resideo	reside	pater	reside		0	0	0	0	0	1	1	0	0	0	2
62.	resideo	reside	genitor	residde		0	0	0	0	0	1	1	0	0	0	2
63.	sum	be	genitor	fuisset		0	0	0	0	0	1	0	0	0	0	1
64.	sum	be	nullus	fuerit		0	0	0	0	0	0	0	0	0	0	0
65.	video	seem (to dwell)	genitor	visus	avitare	0	0	0	0	0	1	0	0	0	0	1
66.	vivo	live	homo	uiuit		0	0	0	0	0	1	1	0	0	0	2
67.	volo	want (to stay)	presbiter	voluerit	conversare	0	0	0	0	0	1	0	0	0	0	1
					Σ	0	0	3	3	0	14	4	0	0	0	24
					%	0	0	18	18	0	82	24	0	0	0	14

51. MED 374 dum Deo volente ..... <u>soror</u> mea advixerimus

52. MED 452 et [si] ille <u>homo</u>, qui in ipsa casa resederit, vel si ego ibi avitavero
53. CDL 222 dum <u>unus</u> ex nobis aduiuere merueremus
54. CDL 93 dum usque aduiuere meruerimus <u>unus</u> aut ambo
55. CDL 24 si ipse <u>abbas</u> custus de hac luce migraueret
56. CDL 28 quid si ... et <u>abbas</u> qui iniui constitutus fueret de hac luce migraueret
57. CDL 178 nullus <u>sacerdos</u> iuidem auitare presumat
58. CDL 30 nec nullus <u>sacerdus</u> ibidem habitare presummat
59. CDL 138 et si qualiter Deo fuerit preceptione et <u>unus</u> aut plures de nus de seculo recesseret
60. CDL 273 in ipsa casa ubi ipsi <u>genitor</u> tuus antea residet
61. CDL 280 una ca<sa> mea quem haure uisum sum prope eclesia sancti Pauli ubi <u>pater</u> tuo reside antea
62. MED 461 in ipsa casa et res ubi antea ipsi <u>genitor</u> meus residde et ego ipsi usque modo avitavit
63. CDT 45 quod <u>genitor</u> noster servus fuisset sancte Marie
64. CDL 179 et si forsitan <u>nullus</u> ex filiis aut nepotibus meis fuerit qui dignus sit
65. MED 427 res illa ... ubi ipsi <u>genitor</u> meus avitare visus fuit
66. CDL 183 dum in hoc seculo uiuit <u>homo</u>
67. CDT 14 dum predictus <u>presbiter</u> voluerit in suprascripto monasterio conversare

A4. Accusative-form animate singular subjects of unaccusative clauses (all 2)

A4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
68.	recedo	recede (= die)	filio	recessissit		0	0	1	1	0	1	0	0	0	0	3
69.	vivo	live	coniuge	uixere		0	0	0	0	0	1	0	0	0	0	1
					Σ	0	0	1	1	0	2	0	0	0	0	m. 2

68. CDL 186 et sicut Redemturi meo fuit iussionem et ipse <u>filio</u> meo de seculo recessissit
69. CDL 171 et si ipsa <u>coniuge</u> mea super me uixere et lecto meo costodierit

A5. Nominative-form personal-name singular subjects of unaccusative clauses (25 in 115)

A5	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
70.	adsum	be present	Alpulus	adfuit		0	0	0	0	0	1	1	0	0	0	2
71.	adsum	be present	Agiprandus	aderant		0	0	0	0	0	1	1	0	0	0	2
72.	adsum	be present	Alpertus	aderant		0	0	0	0	0	1	1	0	0	0	2
73.	adsum	be present	Alpulus	aderant		0	0	0	0	0	1	1	0	0	0	2
74.	adsum	be present	Arochis	aderant		0	0	0	0	0	1	1	0	0	0	2
75.	adsum	be present	Cristianus	aderant		0	0	0	0	0	1	1	0	0	0	2
76.	adsum	be present	Flaipertus	aderant		0	0	0	0	0	1	1	0	0	0	2
77.	adsum	be present	Frotpaldus	aderant		0	0	0	0	0	1	1	0	0	0	2
78.	adsum	be present	Gumpertus	aderant		0	0	0	0	0	1	1	0	0	0	2
79.	adsum	be present	Lamfridus	aderant		0	0	0	0	0	1	1	0	0	0	2
80.	adsum	be present	Petrisundus	aderant		0	0	0	0	0	1	1	0	0	0	2
81.	adsum	be present	Rachiprandus	aderant		0	0	0	0	0	1	1	0	0	0	2
82.	adsum	be present	Sanitas	aderant		0	0	0	0	0	1	1	0	0	0	2
83.	adsum	be present	Stavilis	aderant		0	0	0	0	0	1	1	0	0	0	2
84.	adsum	be present	Teddulus	aderant		0	0	0	0	0	1	1	0	0	0	2

	habito	dwell	Dammianus	habitavit		0	0	0	0	0	1	1	0	0	0	2
85.	habito	dwell	Dammianus	habitavit		0	0	0	0	0	1	1	0	0	0	2
86.	intersum	be involved	Paulus	interfui		0	0	0	0	0	1	1	0	0	0	2
87.	recedo	recede (= die)	Gumpertus	recessisse		0	0	1	1	0	1	0	0	0	0	3
88.	resideo	reside	Pipulus	resedit		0	0	0	0	0	1	1	0	0	0	2
89.	resideo	reside	Dammianus	residde		0	0	0	0	0	1	1	0	0	0	2
90.	sum	be	Pastor	fuit		0	0	0	0	0	1	1	0	0	0	2
91.	sum	be	Teupertus	sunt		0	0	0	0	0	1	1	0	0	0	2
92.	video	seem (to be)	Garimundus	uidetur	esse	0	0	0	0	0	1	0	0	0	0	1
93.	video	seem (to be)	Gumprandus	videtur	esse	0	0	0	0	0	1	0	0	0	0	1
94.	video	seem (to be)	Ostrifonsus	videtur	esse	0	0	0	0	0	1	0	0	0	0	1
				∑		0	0	1	1	0	25	21	0	0	0	48
				%		0	0	4	4	0	100	84	0	0	0	19

70. MED 385 resedente me in domo sanctae matris ecclesiae ... adfuit ante nos memoratus <u>Alpulus</u>
71. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est <u>Agiprandus</u> arcidiaconus
72. MED 385 ubi nobiscum aderant sacerdotes et filii sanctae ecclesie, id est ... <u>Alpertus</u> clericus
73. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Alpulus</u> presbiter
74. MED 385 ubi nobiscum aderant sacerdotes et filii sancte ecclesie, id est ... <u>Arochis</u> wassi domni nostri regis
75. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Cristianus</u> subdiaconus
76. MED 244 ubi aderant Sicheradus clericus ... <u>Flaipertus</u> et Ghisiprandus presbiter
77. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Frotpaldus</u> gastaldius
78. CDT 45 ubi aderant nobiscum Thomas diaconus ... <u>Gumpertus</u> filius quondam Ursi
79. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Lamfridus</u>
80. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Petrisundus</u>
81. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ...<u>Rachiprandus</u>
82. CDT 45 ubi aderant nobiscum Thomas diaconus ... <u>Sanitas</u> notarius
83. MED 385 ubi nobiscum aderant sacerdotes et filii sancte ecclesie, id est ... <u>Stavilis</u> presbiter
84. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Teddulus</u> presbiter
85. MED 202 casa ipsa ubi <u>Dammianus</u> presbiter socero meo habitavit
86. MED 564 <u>Paulus</u> notarius domni imperatoris interfui
87. MED 238 dum ... ipse <u>Gumpertus</u> de hoc seculo recessisse
88. MED 275 in una casa suprascripte ecclesie vestre ... ubi antea quondam <u>Pipulus</u> resedit
89. MED 202 quia ipsi <u>Dammianus</u> presbitero in ipsa basilica nostra sancti Petri residde
90. MED 410 hec fuit ante Sicherado clerico, <u>Pastor</u>, Trasimundus, Petrus, Firmo, Grauso, Petrus notarius
91. MED 549 id sunt <u>Teupertus</u> diaconus et Anspaldo clericus qui previderunt ac renuntiaverunt
92. CDL 114 eclesia sancte Reparate ubi <u>Garimundus</u> clericus custor esse uidetur
93. MED 307 ecclesie ... ubi nunc <u>Gumprandus</u> diaconus rector esse videtur
94. MED 380 ecclesie beati Frigiani ... ubi <u>Ostrifonsus</u> diaconus rector esse videtur

A6. Accusative-form personal-name singular subjects of unaccusative clauses (25 in 111)

A6	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	∑
95.	adsum	be present	Aliteu	aderant		0	0	0	0	0	1	1	0	0	0	2
96.	adsum	be present	Anispaldo	aderant		0	0	0	0	0	1	1	0	0	0	2
97.	adsum	be present	Ghiso	aderant		0	0	0	0	0	1	1	0	0	0	2
98.	adsum	be present	Gumprando	aderant		0	0	0	0	0	1	1	0	0	0	2
99.	adsum	be present	Lopo	aderant		0	0	0	0	0	1	1	0	0	0	2
100.	adsum	be present	Rachipertu	aderant		0	0	0	0	0	1	1	0	0	0	2
101.	adsum	be present	Teodingo	aderant		0	0	0	0	0	1	1	0	0	0	2

102.	adsum	be present	Walprando	aderant		0	0	0	0	0	1	1	0	0	0	2
103.	debeo	have to (dwell)	Atrifuso	debeamus	residere/abitare	0	0	0	0	0	1	0	0	0	0	1
104.	resideo	reside	Achipertulo	resede		0	0	0	0	0	1	1	0	0	0	2
105.	resideo	reside	Benedictulo	residde		0	0	0	0	0	1	1	0	0	0	2
106.	resideo	reside	Dammiano	residdet		0	0	0	0	0	1	1	0	0	0	2
107.	resideo	reside	Lucerulo	residet		0	0	0	0	0	1	1	0	0	0	2
108.	resideo	reside	Magnulo	residet		0	0	0	0	0	1	1	0	0	0	2
109.	resideo	reside	Petrulo	resedet		0	0	0	0	0	1	1	0	0	0	2
110.	resideo	reside	Bonulo	residdetur		0	0	0	0	0	1	1	0	0	0	2
111.	resideo	reside	Pertulo	residit		0	0	0	0	0	1	1	0	0	0	2
112.	sum	be	Firmo	fuit		0	0	0	0	0	1	1	0	0	0	2
113.	sum	be	Ardo	erant		0	0	0	0	0	1	1	0	0	0	2
114.	sum	be	Liutperto	erant		0	0	0	0	0	1	1	0	0	0	2
115.	video	seem (to head)	Alateo	uidetur	preesse	0	0	0	0	0	1	0	0	0	0	1
116.	video	seem (to be)	Aripaldo	videtur	esse	0	0	0	0	0	1	0	0	0	0	1
117.	video	seem (to be)	Ermipertu	videtur	esset	0	0	0	0	0	1	0	0	0	0	1
118.	video	seem (to be)	Iacobo	videtur	esse	0	0	0	0	0	1	0	0	0	0	1
119.	video	seem (to be)	Uernulo	uidetur	esse	0	0	0	0	0	1	0	0	0	0	1
					Σ	0	0	0	0	0	25	19	0	0	0	44
					%	0	0	0	0	0	100	76	0	0	0	18

95. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... Ghiso subdiaconus, <u>Aliteu</u> subdiaconus
96. MED 335 ubi nobiscum aderant Saripertus presbiter ... <u>Anispaldo</u> clericus
97. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae, id est ... <u>Ghiso</u> subdiaconus, Aliteu subdiaconus
98. CDT 45 ubi aderant nobiscum Thomas diaconus ... <u>Gumprando</u> et Walprando germanis
99. MED 385 ubi aderant nobiscum sacerdotes et filii sancte ecclesie, id est ... <u>Lopo</u>
100. CDT 45 ubi aderant nobiscum Thomas diaconus ... <u>Rachipertu</u> clericus
101. CDT 45 ubi aderant nobiscum Thomas diaconus ... <u>Teodingo</u> clericus
102. CDT 45 ubi aderant nobiscum Thomas diaconus ... Gumprando et <u>Walprando</u> germanis
103. MED 714 nos vel nostris heredibus sive <u>Atrifuso</u> germano nostro vel eius heredes ... in suprascriptis casis et rebus residere et abitare debeamus
104. MED 372 in terra, qui pertenit de casa nostra, ubi resede <u>Achipertulo</u>
105. MED 468 in unam casam massaricia ... ubi antea residde <u>Benedictulo</u>
106. MED 281 casa illa sancti Angeli in Monte ubi residdet quondam <u>Dammiano</u> presbitero
107. CDL 238 in casa ecclesiae uestrae in loco Lusciano ubi antea residet <u>Lucerulo</u> germanus Taniperti actori uestri
108. CDL 149 casa ... ubi residet <u>Magnulo</u> massarius noster
109. MED 694 una ex ipse est in Conflenti qui regitur per Dominico, illa alia est in Furnulo ubi resedet <u>Petrulo</u>
110. MED 589 id est casa et res mea illa ... ubi residdetur <u>Bonulo</u>
111. CDL 264 in casa ... ubi residit quondam <u>Pertulo</u>
112. MED 410 hec fuit ante Sicherado clerico, Pastor, Trasimundus, Petrus, <u>Firmo</u>, Grauso, Petrus notarius
113. MED 742 erantque nobiscum Ademarius, Teodemundus vassi idem augusti, <u>Ardo</u>, Cunimundus scavinis
114. MED 742 erantque nobiscum Offo, Minto, <u>Liutperto</u>, Rumualdo, Gisperto ... Iohannes et reliqui multis
115. CDL 171 eclesia ... ubi <u>Alateo</u> arcidiacono preesse uidetur
116. MED 741 ecclesie ... ubi modo <u>Aripaldo</u> presbitero rector esse videtur
117. MED 196 monasterio ... uvi vir venerabilis <u>Ermipertu</u> presbiter esset videtur
118. MED 306 ecclesie ... ubi ... <u>Iacobo</u> in Dei nomine episcopo esse videtur
119. CDL 157 eclesie beati sancti Pauli ubi presbiter <u>Uernulo</u> esse uidetur

