UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Early Medieval Scribes' Command of Latin Spelling and Grammar : A Quantitative Approach

Korkiakangas, Timo

Dahlgren, Sonja; Leiwo, Martti; Vierros, Marja

2024-12

http://hdl.handle.net/10138/590879

Korkiakangas, T 2024, Early Medieval Scribes' Command of Latin Spelling and Grammar : A Quantitative Approach. in S Dahlgren, M Leiwo & M Vierros (eds), Scribes and Language Use in the Graeco-Roman World. Commentationes Humanarum Litterarum, vol. 147, Societas Scientiarum Fennica, Helsinki, pp. 253-278.

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

10. Early Medieval Scribes' Command of Latin Spelling and Grammar: A Quantitative Approach

Timo Korkiakangas

1 Background, objectives and hypotheses

This Chapter investigates the association between spelling variation and certain classical Latin grammatical categories in early medieval documents of Tuscany. The investigation will include linguistic, sociolinguistic, cognitive and educational aspects. The Latin of the Tuscan documents of the eighth and ninth centuries is often highly nonstandard in terms of both grammar and spelling. Although documentary Latin can be described as a language variety of its own with several special terms and fixed expressions (Bartoli Langeli 2006: 25ff.; Larson 2012: 66; cf. Herman 1992: 185), it can also be justly compared to the classical Latin spelling and morphological norm, which the scribes still seemed to pursue to varying extents and which some scribes managed quite well (Korkiakangas 2016: 36; Amsler 1993). The actual models were the Bible and the Christian authors of late antiquity whose Latin can be described post-classical in stylistic terms but is essentially classical in regard to grammar. Classical and standard Latin will be used as synonyms in the following Sections.

Leaving aside the few tentative attempts at recording the vernacular in writing that have come down to us (Castellani 1976), classical Latin was the only language variety with an established writing system that was available. Written Italian vernaculars only came into broad use later in the Middle Ages. Thus, the considerable linguistic variation found in the documents results from the fact that the spoken language (the scribes' L1) and written Latin (L2 to be learnt) had been deviating for a significant period, and the documentary scribes allowed their L1 to interfere with the L2 to varying degrees. In general, it can be assumed that the degree to which scribes produced classical Latin forms and constructions reflected their learned language skills, i.e. the mastery that scribes had acquired in the L2 under their education.

The objective of this Chapter is to examine quantitatively whether the variation in spelling and the variation in employing grammatical categories are associated with each other. In other words, do skilled spellers utilise classical grammar better than do less skilled ones? My working hypothesis is that the more

nonstandardly a scribe spelled, the more frequently he produced novel Romance-type forms and constructions (innovative L1 features) and the poorer was his command of those classical Latin forms and constructions that had vanished or were in decline in late Latin (conservative L2 features). This hypothesis will be statistically tested in Section 5. As mentioned previously, the innovative features had crept into the written code from the spoken idiom of the time, while the conservative features were derived from centuries-old legal Latin. The situation was likely to differ fundamentally from modern-language acquisition situations in which oral usage is also learnt, unlike with documentary Latin.

The conservative grammatical features were memorised during L2 acquisition in school and/or by immersion in authoritative model texts written in standard Latin. The same applies to spelling, which is likely to have occupied an important position in instruction (de Paolis 2010: 229–91; Law 1982: 40–41; cf. Cribiore 2001: 186). My working hypothesis assumes that the scribes mastered the conservative school-learned, memorisable features more or less as a bundle and, consequently, these features pattern with each other and differ from the innovative spoken-language features. Here, the particular linguistic situation in early medieval Italy has to be taken into due consideration: standard Latin was exclusively a medium of written communication and was not meant to be spoken, while all oral communication took place in the vernacular variety, which must have resembled Old Italian to a significant degree (Herman 1991: 39–41; Walsh 1991: 205–7; Zamboni 2000: 182; Lüdtke 1964). Only by the Carolingian ecclesiastical reforms at the end of the eighth century did the emperor require that liturgy and prayers were recited in a restored Latin. However, we do not know what this meant in practice, or what consequences the imperial order had in the kingdom of Italy (Bartoli Langeli 2006: 26–8). It is also often stated that the conceptual differentiation between spoken vernacular and written Latin took place in Gaul in the early ninth century as a result of the Carolingian reforms (e.g. Wright 1991: 109). In Italy, this conceptual rupture is likely to have happened later, given that the first substantial vernacular texts only appear from the tenth or eleventh century onwards (Castellani 1976).