278

B. Unergative clauses

B1. Nominative-form animate singular subjects of unergative clauses (all 24)

B1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
120.	appropinquo	approach	diabolus	adpropinquat		0	1	0	0	1	0	1	1	0	0	4
121.	appropinquo	approach	diauulus	adpropinquat		0	1	0	0	1	0	1	1	0	0	4
122.	appropinquo	approach	fur	adpropinquat		0	1	0	0	1	0	1	1	0	0	4
123.	appropinquo	approach	fur	adpropinquat		0	1	0	0	1	0	1	1	0	0	4
124.	debeo	have to (come)	homo	debeas	venire	0	1	1	0	1	1	0	0	0	0	4
125.	debeo	have to (pray)	sacerdos	deueas	deprecare	0	1	0	0	1	1	0	0	0	0	3
126.	debeo	have to (re-enter)	unus	debeat	reintroire	0	1	1	0	1	1	0	0	0	0	4
127.	eo	go	nullus	ire		0	1	1	0	1	0	0	1	0	0	4
128.	exeo	leave	dominus	exierit		0	1	1	0	1	1	0	1	0	0	5
129.	nolo	not want (to serve)	nullus	nolueret	seruire	0	1	0	0	1	0	0	0	0	0	2
130.	possum		homo	potuerit	introire	0	1	1	0	1	1	0	0	0	0	4
131.	praesumo		homo	presumpserimus/potuerimus	ire/defensare	0	1	1	0	1	1	0	0	0	0	4
132.	spondeo		filius	spondimus	esset venturi	0	1	1	0	1	0	1	1	0	0	5
133.	tento		homo	temtaueri	ire	0	1	1	0	1	1	0	0	0	0	4
134.	venio	come	homo	vinnere		0	1	1	0	1	1	0	1	0	0	5
135.	venio	come	homo	venerit		0	1	1	0	1	1	0	1	0	0	5
136.	venio	come	homo	venerit		0	1	1	0	1	1	0	1	0	0	5
137.	venio	come	rector	uenerit		0	1	1	0	1	1	0	1	0	0	5
138.	venio	come	homo	venerit		0	1	1	0	1	1	0	1	0	0	5
139.	venio	come	homo	venerit		0	1	1	0	1	1	0	1	0	0	5
140.	venio	come	homo	uineris		0	1	1	0	1	1	0	1	0	0	5
141.	venio	come	homo	uineris		0	1	1	0	1	1	0	1	0	0	5
142.	venio	come	homo	vineri		0	1	1	0	1	1	0	1	0	0	5
143.	venio	come	omo	uineri		0	1	1	0	1	1	0	1	0	0	5
					Σ	0	24	18	0	24	17	5	17	0	0	105
					%	0	100	75	0	100	71	21	71	0	0	44

120. CDL 30 sed thensaurizate uobis thesaurum in caelum ubi fur, id est <u>diabolus,</u> non adpropinquat
121. CDL 178 set thensaurizate uobis thensaurum in celo ubi fur, id est <u>diauulus</u>, non adpropinquat
122. CDL 30 sed thensaurizate uobis thesaurum in caelum ubi <u>fur</u>, id est diabolus, non adpropinquat
123. CDL 178 set thensaurizate uobis thensaurum in celo ubi <u>fur</u>, id est diauulus, non adpropinquat
124. MED 437 et ille <u>homo</u> ... semper ad mandato nostro venire debeas ad iustitiam faciendo
125. CDL 261 ut <u>sacerdos</u> ... mihi pro salute anime messarum precibus a Domino deprecare deueas
126. MED 497 tunc <u>unus</u> alterius in suum cambium reintroire debeat
127. CDT 46 <u>nullus</u> de heredibus, posteris vel parentibus nostris ne qualibe hominum genus ... contra hanc nostra decritione ire
128. CDL 146 et si ... alter <u>dominus</u> exierit
129. CDL 204 et si forsitans <u>nullus</u> de heredibus meis, ut supra dixi, in ipsa ecclesia seruire nolueret
130. MED 187 et si quicumque <u>homo</u> vobis in ipso monasterio ... in qualivet portionem introire potuerit
131. CDL 174 si quis amodo nos qui supra uinditor uel heredes nostros aut aliquis <u>homo</u> contra hanc uinditionem ... ire presumpserimus et menime ab omnem homine defensare potuerimus
132. CDL 205 numquam ego qui supra Ausulo nec <u>filius</u> heredes meus atuersus ipsa eclesia ... spondimus esset uenturi
133. CDL 92 si quis ego qui supra uinditor aud eridis mei aud aliquis <u>homo</u> contra hanc uinditione ... ire temtaueri

279

134. MED 181 excepto si quicunque <u>homo</u> vinnere
135. MED 384 nam si alter quisqumque <u>homo</u> venerit absque omni nostro conludio
136. MED 401 nam si alter quiscumque <u>homo</u> venerit absque nostro conludio
137. CDL 256 quando ad ipsum usumfructum ibidem uenerit <u>rector</u> ipsius ecclesiae uel missus eius
138. MED 437 et si ... ipsi <u>homo</u> qui in ipsa casa resederit ad mandato nostro venerit
139. MED 497 nam si alter quiscumque <u>homo</u> venerit absque nostro ..... vestro conludio
140. CDL 97 si quando nus qui uindituri sumus aut aliquis <u>homo</u> uineris
141. CDL 141 si quis amodo ego qui supra uinditor uel heredes meis aut aliquis <u>homo</u> uineris
142. CDT 25 si amodo ego qui supra vindetrix vel heredes meis aut aliquis <u>homo</u> vineri
143. CDL 185 si quis amodo ego qui supra uinditor uel heredes meis aut aliquis <u>omo</u> uineri

B2. Accusative-form animate singular subjects of unergative clauses (all 2)

B2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
144.	habeo	have to (pray)	sacerdote	auead	deprecare	0	1	0	0	1	1	0	0	0	0	3
145.	vado	go	mulierem	vadam		0	1	1	0	1	1	0	1	0	0	5
					Σ	0	2	1	0	2	2	0	1	0	0	m. 4

144. CDL 171 <u>sacerdote</u> ... pro anima mea de pondera peccata mea Deo deprecare auead Deo omni tempore
145. CDT 89 quam filiis qua <u>mulierem</u> tua vadam ubi volueris

B3. Nominative-form personal-name singular subjects of unergative clauses (all 15)

B3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
146.	aspiro	inspire	Deus	adspiraberit		0	1	0	0	1	1	0	1	0	0	4
147.	consigno	confirm	Arnulfus	consignabit		0	1	1	0	1	1	1	1	0	0	6
148.	dignor	condescend (to look at)	Dominus	dignatus	respicere	0	1	0	0	1	1	1	1	0	0	5
149.	loquor	speak	Dominus	loquitur		0	1	0	0	1	1	1	1	0	0	5
150.	praecipio	order	Rachinardus	preceperat		0	1	1	0	1	1	1	1	0	0	6
151.	praecipio	order	Rachinardus	preceperat		0	1	1	0	1	1	1	1	0	0	6
152.	subscribo	subscribe	Filuartus	subscripsi		0	1	1	0	1	1	1	1	0	0	6
153.	subscribo	subscribe	Barsucis	suscripsi		0	1	1	0	1	1	1	1	0	0	6
154.	venio	come	Alprandus	venit		0	1	1	0	1	1	1	1	0	0	6
155.	venio	come	Garipertus	venerunt		0	1	1	0	1	1	1	1	0	0	6
156.	venio	come	Arnulfus	venerunt		0	1	1	0	1	1	1	1	0	0	6
157.	venio	come	Benedictus	venit		0	1	1	0	1	1	1	1	0	0	6
158.	venio	come	Giselmarius	venerunt		0	1	1	0	1	1	1	1	0	0	6
159.	venio	come	Gisulfus	venerunt		0	1	1	0	1	1	1	1	0	0	6
160.	venio	come	Petrus	uenit		0	1	1	0	1	1	1	1	0	0	6
					Σ	0	15	12	0	15	15	14	15	0	0	86
					%	0	100	80	0	100	100	93	100	0	0	57

146. CDT 14 tantum si mihi <u>Deus</u> adspiraberit, ut in monasterio ... intrare voluero
147. CDT 45 sicut <u>Arnulfus</u> vicedomoi per testimonia consignabit

148. CDL 248 si mihi <u>Dominus</u> respicere dignatus fuerint et filios aut filias procreauero

149. CDL 178 unde <u>Dominus</u> per semetipsum loquitur dicens

150. MED 309 et ipse <u>Rachinardus</u> preceperat Harnolfo .....

151. MED 385 et ipse <u>Rachinardus</u> preceperat Arnolfi vicedomino

152. MED 781 <u>Filuartus</u> schavinus subscripsi

153. CDL 56 <u>Barsucis</u> uir clarissimus ciuis Lunensis ... testis suscripsi

154. MED 202 venit ante nos <u>Alprandus</u>

155. MED 774 ibique in nostra venerunt presencia, id est <u>Garipertus</u> diaconus

156. CDT 45 venerunt ibi ante nos <u>Arnulfus</u> vicedomoi nec non et Rotprandulu, Aspertulu clericus et Perticausulu

157. MED 397 ibique venit ante nos <u>Benedictus</u> advocato ecclesie sancti Martini ... et ex alia parte Suave

158. MED 564 venerunt ibi in nostri presencia <u>Giselmarius</u> vassus domni imperatoris

159. MED 742 ibique in nostri venerunt presentia, id est <u>Gisulfus</u> advocatus

160. CDL 255 qualiter uenit ante me Peredeo ... Alitroda mulier ... simul et <u>Petrus</u> clericus

B4. Accusative-form personal-name singular subjects of unergative clauses (all 12)

B4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
161.	accedo	access	Ghiso	accessit		0	1	1	0	1	1	1	1	0	0	6
162.	advenio	arrive	Liliopincto	advenit		0	1	1	0	1	1	1	1	0	0	6
163.	deservio	serve	Aspertulu	deservire		0	1	0	0	1	1	0	0	0	0	3
164.	deservio	serve	Perticausulu	deservire		0	1	0	0	1	1	0	0	0	0	3
165.	deservio	serve	Rotprandulo	deservire		0	1	0	0	1	1	0	0	0	0	3
166.	subscribo	subscribe	Fraimanno	subscripsi		0	1	1	0	1	1	1	1	0	0	6
167.	subscribo	subscribe	Fraiperto	subscripsi		0	1	1	0	1	1	1	1	0	0	6
168.	subscribo	subscribe	Luciprando	subscripsi		0	1	1	0	1	1	1	1	0	0	6
169.	venio	come	Aspertulu	venerunt		0	1	1	0	1	1	1	1	0	0	6
170.	venio	come	Perticausulu	venerunt		0	1	1	0	1	1	1	1	0	0	6
171.	venio	come	Rotprandulu	venerunt		0	1	1	0	1	1	1	1	0	0	6
172.	venio	come	Suave	venit		0	1	1	0	1	1	1	1	0	0	6
					Σ	0	12	9	0	12	12	9	9	0	0	63
					%	0	100	75	0	100	100	75	75	0	0	53

161. MED 188 ubi super hanc cambio ad partibus secundum legem accessit <u>Ghiso</u> misso nostro

162. MED 244 advenit <u>Liliopincto</u> presbiter

163. CDT 45 ut Rotprandulo, <u>Aspertulu</u> clerico et Perticausulu omni tempore deservire [= deservirent]

164. CDT 45 ut Rotprandulo, Aspertulu clerico et <u>Perticausulu</u> omni tempore deservire [= deservirent]

165. CDT 45 ut <u>Rotprandulo</u>, Aspertulu clerico et Perticausulu omni tempore deservire [= deservirent]

166. MED 594 ..... <u>Fraimanno</u> rogatus ab Aloni me teste subscripsi

167. MED 594 ..... <u>Fraiperto</u> rogatus ab Aloni me teste subscripsi

168. MED 544 ..... <u>Luciprando</u> presbiter in anc cartula a me facta manu mea subscripsi

169. CDT 45 venerunt ibi ante nos Arnulfus vicedomoi nec non et Rotprandulu, <u>Aspertulu</u> clericus et Perticausulu

170. CDT 45 venerunt ibi ante nos Arnulfus vicedomoi nec non et Rotprandulu, Aspertulu clericus et <u>Perticausulu</u>

171. CDT 45 venerunt ibi ante nos Arnulfus vicedomoi nec non et <u>Rotprandulu</u>, Aspertulu clericus et Perticausulu

172. MED 397 ibique venit ante nos Benedictus advocato ... et ex alia parte <u>Suave</u> altercationem inter se abentes

C. Transitive clauses

C1. Nominative-form inanimate singular subjects of transitive clauses (25 in 49)

C1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
173.	accommodo	reconcile	amor	adcommodauet	animum	1	1	1	0	1	1	1	1	0	1	8
174.	addo	add	animus	addedet	#	0	1	1	0	1	1	1	1	0	0	6
175.	addo	add	animus	atdidet	hoc	1	1	1	0	1	1	1	1	0	1	8
176.	admitto	permit	uirtus	admiset	#	0	1	1	0	1	1	1	1	0	0	6
177.	admitto	permit	uirtus	admiset	#	0	1	1	0	1	1	1	1	0	0	6
178.	audio	listen	pietas	audeas	me	1	1	0	0	1	1	0	0	0	1	5
179.	concambio	exchange	pars	concambiavemus	omnia	1	1	1	0	1	1	1	1	0	0	7
180.	contineo	contain	textus	contenet	cod	1	0	0	0	0	1	1	0	0	1	4
181.	contineo	contain	textus	contenit	quod	1	0	0	0	0	1	1	0	0	1	4
182.	contineo	contain	iussio	continet	#	0	0	0	0	0	1	1	0	0	0	2
183.	contineo	contain	ratio	contenit	ut clause	0	0	0	0	0	0	1	0	0	0	1
184.	contradico	contest	pars	contradicit	basilicam	1	1	1	0	1	0	1	1	0	1	7
185.	debeo	have to (have)	pars	deberet	abere casa/res	1	0	0	0	0	1	0	0	0	1	3
186.	decerno	order	animus	decreuet	ut clause	0	1	1	0	1	1	1	1	0	0	6
187.	declaro	declare	textus	declarat	#	0	1	1	0	1	1	1	1	0	0	6
188.	detineo	detain	pars	detineret	basilicam	1	0	0	0	1	1	0	1	0	1	5
189.	dico	say	pars	dicebat	direct quotation	0	1	1	0	1	1	1	1	0	0	6
190.	habeo	have	pars	habead	basilicam	1	0	0	0	0	1	0	0	0	1	3
191.	habeo	have (power)	pars	abeant	potestatem	1	0	0	0	0	0	0	0	0	0	1
192.	inquiro	enquire	pars	inquisivi	#	0	1	1	0	1	1	1	1	0	0	6
193.	observo	observe (= conquer)	mortis	obserueat	nus	1	1	1	1	1	1	0	1	1	1	9
194.	permitto	permit	uirtus	permiset	#	0	1	1	0	0	1	1	1	0	0	5
195.	occupo	conquer	mors	occupauerit	me	1	1	1	1	1	1	0	1	1	1	9
196.	refero	return	pars	retulerit	casam/rem	1	1	1	0	1	1	0	1	0	1	7
197.	volo	want (to elect)	congregatio	uoluerit	eligere que	1	1	1	0	1	1	0	1	0	1	7
					Σ	15	17	16	2	17	22	16	17	2	13	137
					%	60	68	64	8	68	88	64	68	8	52	55

173. CDL 25 <u>amor</u> superne uertutis meum animum adcommodauet pro meis facinoribus
174. CDL 62 sic tamen addedet <u>animus</u> meus Filimari
175. CDL 73 et hoc atdidet <u>animus</u> meus
176. CDL 127 ut <u>uirtus</u> admiset
177. CDL 138 ut <u>uirtus</u> admiset
178. CDL APPENDICE domine, audeas [= audiat] me <u>pietas</u> uestra
179. CDT 23 omnia unus ad alterium, <u>pars</u> a partibus in fenitum concambiavemus
180. CDT 27 contra hanc suprascripta vinditiones cod [= quod] <u>textus</u> cartule contenet
181. CDT 27 pro suprascripta vinditiones quod <u>textus</u> cartule contenit
182. CDL 255 secundum ut suprascripta <u>iussio</u> continet
183. CDL 255 nam nec lex nec <u>ratio</u> contenit ut ipsa femina cum custodes ecclesie simul inhabitet
184. MED 742 basilicam ipsam ... <u>pars</u> monasterii Sexto non contradicit
185. MED 774 ut abere deberet <u>pars</u> ipsius ecclesie sancti Cassiani casa et res ipsas unde agebantur
186. CDL 120 et sic decreuet <u>animus</u> eius ut ... ipsa eclesia a fundamentis construere deuirem

282

187. MED 385 sicut <u>textus</u> brevis inferius declarat

188. MED 742 quod basilicam ipsam iniuste <u>pars</u> monasterii Sexto detineret

189. CDL 255 iterum <u>pars</u> ipsius infantuli dicebat

190. MED 742 nescio si <u>pars</u> monasterii Sexto basilicam ipsam habead

191. MED 596 <u>pars</u> ... ecclesie sancti Martini nec pars episcoporum ... nullam abeant potestatem ex ipsa ecclesia

192. MED 742 de basilicam ipsam edificatam ... inquisivi <u>pars</u> monasterii Sexto

193. CDL 267 antequam nus repentina obserueat <u>mortis</u>

194. CDL 28 ut <u>uirtus</u> permiset

195. CDL 171 et si me <u>mors</u> occupauerit

196. MED 216 si ... nobis <u>pars</u> curtis regie ipsam casam et rem retulerit

197. CDL 30 eam que sibi <u>congregatio</u> eligere uoluerit, ipsa in abbatisse ordo succidat

C2. Accusative-form inanimate singular subjects of transitive clauses (all 17)

C2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	∑
198.	debeo	have to (make)	ordinationem	diueas	facere subtractionem	1	1	1	0	1	0	0	0	0	0	4
199.	debeo	have to (defend)	parte	debeat	defendere terra	1	1	0	0	1	1	0	0	0	1	5
200.	do	give	curtes	dare	quem	1	1	1	0	1	1	0	1	0	1	7
201.	do	give	curte	dedet	quem	1	1	1	0	1	1	1	1	0	1	8
202.	do	give	curtes	dedet	quem	1	1	1	0	1	1	1	1	0	1	8
203.	eligo	elect	congrecationem	eligeret	quem	1	1	1	0	1	1	0	1	0	1	7
204.	habeo	have	parte	aberet	petie	1	0	0	0	0	1	0	0	0	1	3
205.	habeo	have	uirtute	haueat	omnia	1	0	0	0	0	1	0	0	0	0	2
206.	habeo	have	uertute	habeat	casa	1	0	0	0	0	1	0	0	0	1	3
207.	habeo	have (power)	ordinationem	haueas	potestatem	1	0	0	0	0	0	0	0	0	0	1
208.	impedio	impede	breve	impedit	#	0	1	0	0	0	0	1	0	0	0	2
209.	impedio	impede	brevem	impedit	#	0	1	0	0	0	0	1	0	0	0	2
210.	impedio	impede	livello	impedit	#	0	1	0	0	0	0	1	0	0	0	2
211.	possideo	possess	uertute	possedeat	medietate	1	0	0	0	0	1	0	0	0	1	3
212.	possideo	possess	uirtutem	possedeas	#	0	0	0	0	0	1	0	0	0	0	1
213.	possideo	possess	uirtutem	possedeat	#	0	0	0	0	0	1	0	0	0	0	1
214.	video	seem (to contain)	breve	videtur	continere #	0	0	0	0	0	1	0	0	0	0	1
					∑	11	9	5	0	6	12	5	4	0	8	60
					%	65	53	29	0	35	71	29	24	0	47	35

198. CDL 138 nec neque nus neque <u>ordinationem</u> nostra ... nulla subtractionem uel semationem facere diueas

199. MED 394 <u>parte</u> prefate ecclesie sancti Petri terra ipsa sibi defendere debeat qualiter potuerit

200. CDL 113 ut ipse uiganium quem <u>curtes</u> regia dare [= daret]

201. CDL 113 et de ipsa terra quem dedet <u>curte</u> regia

202. CDL 113 et ipse edificias quem <u>curtes</u> hic prope ciuitate dedet eclesie sancti Martini

203. CDL 28 quem ipsam <u>congrecationem</u> siui abbatem et priorem eligeret, ipse sit in loco

204. MED 397 ut ipse quattuor petie de vinea aberet <u>parte</u> iam dicte ecclesie sancti Martini

205. CDL 90 et post decesso eius omnia et in omnibus haueat, ut supra, ipsa sancta Dei <u>uirtute</u>

206. CDL 90 suprascripta casa ... uolo adque decerno ut habeat ipse sancta Dei <u>uertute</u> uel sacerdos

207. CDL 138 nullus de nus nec <u>ordinationem</u> ... haueas potestatem nulla femina adducendi

208. MED 309 <u>breve</u> ista mihi non impedit, quia veritas non fuit

209. MED 385 <u>brevem</u> istam mihi non impedit, quia veritas non fuit

210. MED 397 <u>livello</u> isto quas tu ostendis ad parte sancti Martini nulla impedit

211. CDL 90 medietate de ipsa terrola possedeat ipsa sancta Dei <u>uertute</u>

212. CDL 94 sicot supra legitor, securiter possedeas ipsa sancta Dei <u>uirtutem</u>

213. CDL 73 siquod superius legitur, in eodem tenure possedeat ipsa sancta <u>uirtutem</u>
214. MED 239 sicut ..... <u>breve</u> illa continere videtur

C3. Nominative-form animate singular subjects of transitive clauses (25 in 98)

C3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
215.	aedifico	build	pater	aedificabet	quem	1	1	1	0	1	1	1	1	1	1	9
216.	compono	recompense	heridis	conpona	soledos	1	1	1	0	1	1	0	1	0	0	6
217.	confirmo	confirm	genitor	confirmauit	rem	1	1	1	0	1	1	1	1	0	1	8
218.	debeo	have to (possess)	custos	debeat	possedere #	0	0	0	0	0	1	0	0	0	0	1
219.	debeo	have to (give)	filius	debeat	dare denarios	1	1	1	0	1	1	0	0	0	0	5
220.	debeo	have to (possess)	homo	debeatis	abere/possidere/lavorare/ fruere res	1	0	0	0	1	1	0	0	0	1	4
221.	debeo	have to (receive/feed)	presbiter	deveat	suscipere/pascere quos	1	1	1	0	1	1	0	0	0	1	6
222.	debeo	have to (pray)	sacerdus	diueas	hora<re> #	0	1	0	0	1	1	0	0	0	0	3
223.	demando	demand	rex	demandavit	#	0	1	0	0	1	1	1	1	0	0	5
224.	do	give	homo	dedi	quas	1	1	1	0	1	1	1	1	0	1	8
225.	facio	make (ordination)	pontifex	faciat	ordinatione	1	1	1	0	1	1	0	1	0	0	6
226.	habeo	have (power)	homo	abeatis	potestatem vendere medietate	1	0	0	0	0	1	0	0	0	0	2
227.	habeo	have	genitor	abuit	quantum	1	0	0	0	0	1	1	0	0	0	3
228.	habeo	have (the right)	episcopus	aveat	licentiam venire	1	0	0	0	0	1	0	0	0	0	2
229.	habeo	have	sacerdos	habeat	casa	1	0	0	0	0	1	0	0	0	1	3
230.	judico	decide	pater	iudicat	#	0	1	1	0	1	1	1	1	0	0	6
231.	possum	can (alienate)	nullus(que)	possant	extraneare eclesia	1	1	1	0	1	0	0	0	0	1	5
232.	possum	may (interfere)	nullus	possit	inferi molestia	1	1	0	0	1	0	0	0	0	0	3
233.	praesumo	dare (to expel)	posterus	praesumat	molestari te	1	1	0	0	1	0	0	0	1	1	5
234.	praesumo	dare (to perform)	presbiter	presumpserit	agere aliquid	1	1	0	0	1	1	0	1	0	0	5
235.	profiteor	profess	vinditor	profiteor	me suscepisse solidos	1	0	1	0	1	1	1	1	0	1	7
236.	quaero	try (to contest...)	homo	quesierimus	intentionare/retolli/ subtragi #	0	1	1	0	1	1	0	1	0	0	5
237.	quaero	try (to contest...)	homo	quesierit	intentionare/subtragere eam	1	1	1	0	1	1	0	1	0	1	7
238.	spondeo	undertake (to recompense)	homo	spondimus	componere terra	1	1	1	0	1	1	0	1	0	1	7
239.	video	seem (to have)	massarius	videtur	abere quantu	1	0	0	0	0	1	0	0	0	0	2
					Σ	20	17	13	0	19	22	7	12	2	11	123
					%	80	68	52	0	76	88	28	48	8	44	49

215. CDL 194 ecclesie sancte Marie et sancti Donati , quem <u>pater</u> meus aedificabet [= aedificavit]
216. CDL 214 conpona uobis ille <u>heridis</u> meus ... auri soledos nomero mille quinienti
217. CDL 245 quia <u>genitor</u> noster per cartulam omnem rem suam in nobis confirmauit
218. CDL 145 ipsa Dei eclesia uel eius <u>custos</u> inuiolabiliter possedere debeat
219. MED 596 ipse Walperto clerico quam et ipse <u>filius</u> eius ... dare et persolvere debeat ... denarios ... numero triginta
220. MED 424 tu aut ille <u>homo</u> ... suprascripta res ... abere et possidere et lavorare seu fruere debeatis
221. MED 231 quos una die per singulas ebdomadas in ipso senodocio sancti Vitalis suscipere et a mensa pascere deveat ipse <u>presbiter</u>
222. CDL 24 ut pro meis peccatis <u>sacerdus</u> qui iniui ordinatus est aut fueret hora<re> diueas

223. MED 385 ecce paratus sum, iudica inter me et te, sicut domnus <u>rex</u> demandavit

224. MED 401 predicta rem quas tibi, ut supra, in commutationem dedi nos vel ille <u>homo</u>

225. CDL 179 ille <u>pontifex</u> iuidem ordinatione faciat

226. CDL 281 ut post meum decessum tam tu quam et ille <u>homo</u> ... potestatem abeatis uendere et dispensare medietate ex omni re mea

227. MED 658 omnia quantum ad ipsa casa est pertenentes et suprascripto <u>genitor</u> meus exinde ad manum sua abuit

228. MED 231 volo ut ... <u>episcopus</u> huius civitatis ad celebrandas missas cum sacerdotibus ibidem venire licentiam aveat

229. CDL 90 nam suprascripta casa ... uolo adque decerno ut habeat ipse sancta Dei uertute uel <u>sacerdos</u>

230. CDL 287 sicut <u>pater</u> iudicat, in eo moderamen persistat

231. CDL 127 nec per nullo ingenio <u>nullusque</u> de nus ipsa Dei eclesia uel res eidem pertinente aliui extraneare possant

232. CDT 15 <u>nullus</u> de heredes, proheredes nostrus contra hanc cartulam in aliquo possit inferi molestia

233. CDL 35 et numquam nos uel <u>posterus</u> noster te de hanc dicto loco molestari praesumat

234. MED 309 si quis <u>presbiter</u> aut diaconus ad proprio episcopo excommunicatus presumpserit aliquid ministerii agere

235. CDT 27 unde profiteor me qui supra <u>vinditor</u> suscepisse et accepit ... auri solidos triginta

236. MED 673 si ... intentionare aut retolli vel subtragi quesierimus per quolibet ingenium nos aut ille <u>homo</u>

237. MED 291 et si quiscumque <u>homo</u> absque nostro et vestro conludio vobis eam intentionare aut subtragere quesierit

238. MED 384 nos vel ille <u>homo</u> ... spondimus cum nostris heredibus componere tibi vel successoribus tuis predicta terra

239. MED 216 omnia et in omnibus quantu mihi ad suprascripto Gausprando obvenit et ipse <u>massarius</u> ad manus sua abere videtur

C4. Accusative-form animate singular subjects of transitive clauses (all 19)

C4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
240.	compono	recompense	heredes	compona	res	1	1	1	0	1	1	0	1	0	0	6
241.	confirmo	confirm	notario	confirmauerunt	testimonium	1	1	1	0	1	1	1	1	0	1	8
242.	confirmo	confirm	germano	confirmauet	eclesia	1	1	1	0	1	1	1	1	0	1	8
243.	debeo	have to (do)	uno	debeamus	facere (?)	1	1	0	0	1	1	0	0	0	0	4
244.	debeo	have to (do)	unum	debeamus	(angaria/dies) facere	1	1	0	0	1	1	0	0	0	0	4
245.	debeo	should (expel)	erede	debeas	molestare (vos)	1	1	1	0	1	0	0	0	1	1	6
246.	debeo	should (lose)	filio	deberet	perdere (ecclesia)	1	0	1	0	0	1	0	0	0	1	4
247.	debeo	have to (pray)	presbitero	diueas	deprecare (Dominus)	1	1	0	0	1	1	0	0	0	1	5
248.	habeo	have (power)	unum	abeatis	potestatem	1	0	0	0	0	1	0	0	0	0	2
249.	habeo	have	filio	habet	quem	1	0	0	0	0	1	1	0	0	1	4
250.	habeo	have	massario	habet	quem	1	0	0	0	0	1	1	0	0	1	4
251.	habeo	have (the right)	unum	haueant	licentiam (introire)	1	0	0	0	0	1	0	0	0	0	2
252.	judico	decide	unum	iudicaverimus	(#)	0	1	1	0	1	1	0	1	0	0	5
253.	offero	offer	genitore	offeruit	quas	1	1	1	0	1	1	1	1	0	1	8
254.	pono	cause (a negligency)	episcopo	posuerit	neglegentiam	1	1	1	0	0	1	0	0	0	0	4
255.	possum	may (interfere)	homine	possit	(offerta) molestare	1	1	0	0	1	0	0	0	1	1	5
256.	reddo	give (a testimony)	notarium	reddidissent	testimonium	1	1	1	0	1	1	0	1	0	0	6
257.	video	seem (to have)	guvernatore	videor	(quantum) habere	1	0	0	0	0	1	0	0	0	0	2
258.	video	seem (to have)	rectore	videor	(quantum) habere	1	0	0	0	0	1	0	0	0	0	2
					Σ	18	12	9	0	11	17	5	6	2	9	89
					%	95	63	47	0	58	89	26	32	11	47	47

240. MED 277 compona ipsi <u>heredes</u> meus ... omnia in triplum res melioratas

241. CDL APPENDICE et dum testimonia et <u>notario</u> ipsum per euangelia testimonium suum confirmauerunt

242. CDL 170 qualiter iam dicto <u>germano</u> meo in me ipsa eclesia et monasterio sancti Petri confirmauet

243. MED 418 vero vobis facere debeamus <u>uno</u> de nos in suprascripto loco finibus Maritimense et ..... portione , ad que vobis fuerit utilitas

244. MED 301 angaria vero per singulas septimana vobis facere debeamus <u>unum</u> ex nobis aut nostris heridis dies duo

245. CDL 62 et numquam ego Filimari aut nullo <u>erede</u> meo uos de ipsas res in alico molestare debeas

246. MED 202 quod ipsi <u>filio</u> meo per sua culpa ecclesia et res ipsa perdere deberet

247. CDL 136 ut <u>presbitero</u> , qui in psa Dei eclesia ordinatus fueri , pro meis facinoribus Dominus deprecare diueas

248. MED 689 potestatem abeatis tam vos toti insimul dispensatoribus meis quam et duo seu <u>unum</u> ex vobis

249. CDL 154 due filie Fuscule de Tramonte , quem habet ad muliere <u>filio</u> Teudaldi

250. CDL 160 clausura illa in integro , quem ad manus suas habet ... <u>massario</u> eclesie sancte Marie

251. CDL 127 tantum <u>unum</u> de heredis seo de filii nostros per capud haueant licentiam introire

252. MED 253 tam toti insimul quam et <u>unum</u> de nobis ... in quod iudicaverimus de res eius

253. MED 524 quas ipsius ecclesie offeruit ipse <u>genitore</u> nostro

254. MED 231 et si <u>episcopo</u> huius civitatis neglegentiam posuerit faciendo

255. MED 172 ut neque ego qui supra offertor neque meis heredibus neque nullo <u>homine</u> ... possit molestare aut resubtragere mea offerta

256. CDL APPENDICE et dum testimonia ipsa et ipsum <u>notarium</u> taliter testimonium reddidissent

257. MED 195 quantum habere videor [= videtur] <u>rectore</u> et <u>guvernatore</u>

258. MED 195 quantum habere videor [= videtur] <u>rectore</u> et guvernatore

C5. Nominative-form personal-name singular subjects of transitive clauses (25 in 219)

C5	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
259.	affor	speak	Dominus	adfatus	#	0	1	0	0	1	1	1	1	0	0	5
260.	condo	endow/give	Dominus	condedit	arbitrio	1	1	1	0	1	1	1	1	0	0	7
261.	conservo	conserve	Dominus	conseruaueret	nus	1	1	0	0	1	1	0	1	0	1	6
262.	decerno	order	Sichipertus	decrevit	#	0	1	1	0	1	1	1	1	0	0	6
263.	dico	say	Alpulus	dicebat	direct quotation	0	1	1	0	1	1	1	1	0	0	6
264.	dico	say	Dominus	dicit	direct quotation	0	1	1	0	1	1	1	1	0	0	6
265.	dico	say	Gallus	dixit	salutem	1	1	1	0	1	1	1	1	0	0	7
266.	dico	say	Anseramus	disserunt	#	0	1	1	0	1	1	1	1	0	0	6
267.	dico	say	Atus	dixit	#	0	1	1	0	1	1	1	1	0	0	6
268.	dignor	condescend (to donate)	Dominus	dignatus	donare filios	1	1	1	0	1	1	0	1	0	0	6
269.	dignor	condescend (to donate)	Dominus	dignatus	donare filium	1	1	1	0	1	1	0	1	0	0	6
270.	dignor	condescend (to donate)	Dominus	dignatus	donare que	1	1	1	0	1	1	0	1	0	1	7
271.	dirigo	send	Peredeus	direxit	Munualdum	1	1	1	0	1	1	1	1	0	1	8
272.	do	give	Dominus	dederit	quas	1	1	1	0	1	1	0	1	0	1	7
273.	do	give	Rumoaldus	dedit	wadia	1	1	1	0	1	1	1	1	0	0	7
274.	facio	make	Deus	fecit	hominem	1	1	1	0	1	1	1	1	0	0	7
275.	facio	make	Humulus	fecerunt	placitum	1	1	1	0	1	1	1	1	0	0	7
276.	habeo	have	Tachipertus	abuit	quam	1	0	0	0	0	1	1	0	0	1	4
277.	habeo	have	Grecorius	abuit	quantum	1	0	0	0	0	1	1	0	0	0	3
278.	interrogo	interrogate	Rachinardus	interrogaverat	Alpulum	1	1	1	0	1	1	1	1	0	1	8
279.	mitto	put	Britto	misit	ea	1	1	1	0	1	1	1	1	0	1	8
280.	possum	can (prepare)	Dominus	poteest	preparare quod	0	1	1	0	1	1	0	1	1	1	7
281.	reservo	reserve	Deus	reseruauet	notitie	1	1	1	0	1	1	1	1	0	1	8
282.	scribo	write	Osprandus	scripsit	quam	1	1	1	0	1	1	1	1	1	1	9
283.	volo	contest	Adegrimus	voluit	contendere illam	1	1	1	0	1	1	0	1	0	1	7
					Σ	18	23	21	0	23	25	18	23	2	11	164
					%	72	92	84	0	92	100	72	92	8	44	66

259. CDL 124 ideo omnipotens <u>Dominus</u> suo fideli famulo iustoque dispensatori adfatus est dicens

260. MED 196 quia omnipotens <u>Dominus</u> homine livero arbitrio condedit

261. CDL 138 dum nus <u>Dominus</u> in hoc seculo conseruaueret

262. MED 231 qualiter ipsi <u>Sichipertus</u> clericus decrevit per ipso dote

263. MED 309 dicebat prefatus <u>Alpulus</u>

264. CDL 93 <u>Dominus</u> dicit

265. CDL 94 <u>Gallus</u> uir uenerabilis diaconus tibi eclesie Dei ad que beatissimi sancti Reguli ... perpetuam salutem dixit

266. MED 539 Ostrualdus clericus, Gumfridi clericus, <u>Anseramus</u> clericus, Fridianus clericus, Guntelmus clericus similiter disserunt

267. MED 539 dixit ... <u>Atus</u> presbiter similiter

268. CDL 90 et si mihi <u>Dominus</u> filios aut filias donare dignatus fueret

269. MED 372 si nobis <u>Dominus</u> pro sua misericordia filium aut filiam donare dignatus fuerit

270. CDL 186 que iniui ad ipsos Baroncione et Ursus data est aut per qualiuet ingenium iniui <u>Dominus</u> donare fuit dignatus

271. CDL 241 et super hoc cambium direxit <u>Peredeus</u> in Dei nomine episcopus Munualdum

272. MED 621 medietate labore maiore et tertia de minore quas de res ipsa nobis <u>Dominus</u> dederit

273. MED 774 et statim dedit wadia ipse <u>Rumoaldus</u> advocato ipsius Audiprandi ita adprobandum

274. CDL 94 rerum creaturarum omnium creator <u>Deus</u> fecit hominem ad imaginem sue similitudinis

275. CDL 247 fecerunt inter se placitum Gaudiosus presbiter et <u>Humulus</u> clericus

276. MED 393 terra illa quam ipse <u>Tachipertus</u> clericus abuit

277. MED 472 omnia et in omnibus quantum ad predicta ecclesia est pertinentes et ipse <u>Grecorius</u> presbitero ... abuit

278. MED 309 et paululum post infra igitur <u>Rachinardus</u> ... interrogaverat eumdem Alpulum

279. CDL APPENDICE presenti ante nos ipse <u>Britto</u> misit ea in focum et ibi arsit

280. CDL 118 non alius thensaru non est talis quili est illa aeterna uita quod nobis <u>Dominus</u> preparare poteest

281. CDL 267 uita [= vitae] tempore [= temporis] et mortis omnipotens <u>Deus</u> sue reseruauet notitie

282. CDL 208 et ecce breue scriptum ... quam <u>Osprandus</u> diaconus scripsit per eius dictatum

283. MED 539 set <u>Adegrimus</u> wassus domni regis illam voluit contendere ad partem palatii

C6. Accusative-form personal-name singular subjects of transitive clauses (25 in 161)

C6	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
284.	adduco	bring	Magnouiro	atduxerunt	cartula	1	1	1	0	1	1	1	1	0	1	8
285.	confero	bring	Alaperto	contulit	quam	1	1	1	0	1	1	1	1	0	1	8
286.	construo	construct	Ansuartu	construxerunt	ecclesiam	1	1	1	0	1	1	1	1	1	1	9
287.	debeo	have to (give)	Walperto	debeat	dare denarios	1	1	1	0	1	1	0	1	0	0	6
288.	dico	say	Iohanne	dicebant	direct quotation	0	1	1	0	1	1	1	1	0	0	6
289.	dico	say	Liutpertu	dixit	direct quotation	0	1	1	0	1	1	1	1	0	0	6
290.	dico	say	Teudiprando	dixerunt	direct quotation	0	1	1	0	1	1	1	1	0	0	6
291.	do	give	Ciemiccio	dedet	quem	1	1	1	0	1	1	1	1	0	1	8
292.	facio	make	Brittulo	fecit	quas	1	1	1	0	1	1	1	1	0	1	8
293.	habeo	have	Atripaldulo	abuerunt	queque	1	0	0	0	0	1	1	0	0	1	4
294.	habeo	have	Argimo	abuit	ecclesiam	1	0	0	0	0	1	1	0	0	1	4
295.	habeo	have	Poso	abuit	que	1	0	0	0	0	1	1	0	0	1	4
296.	habeo	have	Ursulo	abuit	quam	1	0	0	0	0	1	1	0	0	1	4
297.	habeo	have	Aiolfulo	habuit	quanto	1	0	0	0	0	1	1	0	0	0	3
298.	habeo	have	Bonulo	abuerunt	que	1	0	0	0	0	1	1	0	0	1	4
299.	habeo	have	Pettulo	abuet	#	0	0	0	0	0	1	1	0	0	0	2
300.	habeo	have	Ghinnulo	abuit	quantum	1	0	0	0	0	1	1	0	0	0	3
301.	habeo	have	Ermulo	abuimus	quantum	1	0	0	0	0	1	1	0	0	0	3
302.	mitto	give (explanation)	Wichelmo	mitterent	rationes	1	1	1	0	1	1	0	1	0	1	7
303.	offero	offer	Rigniperto	offeruerunt	ecclesia	1	1	1	0	1	1	1	1	0	1	8

	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
304.	relaxo	ease (= forgive)	Domino	relaxare	pondere/peccata	1	1	1	0	1	1	0	1	0	0	6
305.	respondeo	answer	Iohanne	respondebat	direct quotation	0	1	0	0	1	1	1	1	0	0	5
306.	vendo	sell	Donni	uendidet	ea	1	1	1	0	1	1	1	1	0	1	8
307.	venumdo	sell	Teudectulo	venundavit	casis	1	1	1	0	1	1	1	1	0	1	8
308.	video	seem (to have)	Leulo	videtur	abere quam	1	0	0	0	0	1	0	0	0	1	3
					∑	20	15	14	0	15	25	21	15	1	15	141
					%	80	60	56	0	60	100	84	60	4	60	56

284. CDL 128 qualiter atduxerunt Maurino et Magnouiro presbiteri cartula donationis de eglesia sancti Prosperii

285. MED 595 illa portio quam quondam Alaperto presbitero per cartulam isdem ecclesie Domini Salvatoris contulit

286. MED 291 quia quondam Ansuartu, Ermifridi, Ermualdu, Ansprand, Ermerisci et Ermulau ... construxerunt ecclesiam

287. MED 596 ut tam ipse Walperto clerico quam et ipse filius eius ... dare et persolvere debeat ... denarios ... triginta

288. MED 202 dicebant Iohanne episcopus cum Fratellu advocato suo

289. CDL 171 Liutpertu uir deuotus filio quondam Dondoloni dixit

290. MED 539 Aloni, Ansperto, Teudiprando et Ermiprando germani ... equaliter dixerunt

291. CDL 154 Teuderisciulu, quem dedet nobis Ciemiccio in uiganio

292. CDL 253 ipsa uinditione quas eorum Brittulo fecit

293. MED 714 casis et rebus illis ... queque Atripaldulo et Aggi abuerunt

294. MED 786 ecclesiam ... quas quondam Argimo presbitero ad manus suas abuit

295. MED 637 uno modiorum de terra in Fabrorum que quondam Poso ad manu sua abuit

296. MED 710 casa et res illa in suprascripto loco Vallivu quam Ursulo ad manus suas abuit

297. MED 500 casa ipsa cum fundamento suo et cum orto suo quanto Aiolfulo ibi ad manu sua habuit

298. MED 543 res illa ... que antea ad manus sua abuit Bonulo et Magnulo

299. MED 439 omnia et in omnibus ubi Pettulo massario residere visus fuet et ipsi Pettulo ad manum suam abuet

300. MED 753 omnia quantum ad ipsa casa est pertinentes et suprascripto Ghinnulo exinde ad manus suas abuit

301. MED 787 omnia ... quantum ... suprascripto quidam Ermulo et ego ipsis qui supra Anso presbitero exinde ad manus nostra abuimus

302. MED 595 inter se utrumque Wichelmo et Teufrido advocato diversas mitterent rationes

303. MED 595 Teutperto presbitero et Rigniperto clerico offeruerunt ... interdicta ecclesia sancti Grecorii

304. CDL 171 ut mihi Domino pondere, peccata relaxare

305. MED 202 respondebat ipsi Iohanne episcopus cum Fratello advocato suo

306. CDL 89 portionem eius ... uendere et tradere uisus sum secundum qualiter ea mihi Donni uendidet

307. MED 797 Teudectulo filius quondam Pauli ... venundavit mihi, id est casis et omnibus rebus suis

308. MED 420 unam petiam de vinea mea, quam abeo in loco Vaccule, quam ad lavorandum abere videtur Leulo

Clauses with plural subjects

D. Unaccusative clauses

D1. Nominative-form inanimate plural subjects of unaccusative clauses (all 25)

D1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
1.	affluo	abound	diuitie	affluent		0	0	0	0	0	1	0	0	0	0	1
2.	eo	go/be valid	quales	ierint		0	0	0	0	0	1	0	0	0	0	1
3.	pertineo	pertain	res	pertenuissem		0	0	0	0	0	1	0	0	0	0	1
4.	remaneo	remain	quante	remansere		0	0	0	0	0	1	0	0	0	0	1
5.	sum	be	quales	fuerint		0	0	0	0	0	1	0	0	0	0	1
6.	sum	be	quales	fuerint		0	0	0	0	0	1	0	0	0	0	1
7.	sum	be	quales	fuerint		0	0	0	0	0	1	0	0	0	0	1
8.	sum	be	quales	fuerint		0	0	0	0	0	1	0	0	0	0	1

No.															
9.	sum	be	qualis	fuerint		0	0	0	0	0	1	0	0	0	0	1
10.	sum	be	qualis	fuerint		0	0	0	0	0	1	0	0	0	0	1
11.	sum	be	qualis	fuerint		0	0	0	0	0	1	0	0	0	0	1
12.	sum	be	quale	fuerit		0	0	0	0	0	1	0	0	0	0	1
13.	sum	be	qualis	fuerit		0	0	0	0	0	1	0	0	0	0	1
14.	sum	be	case	sint		0	0	0	0	0	1	0	0	0	0	1
15.	sum	be	res	sint		0	0	0	0	0	1	0	0	0	0	1
16.	sum	be	terre	sint		0	0	0	0	0	1	0	0	0	0	1
17.	sum	be	uineae	sint		0	0	0	0	0	1	0	0	0	0	1
18.	sum	be	curtes	sit		0	0	0	0	0	1	0	0	0	0	1
19.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
20.	sum	be	petiae	sunt		0	0	0	0	0	1	1	0	0	0	2
21.	sum	be	uineae	sunt		0	0	0	0	0	1	1	0	0	0	2
22.	sum	be	petie	sunt		0	0	0	0	0	1	1	0	0	0	2
23.	sum	be	petie	sunt		0	0	0	0	0	1	1	0	0	0	2
24.	valeo	should (remain)	res	ualeat	permanere	0	0	0	0	0	1	0	0	0	0	1
25.	video	seem (to pertain)	quante	videtur	pertinere	0	0	0	0	0	1	0	0	0	0	1
					∑	0	0	0	0	0	25	4	0	0	0	29
					%	0	0	0	0	0	100	16	0	0	0	12

1. CDL 93 diuitie si affluent, nolite cor adponere
2. CDT 43 solidos duo Langobardiscos ... quales tunc per tempore per ista patria ierint
3. MED 202 ipse bassilica sancti Petri vel res ... aliquando ecclesie sancti Martini ... pertenuissem [= pertinuissent]
4. CDL 214 omnis terre mei, quante ad me hic circa ciuitatem ista non iudicate remansere
5. MED 240 ipsos homines cum quibus aut quales tunc fuerint
6. MED 477 duodeci denarios bonos ... quales tunc per tempus expendibiles fuerint
7. MED 478 quattuordeci denarios bonos ... quales tunc per tempus expendibiles fuerint
8. MED 506 denarios septuagenta duo ... quales in diebus illis per caput expendibiles fuerint hic Luca
9. MED 424 quatraginta quinque solidos ... qualis tunc melioris per istam civitate per caput fuerint expendivilis
10. MED 428 solidos quindecim ... qualis tunc melioris per istam civitate per caput fuerint expendiviles
11. MED 516 denarios bonos qualis tunc hic Luca per caput fuerint expendivilis numero viginti septe
12. MED 806 denarios bonos expendiviles quale per civitate ista Lucense ... fuerit expendiviles numero viginti et uno
13. MED 433 dinari nove boni qualis tunc in die illa per capu bene fuerit expendivilis
14. MED 582 ut da admodum iam dicte case cum rebus suis in tua sint potestatem abendi
15. MED 244 cum omnes res tue sint pertenentes ecclesie sancti Ipoliti
16. CDL 161 hec omnia suprascripte terre uel uineae ... in integrum sint in ista sorte
17. CDL 161 hec omnia suprascripte terre uel uineae ... in integrum sint in ista sorte
18. CDL 175 curtes sundriales ... sit in potestate suprascripte eclesiae
19. MED 206 ipsum monasterium cum rebus suis in mea sit potestate habitandi ... et res mobiles et nutrimina
20. CDL 265 ambae ipsae petiae sunt per mensuram plus minus sistariorum uiginti
21. CDL 265 et ipsae uineae sunt per mensuram plus minus sistariorum decem et octo
22. MED 788 et ille petie de vinea et tres petie de terre sunt ad uno tenente
23. MED 788 et ille petie de vinea et tres petie de terre sunt ad uno tenente
24. CDL 40 ipse sanctus monasterio et omnes res iuidem pertenente in tua defensionem et dominio ualeat permanere
25. MED 767 una cum omnes moniminas quante mihi nunc ex suprascriptis rebus pertinere videtur

289

D2. Accusative-form inanimate plural subjects of unaccusative clauses (all 20)

D2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	∑
26.	convenio	be convenient	quas	conuenit		0	0	0	0	0	1	1	0	0	0	2
27.	debeo	have to (be)	res	debeant	esse	0	0	0	0	0	1	0	0	0	0	1
28.	obvenio	come down	casas	obvine		0	0	1	0	0	1	1	0	0	0	3
29.	obvenio	come down	ris	obvine		0	0	1	0	0	1	1	0	0	0	3
30.	permaneo	remain	casas	permaneat		0	0	0	0	0	1	0	0	0	0	1
31.	sum	be	petias	sint		0	0	0	0	0	1	0	0	0	0	1
32.	sum	be	res	sint		0	0	0	0	0	1	0	0	0	0	1
33.	sum	be	res	sint		0	0	0	0	0	1	0	0	0	0	1
34.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
35.	sum	be	casas	sit		0	0	0	0	0	1	0	0	0	0	1
36.	sum	be	casas	sit		0	0	0	0	0	1	0	0	0	0	1
37.	sum	be	curtis	sit		0	0	0	0	0	1	0	0	0	0	1
38.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
39.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
40.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
41.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
42.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
43.	sum	be	res	sit		0	0	0	0	0	1	0	0	0	0	1
44.	sum	be	quercias	sunt		0	0	0	0	0	1	1	0	0	0	2
45.	sum	be	res	sunt		0	0	0	0	0	1	1	0	0	0	2
				∑		0	0	2	0	0	20	5	0	0	0	27
				%		0	0	10	0	0	100	25	0	0	0	14

26. CDL 288 solidos septe finitum pretium quas inter nouis bono animo conuenit
27. CDL 131 ut cunctis diebus ipsas res in potestate de predictas eclesias esse debeant
28. CDT 23 mihi ipsas casas vel ris da quondam Cuntiperti consobrino tuo obvine
29. CDT 23 mihi ipsas casas vel ris da quondam Cuntiperti consobrino tuo obvine
30. MED 372 suprascriptas casas ... permaneat semper in potestate predicte Dei ecclesie
31. MED 507 ut infra isti quinque anni venturi sint ibidem ipsas ambas petias vinea levata et propaginata
32. CDL 197 ipsas res sint in potestatem in ipsa Dei ecclesia
33. CDL 197 ut omnia ipsas res ad ipsas casas pertenentem in mea sint potestatem lauorandi
34. CDT 12 omnis res meas ... omni tempore in eorum sit potestate
35. CDL 175 casas massaricias et aldionales ... sit in potestate suprascripte eclesiae
36. MED 277 ut omnia casas et res meam in mea sit potestate vendendum
37. MED 425 curtis ... in integrum in mea sit potestate abendi
38. CDL 100 ecclesiam et omnia ea ipsas res suprascriptas in mea sit potestatem ordinandi
39. CDL 140 nam alias res meas, ut dixi, omnia sit in potestate suprascripte ecclesie
40. CDL 287 simul et omnis res illas ... usufructuandi in tua sit potestatem
41. CDL 287 simeliter et res illas ... usufructuandi in tua sit potestatem
42. MED 231 et omnes res ipsas et homines ... sit potestatem in omnibus
43. CDL 261 omnes res meas in mea sit potestatem usufructuandum
44. CDL 146 super illu est alius testuclu tesseratu et super illu duo testucli sunt duo quercias
45. MED 774 quia casa et res ipsas proprie sunt ipsius ecclesie sancti Cassiani

290

D3. Nominative-form animate plural subjects of unaccusative clauses (25 in 39)

D3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
46.	adsum	be present	alii	aderant		0	0	0	0	0	1	1	0	0	0	2
47.	adsum	be present	homines	aderant		0	0	0	0	0	1	1	0	0	0	2
48.	adsum	be present	filii	aderant		0	0	0	0	0	1	1	0	0	0	2
49.	adsum	be present	lociservatores	aderant		0	0	0	0	0	1	1	0	0	0	2
50.	adsum	be present	plures	aderant		0	0	0	0	0	1	1	0	0	0	2
51.	adsum	be present	plures	aderant		0	0	0	0	0	1	1	0	0	0	2
52.	adsum	be present	sacerdotes	aderant		0	0	0	0	0	1	1	0	0	0	2
53.	adsum	be present	sacerdotes	aderant		0	0	0	0	0	1	1	0	0	0	2
54.	adsum	be present	sacerdotes	aderant		0	0	0	0	0	1	1	0	0	0	2
55.	adsum	be present	aremannis	adessent		0	0	0	0	0	1	0	0	0	0	1
56.	debeo	have to (be)	omnes	debeant	esse	0	0	0	0	0	1	0	0	0	0	1
57.	debeo	have to (be)	heredis	deveamus	esse	0	0	0	0	0	0	0	0	0	0	0
58.	debeo	have to (be)	successoris	deveamus	esse	0	0	0	0	0	0	0	0	0	0	0
59.	promitto	promit (to remain)	filiis	promittimus	permanere	0	0	0	0	1	1	1	0	0	0	3
60.	recedo	recede (= die)	plures	recesseret		0	0	1	1	0	1	0	0	0	0	3
61.	sum	be	reliqui	erant		0	0	0	0	0	1	1	0	0	0	2
62.	sum	be	sacerdotes	fuerant		0	0	0	0	0	1	1	0	0	0	2
63.	sum	be	sacerdotes	fuerant		0	0	0	0	0	1	1	0	0	0	2
64.	sum	be	viri	fuissemus		0	0	0	0	0	1	0	0	0	0	1
65.	sum	be	omnes	sint		0	0	0	0	0	1	0	0	0	0	1
66.	sum	be	pauperes	sint		0	0	0	0	0	1	0	0	0	0	1
67.	sum	be	homines	sint		0	0	0	0	0	1	0	0	0	0	1
68.	sum	be	omnis	sumus		0	0	0	0	0	1	1	0	0	0	2
69.	volo	want (to consent)	filii	uolueret	consintire	0	0	0	0	1	0	0	0	0	0	1
70.	volo	want (to reside)	filiis	volueris	resederem	0	0	0	0	0	1	0	0	0	0	1
				Σ		0	0	1	1	2	22	14	0	0	0	40
				%		0	0	4	4	8	88	56	0	0	0	16

46. MED 309 ubi nobiscum aderant sacerdotes et filii sanctae ecclesiae ... et alii <u>plures</u>
47. MED 385 tunc ibidem nobiscum aderant idonei <u>homines</u> qui dixerunt
48. MED 309 ubi nobiscum aderant sacerdotes et <u>filii</u> sanctae ecclesiae
49. MED 202 ubi nobiscum aderant sacerdotes et <u>lociservatores</u> seu haremannos
50. CDL 182 ubi aderant insimul Rachiprando presbiter ... Petto de curte domni regi et alii <u>plures</u>
51. MED 385 ubi nobiscum aderant sacerdotes et ... Christianus subdiaconus et alii <u>plures</u>
52. CDL APPENDICE ubi nobiscum aderant <u>sacerdotes</u> et aremannos huius Lucane ciuitatis
53. MED 202 ubi nobiscum aderant <u>sacerdotes</u> et lociservatores seu haremannos
54. MED 309 ubi nobiscum aderant <u>sacerdotes</u> et filii sanctae ecclesiae
55. MED 397 adessent ibidem suprascripti <u>aremannis</u>
56. CDL 114 uolo ut liueri <u>omnes</u> esse debeant et a ius patronati absoluti
57. MED 253 ut neque nos ... neque <u>heredis</u> vel successoris nostri restauratoris esse non deveamus
58. MED 253 ut neque nos ... neque heredis vel <u>successoris</u> nostri restauratoris esse non deveamus
59. CDT 43 ita et ego qui supra Raghipertus vel meis <u>filiis</u> in omnia ... permanere promittimus
60. CDL 138 et si qualiter Deo fuerit preceptione et unus aut <u>plures</u> de nus de seculo recesseret
61. MED 742 erantque nobiscum Offo, Minto, Liutperto, Rumualdo ... et <u>reliqui</u> multis
62. MED 309 cum ipso Rachinardo fuerant plurimis suis <u>sacerdotes</u>

63. MED 385 Et quando hanc professionem fecerat , cum ipso Rachinardo fuerant plurimis suis sacerdotes .
64. MED 186 et omnia in mea esse [= essent] potestatem, tamquam <u>viri</u> duo fratres fuissemus
65. CDL 93 ut <u>omnes</u> sint ... liueri et absoluti ab omni nexu condicionis uel a iugo seruitutis
66. MED 231 ut sint insimul <u>pauperes</u> duodecim
67. MED 240 illi <u>homines</u> qui mihi a germanis meis competunt in tua sint potestatem
68. CDL 230 quoniam incerti sumus <u>omnis</u> de Dei iudicio
69. CDL 254 si ... <u>filii</u> mei ipsei consintire non uolueret
70. CDT 89 et si <u>filiis</u> tui masculinis volueris [= voluerint] resederem ad suprascripta ordinazionem

D4. Accusative-form animate plural subjects of unaccusative clauses (all 18)

D4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
71.	adsum	be present	aremannos	aderant		0	0	0	0	0	1	1	0	0	0	2
72.	adsum	be present	haremannos	aderant		0	0	0	0	0	1	1	0	0	0	2
73.	adsum	be present	homines	aderant		0	0	0	0	0	1	1	0	0	0	2
74.	adsum	be present	plures	aderant		0	0	0	0	0	1	1	0	0	0	2
75.	adsum	be present	plures	aderant		0	0	0	0	0	1	1	0	0	0	2
76.	mereo	deserve (to live on)	filias	meruerint	aduiuere	0	0	0	0	0	1	0	0	0	0	1
77.	sum	be	aldias	sint		0	0	0	0	0	1	0	0	0	0	1
78.	sum	be	aldiones	sint		0	0	0	0	0	1	0	0	0	0	1
79.	sum	be	aldiones	sint		0	0	0	0	0	1	0	0	0	0	1
80.	sum	be	ancillas	sint		0	0	0	0	0	1	0	0	0	0	1
81.	sum	be	ancillas	sint		0	0	0	0	0	1	0	0	0	0	1
82.	sum	be	homines	sint		0	0	0	0	0	1	0	0	0	0	1
83.	sum	be	seruos	sint		0	0	0	0	0	1	0	0	0	0	1
84.	sum	be	servos	sint		0	0	0	0	0	1	0	0	0	0	1
85.	sum	be	aldiones	sit		0	0	0	0	0	1	0	0	0	0	1
86.	sum	be	ancillas	sit		0	0	0	0	0	1	0	0	0	0	1
87.	sum	be	homines	sit		0	0	0	0	0	1	0	0	0	0	1
88.	sum	be	seruos	sit		0	0	0	0	0	1	0	0	0	0	1
				Σ		0	0	0	0	0	18	5	0	0	0	23
				%		0	0	0	0	0	100	28	0	0	0	13

71. CDL APPENDICE ubi nobiscum aderant sacerdotes et <u>aremannos</u> huius Lucane ciuitatis
72. MED 202 ubi nobiscum aderant sacerdotes et lociservatores seu <u>haremannos</u>
73. MED 397 ubi nobiscum aderant aremannos ... et ... <u>homines</u> Franciscos et alii plures
74. CDT 45 ubi aderant nobiscum Thomas diaconus , Otus presbyter ... et alios <u>plures</u>
75. MED 397 ubi nobiscum aderant aremannos ... et ... homines Franciscos et alii <u>plures</u>
76. CDL 178 et dum Gumpranda aut <u>filias</u> meas aduiuere meruerint
77. MED 240 et homines meos omnes, servos et ancillas, aldiones adque <u>aldias</u> ... in tua sint potestatem
78. CDL 175 seruos uero uel ancillas seu <u>aldiones</u> meos ... sint omnes in potestate suprascripte eclesiae
79. MED 240 et homines meos omnes, servos et ancillas, <u>aldiones</u> adque aldias ... in tua sint potestatem
80. CDL 175 seruos uero uel <u>ancillas</u> seu aldiones meos ... sint omnes in potestate suprascripte eclesiae
81. MED 240 et homines meos omnes, servos et <u>ancillas,</u> aldiones adque aldias ... in tua sint potestatem
82. MED 240 et <u>homines</u> meos omnes, servos et ancillas, aldiones adque aldias ... in tua sint potestatem
83. CDL 175 <u>seruos</u> uero uel ancillas seu aldiones meos ... sint omnes in potestate suprascripte eclesiae
84. MED 240 et homines meos omnes, <u>servos</u> et ancillas, aldiones adque aldias ... in tua sint potestatem
85. CDL 175 omnes seruos uel ancillas seu <u>aldiones</u> meos ... in eius sit potestate
86. CDL 175 omnes seruos uel <u>ancillas</u> seu aldiones meos ... in eius sit potestate
87. MED 231 omnes res ipsas et <u>homines</u> de ipsam congregationem monacarum ... sit potestatem

88. CDL 175 omnes <u>seruos</u> uel ancillas seu aldiones meos ... in eius sit potestate

E. Unergative clauses

E1. Nominative-form inanimate plural subjects of unergative clauses (all 4)

E1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
89.	conjungo	convene	partes	coniunxerunt		0	1	1	0	1	1	1	1	0	0	6
90.	revorto	return	partes	reversi		0	1	1	0	1	1	1	1	0	0	6
91.	venio	come	partes	venerunt		0	1	1	0	1	1	1	1	0	0	6
92.	venio	come	partes	venerunt		0	1	1	0	1	1	1	1	0	0	6
					Σ	0	4	4	0	4	4	4	4	0	0	m. 6

89. CDL 182 et in constituta die iteratim ambe <u>partes</u> ante nos coniunxerunt
90. MED 202 et dum utraque <u>partes</u> in constituto reversi fuerunt
91. MED 742 in constituta die in eadem curte ducalae ... utraque <u>partes</u> venerunt
92. MED 564 in statuta die venerunt ambe <u>partes</u> denuo nostri presencia

E2. Accusative-form inanimate plural subjects of unergative clauses (all 2)

E2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
93.	revorto	return	partes	reversi		0	1	1	0	1	1	0	1	0	0	5
94.	revorto	return	partes	reversi		0	1	1	0	1	1	0	1	0	0	5
					Σ	0	2	2	0	2	2	0	2	0	0	m. 5

93. MED 397 dum in constituto ambas <u>partes</u> ante nos reversi fuisserunt in iudicio
94. MED 397 dum in constituto ambas <u>partes</u> ante nos reversi fuisserunt in iudicio

E3. Nominative-form animate plural subjects of unergative clauses (all 15)

E3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
95.	concurro	hasten	cunctis	concurret		0	1	0	0	1	1	0	1	0	0	4
96.	debeo	have to (serve)	filii	deueas	seruire	0	1	0	0	1	1	0	0	0	0	3
97.	exeo	leave	filiis	exeas		0	1	1	0	1	1	0	1	0	0	5
98.	festino	hasten	cunctis	festinet	concurrere	0	1	0	0	1	1	0	1	0	0	4
99.	nolo	not want (to serve)	filii	nuluere	seruire	0	1	0	0	1	0	0	0	0	0	2
100.	possum	can (contest)	filiis	possa	contrare	0	1	0	0	1	0	0	0	0	0	2
101.	praesumo	dare to (injure)	heridis	presumserimus	molestare	0	1	0	0	1	1	0	1	0	0	4
102.	tento	try (to go)	successores	temtauerit	ire	0	1	1	0	1	1	0	1	0	0	5

	verb lemma	verb sense	subject	verb		2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
103.	testificor	testify	sacerdotes	testificati		0	1	1	0	1	1	0	1	0	0	5
104.	testificor	testify	sacerdotes	testificati		0	1	1	0	1	1	0	1	0	0	5
105.	testificor	testify	sacerdotes	testificati		0	1	1	0	1	1	1	1	0	0	6
106.	testor	testify	clerici	testati		0	1	1	0	1	1	1	1	0	0	6
107.	vado	go	filiis	vadam		0	1	1	0	1	1	0	1	0	0	5
108.	venio	come	homines	veneritis		0	1	1	0	1	1	0	1	0	0	5
109.	venio	come	missi	veneritis		0	1	1	0	1	1	0	1	0	0	5
					Σ	0	15	9	0	15	13	2	12	0	0	66
					%	0	100	60	0	100	87	13	80	0	0	44

95. CDL 248 quoniam uoluntas Dei [est] ut cunctis pro anima sua remedium concurret
96. CDL 31 ut filii nostri ibidem in ipso monasterio Domino seruire deueas
97. CDL 192 tunc exeas ipsis filiis meis cum omnis ris muuilem de ipsa casa
98. CDT 46 bolumtas Dei est ut cunctis pro anime sue remedio festinet concurrere
99. CDL 261 et si forsitans ipsi filii mei ad ipsa ecclesia seruire nuluere
100. CDT 88 ut nequem ego Ascolfus nequem filiis ... contra hanc datio mea contrare possa
101. CDL 134 si ... cuntra hanc cartulam ... in alicho molestare presumserimus [nos] uel heridis nostris
102. CDL 160 nos seo heredes uel successores nostri contra hanc cartulas ire temtauerit
103. MED 309 et dum prefati sacerdotes taliter testificati fuissent
104. MED 385 et dum prefati sacerdotes taliter testificati fuissent
105. MED 385 sicut ipsi sacerdotes nostri de ipso iurando testificati sunt
106. MED 539 isti omnes prefati clerici per sagramentum ad sancta Dei evangelia unianimiter testati sunt
107. CDT 89 quam filiis qua mulierem tua vadam [= vadant] ubi volueris [= voluerint]
108. MED 461 et quando inibi vos aut missi tui, id est homines tres, super vendemmia veneritis
109. MED 461 et quando inibi vos aut missi tui, id est homines tres, super vendemmia veneritis

E4. Accusative-form animate plural subjects of unergative clauses (all 3)

E4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
110.	debeo	have to (come)	homines	debeant	venire	0	1	1	0	1	1	0	0	0	0	4
111.	debeo	have to (come)	homines	debeant	venire	0	1	1	0	1	1	0	0	0	0	4
112.	debeo	have to (come)	homines	debeas	venire	0	1	1	0	1	1	0	0	0	0	4
					Σ	0	3	3	0	3	3	0	0	0	0	m. 4

110. MED 710 et homines illos qui in suprascripta casa abitantes fuerit a mandato vestro venire debeant
111. MED 725 et homines illos qui in suprascriptam casa abitantes fuerint a mandato vestro venire debeant
112. MED 781 et homines illos qui in suprascripta casa abitantes fuerit ad mandato vestro venire debeas

F. Transitive clauses

F1. Nominative-form inanimate plural subjects of transitive clauses (all 5)

F1	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
113.	acquiro	gain	anime	adquirat	medillam	1	0	1	1	0	1	0	0	0	1	5
114.	capio	gather	mentis	capiant	#	0	0	0	0	0	1	0	0	0	0	1

						2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
115.	contineo	contain	cartulae	contenet	cod	1	0	0	0	0	1	1	0	0	1	4
116.	eligo	elect	partis	eligerent	quem	1	1	1	0	1	1	0	1	0	1	7
117.	profiteor	profess	partes	professi	ACI	1	1	1	0	1	1	1	1	0	0	7
					Σ	4	2	3	1	2	5	2	2	0	3	m. 4.8

113. CDL 267 qualiter anime nostre medillam adquirat [= acquirant]
114. CDL 16 suoque arbitrio perfectissime capiant homane <u>mentis</u>
115. CDT 31 at tale husuo [= usum], quale mihi persolvere dibuit, cod <u>cartulae</u> ..... contenet
116. CDL 24 quem iniui priorem et gubernatorem due <u>partis</u> ex nouis una cum monaci ipsi eligerent
117. MED 564 dum ambe <u>partes</u> nullum testem propter longinquitatem inde dare professi sunt

F2. Accusative-form inanimate plural subjects of transitive clauses (all 2)

F2	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
118.	rogo	request	parti	rogaverunt	ACI	1	1	0	0	1	1	1	1	0	0	6
119.	rogo	request	partis	rogavimus	ACI	1	1	0	0	1	1	1	1	0	0	6
					Σ	2	2	0	0	2	2	2	2	0	0	m. 6

118. CDT 23 hanc cartulam ambas <u>parti</u> fieri rogaverunt
119. MED 252 et Rachipert presbitero ambas <u>partis</u> scrivere rogavimus

F3. Nominative-form animate plural subjects of transitive clauses (all 25)

F3	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
120.	compono	recompense	filii	componamus	solidos	1	1	1	0	1	1	0	1	0	0	6
121.	debeo	have to (have)	monachi	debeat	hauire ortum	1	0	0	0	0	0	1	0	0	0	2
122.	debeo	have to (pay)	persone	debeatis	reddere denarios	1	1	1	0	1	1	0	1	0	0	6
123.	debeo	have to (pray)	sacerdotes	debeant	deprecare Dominum	1	1	0	0	1	1	0	0	0	1	5
124.	debeo	have to (give)	filii	deueas	dare #	0	1	1	0	1	1	0	0	0	0	4
125.	dico	say	sacerdotes	dixerunt	ut clause	1	1	1	0	1	1	1	1	0	0	7
126.	eligo	elect	monaci	eligere	quem (< ipsum)	1	1	1	0	1	1	0	1	1	1	8
127.	facio	make (a will)	parentes	fecimus	uoluntatem	1	1	1	0	1	1	1	1	0	0	7
128.	facio	make (a will)	parentes	fecerunt	voluntate	1	1	1	0	1	1	1	1	0	0	7
129.	facio	do	parentis	ficerunt	quas (< case)	1	1	1	0	1	1	1	1	1	1	9
130.	habeo	have	parentes	habeant	potestate	1	0	0	0	0	1	0	0	0	0	2
131.	habeo	have	massarii	abuerunt	quantum	1	0	0	0	0	1	1	0	0	0	3
132.	habeo	have	filii	haueas	licentiam	1	0	0	0	0	0	0	0	0	0	1
133.	habeo	have	monachi	abent	congregationem	1	0	0	0	0	1	1	0	0	0	3
134.	habeo	have	filii	haueas	licentia	1	0	0	0	0	1	0	0	0	0	2
135.	persolvo	discharge	filiis	persoluant	redditum	1	1	1	0	1	1	0	1	0	1	7
136.	possideo	possess	filiis	possedeas	#	0	0	0	0	0	1	0	0	0	0	1
137.	praesumo	dare (to expel)	rectores	presumserimus	expellere uos	1	1	1	0	1	1	0	1	1	1	8

138.	promitto	promise	posteris	promitto	ACI	1	1	1	1	1	1	1	1	0	0	8
139.	recordor	recall	omnes	recordati	#	0	0	0	0	0	1	1	0	0	0	2
140.	reddo	pay	nepotes	redderimus	solidum	1	1	0	0	1	0	0	1	0	1	5
141.	superimpono	overburden	posteris	superinposuerimus	angaria	1	1	0	0	1	1	0	1	0	0	5
142.	video	seem (to have)	massarii	videtur	habere quantum	1	0	0	0	0	1	0	0	0	0	2
143.	volo	want (to permit)	filii	uolueret	consentire #	1	0	0	0	0	0	0	0	0	0	1
144.	volo	want (to pay)	filii	uoluerint	reddere oleum	1	1	1	0	1	1	0	1	0	1	7
					∑	22	15	12	1	15	21	9	13	3	7	118
					%	88	60	48	4	60	84	36	52	12	28	47

120. MED 409 ut ego aut suprascripti <u>filii</u> mei componamus tibi ..... solidos sexaginta
121. CDL 28 ubi sibi abbas uel <u>monachi</u> iniui consistentis ortum uel pigmentarium hauire debeat
122. MED 413 tu et ille alie duo <u>persone</u> ... reddere debeatis semper in mense magio argentum bonos denarios
123. CDL 114 ut <u>sacerdotes</u> ... pro facinoribus meis Dominum deprecare debeant
124. CDL 261 et post meo decesso <u>filii</u> mei similiter dare deueas, sicut ego supra premisi
125. MED 309 prenominati <u>sacerdotes</u> iurando dixerunt ut veritatem fuissent
126. CDL 31 quem siui ipsi <u>monaci</u> de ea congrecationem eligere, ipsum aueat ordinatum
127. CDL 246 tam <u>parentes</u> mei quam et ego uoluntatem et imperationem fecimus
128. MED 335 qualiter <u>parentes</u> mei pro ipsa re et ecclesia fecerunt voluntate
129. CDL 85 cunsuetudo ipseius case quas <u>parentis</u> nostris ficerunt per singulus annus
130. CDL 93 potestate habeant heredes aut <u>parentes</u> de ipsi diacones prendere et tenere et defendere
131. MED 805 omnia quantum ... suprascripti <u>massarii</u> exinde ad manus suas abuerunt
132. CDL 261 et hoc uolo <u>filii</u> mei post parte sua tulta licentiam non haueas alii homini uendere
133. CDT 14 in monasterio ubi <u>monachi</u> abent congregationem
134. CDL 254 ut licentia haueas toti <u>filii</u> mei comuniter in ipsa Dei eclesia presbitero ordinando
135. CDL 131 eorum <u>filiis</u> uel nepotes ... per uno queque anno persoluant redditum case
136. CDL 61 ita ut ab hodierna dies securiter possedeas tu uel <u>filiis</u> tuis
137. CDL 242 si ... ego aut <u>rectores</u>... uos de ipsa casa et rem foris expellere ... presumserimus
138. CDT 86 conponere promitto me ego ... vel <u>posteris</u> nostris ... solidos numero quinquagenta
139. MED 309 qui <u>omnes</u> unanimiter recordati sunt dicentes
140. CDL 285 si ego ... uel ipsi predicti <u>nepotes</u> mei ... ipsum solidum ... non redderimus
141. CDT 86 si ego ... vel <u>posteris</u> nostris ... plus angaria vel pensione superinposuerimus
142. MED 756 omnia quantum ... suprascripti <u>massarii</u> exinde ad manus suas abere videtur
143. CDL 254 et si ... <u>filii</u> mei ipsei consintire non uolueret
144. CDL 276 et si <u>filii</u> mei praedictum oleum in ipsa ecclesia reddere uoluerint

F4. Accusative-form animate plural subjects of transitive clauses (all 19)

F4	verb lemma	verb sense	subject	verb	object	2 participants	action	telicity	punctuality	volitionality	affirmation	realis mode	agency	affected O	individuat. O	Σ
145.	accipio	receive	eginos	adcipiant	consulationem	1	0	1	0	0	1	0	0	0	0	3
146.	accipio	receive	peregrinos	adcipiant	consulationem	1	0	1	0	0	1	0	0	0	0	3
147.	accipio	receive	eginos	accepiant	consulationem	1	0	1	0	0	1	0	0	0	0	3
148.	accipio	receive	pauperos	accepiant	consulationem	1	0	1	0	0	1	0	0	0	0	3
149.	accipio	receive	filias	accipiant	medietatem	1	0	1	0	0	1	0	0	0	1	4
150.	compono	recompense	supcessore	componam	soledus	1	1	1	0	1	1	0	1	0	0	6
151.	compono	recompense	posterus	componamus	solidos	1	1	1	0	1	1	0	1	0	0	6
152.	compono (sum)	recompense	successores	(sint) componiturus	#	0	1	1	0	1	1	0	0	0	0	4
153.	compono (sum)	recompense	successores	(sit) componiturus	solidos	1	1	1	0	1	1	0	0	0	0	5
154.	debeo	have to (have)	successores	debeamus	abere solidos	1	0	0	0	0	1	0	0	0	0	2

155.	habeo	have	successores	abeatis	potestatem	1	0	0	0	0	1	0	0	0	0	2
156.	habeo	have	ancilla	aueat	potestatem	1	0	0	0	0	1	0	0	0	0	2
157.	habeo	have	seruos	aueat	potestatem	1	0	0	0	0	1	0	0	0	0	2
158.	habeo	have	filius	habeat	potestatem	1	0	0	0	0	1	0	0	0	0	2
159.	quaero	try (to violate)	filios	quesieremus	disrumpi cartulam	1	1	1	0	1	1	0	1	1	1	8
160.	repromitto	promit	meus	repromitto	ut clause	1	1	1	1	1	1	1	1	0	0	8
161.	volo	want (to invite)	filias	uoluerint	inuitare quem	1	1	1	0	1	0	0	1	0	1	6
162.	volo	want (to invite)	ancillas	uoluerint	inuitare quem	1	1	1	0	1	0	0	1	0	1	6
163.	volo	want (to decree)	posterus	uolueris	iudicare quem	1	1	1	0	1	1	0	1	0	1	7
					Σ	18	9	14	1	9	17	1	7	1	5	82
					%	95	47	74	5.3	47	89	5.3	37	5.3	26	43

145. CDL 127 et uolomus ... ut peregrinos adque eginos cotidie consulationem adcipiant
146. CDL 127 et uolomus ... ut peregrinos adque eginos cotidie consulationem adcipiant
147. CDL 127 et iniui pauperos adque eginos modernos et futures temporibus consulationem accepiant
148. CDL 127 et iniui pauperos adque eginos modernos et futures temporibus consulationem accepiant
149. CDL 248 medietatem accipiant filiis aut filias meas
150. MED 217 componam ego ... aut meus postero supcessore ... soledus numerum quinquagentas
151. MED 180 componamus ... ego Lampert presbitero vel posterus meus auri solidos nomiro quadraginta
152. MED 425 volo ut sint componiturus ipse heredes vel successores meos
153. MED 530 tunc sit componiturus ipse heredes vel successores meos tibi ... penam argentum solidos octuagenta
154. MED 428 ego vel heredes aut successores meos suprascriptos solidos aput nos abere debeamus
155. MED 401 ut tu vel successores tuos potestatem abeatis reintroire
156. CDL 40 et seruos uel ancilla ... uolo ut liueram aueat potestatem
157. CDL 40 et seruos uel ancilla ... uolo ut liueram aueat potestatem
158. CDL 30 nam filius meus uel heredis meus nullam ibidem habeat potestatem dominandi
159. CDT 12 si ... ego Ratchausu aut quolivet heredes vel filios meo hanc cartulam disrumpi quesieremus
160. MED 398 repromitto ego qui supra Gundulu vel meus ..... ut si ... non adimpleverimus
161. CDL 178 nullus sacerdos ... auitare presumat, nisi quem ... Gumpranda aut filias meas inuitare uoluerint
162. CDL 30 nullus sacerdus ibidem habitare presummat, nisi quem ipsas ancillas Dei inuitare uoluerint
163. CDL 144 omnia ... quem tu ... iudicare uel facere uolueris tu ... uel posterus tuo

Appendix 4.4. Construction type as a function of transitivity degree.

Construction type		Transitivity degree									Total
		0-1	2	3	4	5	6	7	8	9	
S_O	N	114	83	10	0	0	0	0	0	0	207
	%	91.2%	76%	28%	0%	0%	0%	0%	0%	0%	43.9%
	residual	12.4	7.7	-2.0	-5.6	-6.2	-7.5	-4.7	-4.5	-2.2	
S_A	N	0	3	6	17	24	27	0	0	0	77
	%	0%	3%	17%	46%	55%	44%	0%	0%	0%	16%
	residual	-5.8	-4.4	0.1	5.1	7.2	6.2	-2.4	-2.3	-1.1	
A	N	11	23	20	20	20	35	27	25	6	187
	%	9%	21%	56%	54%	46%	57%	100%	100%	100%	39.7%
	residual	-8.2	-4.5	2.0	1.9	0.8	2.9	6.6	6.3	3.0	
Total	N	125	109	36	37	44	62	27	25	6	471
Chi-square		$\chi^2 = 410.72$, $df = 16$, $p < 0.001$									

Appendix 4.5. Animacy as a function of transitivity degree.

Animacy		Transitivity degree										Total
		0	1	2	3	4	5	6	7	8	9	
inanimate	N	0	78	26	8	4	7	13	8	4	2	150
	%	0%	65%	24%	22%	11%	16%	21%	30%	16%	33%	32 %
	residual	-1.5	9	-2	-1.3	-2.9	-2.4	-2	-0.3	-1.7	0.1	
animate	N	5	33	42	19	26	33	15	9	10	2	194
	%	100%	28%	39%	53%	70%	75%	24%	33%	40%	33%	41 %
	residual	2.7	-3.5	-0.6	1.5	3.7	4.8	-2.9	-0.9	-0.1	-0.4	
personal	N	0	9	41	9	7	4	34	10	11	2	127
	%	0%	7.5%	38%	25%	19%	9.1%	55%	37%	44%	33%	27 %
	residual	-1.4	-5.6	2.9	-0.3	-1.1	-2.8	5.3	1.2	2	0.4	
Total	N	5	120	109	36	37	44	62	27	25	6	471
Chi-square		$\chi^2 = 145.50$, $df = 18$, $p < 0.001$										

Appendix 5.1. Contingency table of subject case and linear distance from the NP head in structural category A1.

Case		Linear distance from the NP head								Total
		1	2	3	4	5	6	7	8–10	
nominative	N	238	192	62	46	11	11	4	7	571
	%	45.9%	61.0%	46%	51%	36%	79%	25%	37%	50.1%
	residual	**-2.6**	**4.5**	-1.0	0.1	-1.7	**2.1**	**-2.0**	-1.2	
accusative	N	280	123	73	45	20	3	12	12	568
	%	54.1%	39%	54%	50%	65%	21%	75%	63%	49.9%
	residual	**2.6**	**-4.5**	1.0	-0.1	1.7	**-2.1**	**2.0**	1.2	
Total	N	518	315	135	91	31	14	16	19	1,139
Chi-square		$\chi^2 = 31.92$, $df = 7$, $p < 0.001$								

Appendix 5.2. Frequencies of subjects as a function of subject NP length in LLCT ($\chi^2 = 7.05$, $df = 8$, $p = 0.531$).
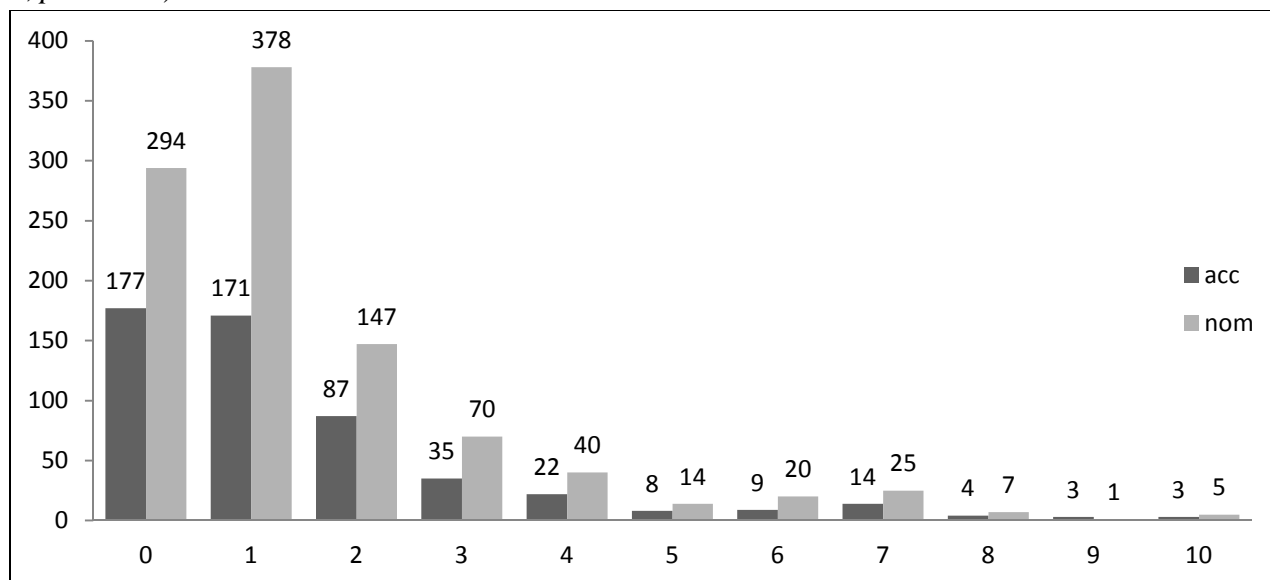
Appendix 5.3. Contingency table of subject case and distance in LLCT.

Case		Distance										Total
		-8	-7	-6	-5	-4	-3	-2	-1	1	2	
nominative	N	39	37	36	59	71	76	92	134	94	44	682
	%	74%	73%	55%	62%	54%	68%	80%	82.2%	62%	56%	67.0%
	residual	1.0	0.9	**-2.2**	-1.2	**-3.3**	0.2	**3.1**	**4.5**	-1.5	**-2.2**	
accusative	N	14	14	30	37	60	36	23	29	58	35	336
	%	26%	28%	46%	39%	45.8%	32.1%	20.0%	17.8%	38.2%	44%	33.0%
	residual	-1.0	-0.9	**2.2**	1.2	**3.3**	-0.2	**-3.1**	**-4.5**	1.5	**2.2**	
Total	N	53	51	66	96	131	112	115	163	152	79	1,018
Chi-Square		$\chi^2 = 49.69$, $df = 9$, $p < 0.001$										

Appendix 5.4. Decision tree for subject case (dependent) and distance (independent) in the entire LLCT.
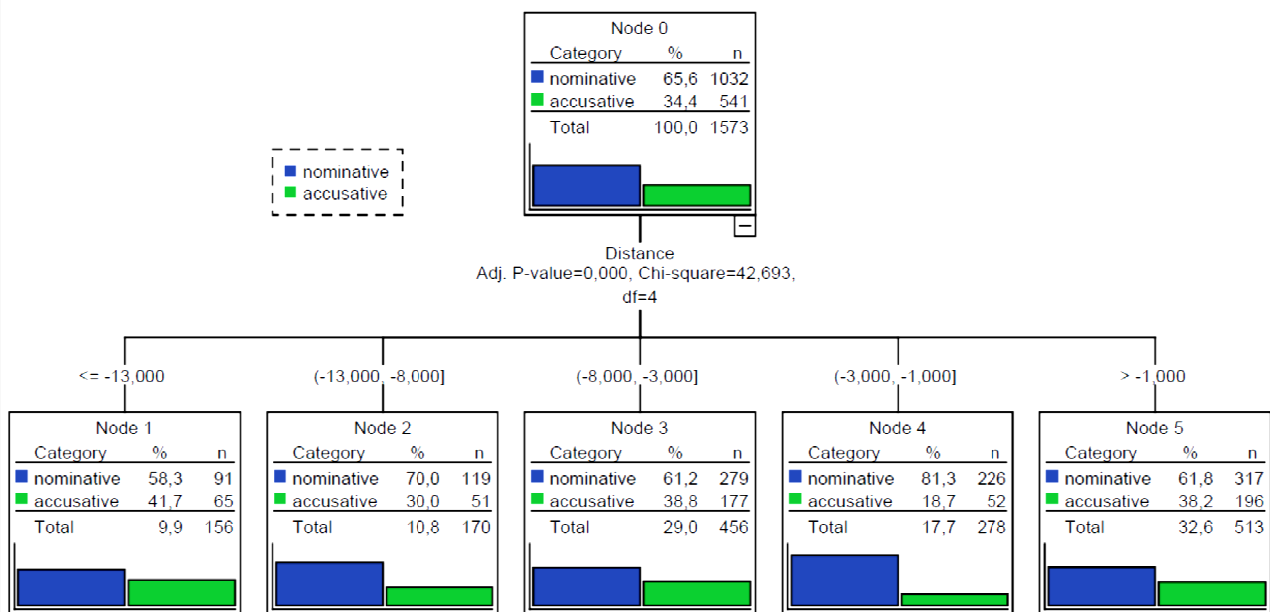
Square bracket ']' stands for 'up to and including' and round bracket '(' stands for 'from but not including'. E.g. (-3.000, -1.000] in node 4 stands for positions -2 and -1.

Appendix 5.5. Decision tree for subject case (dependent) and distance (independent) in the subcorpus.
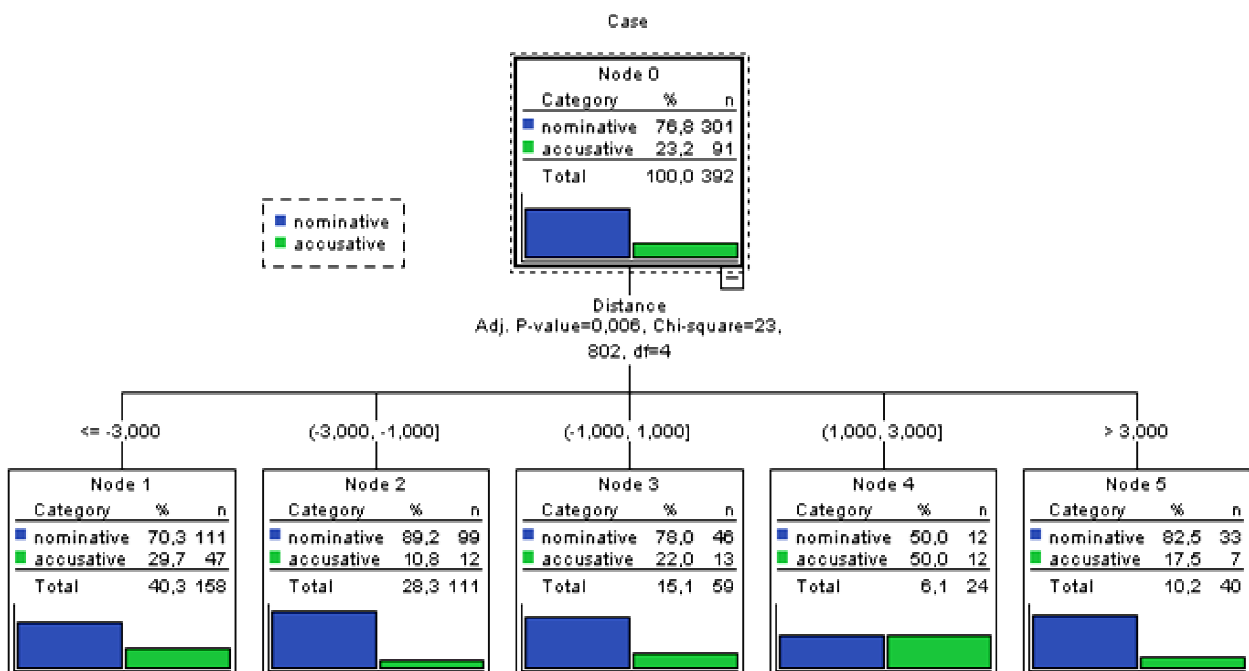
Appendix 5.6. Contingency table of subject case and distance in structural category A1 (range -5 to +2).

Case		Distance							Total
		-5	-4	-3	-2	-1	1	2	
nomina-tive	N	47	61	53	83	120	83	36	483
	%	63%	54%	71%	81%	81.6%	61%	53%	67.6%
	residual	-1.0	**-3.4**	0.6	**3.2**	**4.1**	-1.7	**-2.7**	
accusative	N	28	52	22	19	27	52	32	232
	%	37%	46%	29%	19%	18%	39%	47%	32.4%
	residual	1.0	**3.4**	-0.6	**-3.2**	**-4.1**	1.7	**2.7**	
Total	N	75	113	75	102	147	135	68	715
Chi-square		$\chi^2 = 41.72, df = 6, p < 0.001$							

Appendix 5.7. Contingency table of construction type and distance in LLCT (range -8 to +2).

Construction type		Distance										Total
		-8	-7	-6	-5	-4	-3	-2	-1	1	2	
A	N	27	26	37	73	64	50	65	51	40	47	480
	%	51%	51%	56%	76%	49%	45%	57%	31%	26%	60%	47.2%
	residual	0.6	0.6	1.5	**6.0**	0.4	-0.6	**2.1**	**-4.4**	**-5.6**	**2.3**	
S_A	N	4	7	6	3	2	8	6	14	3	2	55
	%	8%	14%	9%	3%	2%	7%	5%	9%	2%	3%	5%
	residual	0.7	**2.7**	1.4	-1.0	**-2.1**	0.9	-0.1	**2.0**	**-2.0**	-1.2	
S_O	N	16	15	19	17	48	39	33	71	107	28	393
	%	30%	29%	29%	18%	37%	35%	29%	44%	70%	35%	38.6%
	residual	-1.3	-1.4	-1.7	**-4.4**	-0.5	-0.9	**-2.3**	1.4	**8.7**	-0.6	
S_P	N	6	3	4	3	17	15	11	27	2	2	90
	%	11%	6%	6%	3%	13%	13%	10%	17%	1%	3%	9%
	residual	0.7	-0.8	-0.8	**-2.1**	1.8	1.8	0.3	**3.8**	**-3.5**	**-2.1**	
Total	N	53	51	66	96	131	112	115	163	152	79	1,018
Chi-square		$\chi^2 = 161.08, df = 27, p < 0.001$										

Appendix 5.8. Contingency table of animacy and distance in LLCT (range -8 to +2).

Animacy		Distance -8	-7	-6	-5	-4	-3	-2	-1	1	2	Total
inanimate	N	18	15	17	20	32	46	39	98	86	20	391
	%	34%	29%	26%	21%	24%	41%	34%	60%	57%	25%	38.4%
	residual	-0.7	-1.4	**-2.2**	**-3.7**	**-3.5**	0.6	-1.1	**6.2**	**5.0**	**-2.5**	
animate	N	15	20	13	15	16	29	26	21	16	12	183
	%	28%	39%	20%	16%	12%	26%	23%	13%	11%	15%	18.0%
	residual	**2.0**	**4.1**	0.4	-0.6	-1.8	**2.3**	1.4	-1.8	**-2.6**	-0.7	
personal	N	20	16	36	61	83	37	50	44	50	47	444
	%	38%	31%	55%	64%	63%	33%	44%	27%	33%	60%	43.6%
	residual	-0.9	-1.8	1.9	**4.1**	**4.9**	**-2.4**	0	**-4.7**	**-2.9**	**3.0**	
Total	N	53	51	66	96	131	112	115	163	152	79	1,018
Chi-square		$\chi^2 = 133.53$, $df = 18$, $p < 0.001$										

Appendix 5.9. Percentage distribution of *esse* as a function of linear position in LLCT.

45% 40% 35% 30% 25% 20% 15% 10% 5% 0% ≤ -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 1 2 3 4 5 6 7 8 9 ≥ 10