Since the scribes were not and did not seek to become speakers of the L2 they wrote, the mastery of both L2 spelling and grammar was the result of learning; thus, it is justifiable to assume that the spelling- and grammar-related features pattern in a similar way. By contrast, the vast majority of modern-language learners also learn to speak the language, and oral communication is often the most important motivation to learn a language. Consequently, some fluent L2 speakers of modern English, for example, may have invested much less effort in learning

its correct spelling. Thus, for spoken modern languages in which the grapheme/phoneme correspondence is irregular, the connection between grammatical and spelling competence is not likely to be as close as it was for written early medieval Tuscan Latin, in which both types of competences served written communication exclusively while vernacular was used for oral communication.

It is true that the grammatical systems of Latin and Old Italian were not very different from each other; thus, L2 acquisition was not necessarily a particularly difficult task. On the other hand, the distance between the two languages may have appeared smaller than it actually was, which may have led to indifference to and a failed recognition of the differences that existed in reality. This non-recognition is also likely to have been the reason for the long-standing conceptual non-differentiation between Latin and the vernacular in Italy.

2 Data

Early medieval documents can be considered a more useful resource for variationist linguistic analysis than can most literary texts. As most are originals, the documents have not been subjected to emendations and modifications during textual transmission, unlike texts that have survived in copied manuscripts. The documents also allow addressing sociolinguistically inspired research questions because the writers' names, the places of writing and the dates of almost all the documents are known. Early medieval Italian documents show notable linguistic and formula variation. Along with the fact that no contemporary formulary book is known from Italy, this variation suggests that early medieval Italian documents were not copied slavishly from model document collections, but instead the scribes memorised the legally required formulae, which they then applied and modified as necessary (Schiaparelli 1933: 3). It is easy to see how such a practice resulted in variation at all levels of language, notwithstanding the underlying formulaicity. Besides, standard formulae could not be used to express those details that varied from one legal transaction to another. The scribes had to improvise the case-specific contents on their own, often drawing on the spoken language of the time (Sabatini 1965).

The data of this study is the *Late Latin Charter Treebank* (LLCT). LLCT is a lemmatised and morphologically and syntactically annotated corpus which consists of 519 hand-written private documents (225,834 tokens). They record the selling and buying of landed property and were written by 176 scribes in Lombard and Carolingian Tuscany (mainly Lucca, Siena and Pisa) between 714

and 869 CE. The documents survive as originals, with the exception of a few contemporary copies. The annotation follows the *Guidelines for the Syntactic Annotation of Latin Treebanks* (Bamman *et al.* 2007), with the additions and modifications described in Korkiakangas and Passarotti (2011). The LLCT is based on three copyright-free diplomatic editions and is available for download in Prague Markup Language format (PML) (see References).

3 Spelling variation

Spelling correctness is utilised widely in modern-language studies (e.g. Burt 2006). As for Latin, the rarity of coherent text corpora with original spelling has constrained this kind of research until recently. This Chapter exploits the method of quantifying spelling variation developed in Korkiakangas (2017). I quantify spelling variation as a continuous variable that indicates the percentage of all nonstandardly spelled characters in all characters written by a scribe (e.g. 187 nonstandard spellings in 1,306 characters of the document *ChLA* 732 corresponds to a spelling correctness of 85.7%).

First, the nonstandard word forms of LLCT are normalised into respective standard-Latin forms. An edit distance value is then calculated between all the attested word forms and their standard-Latin counterparts, regardless of whether the attested ones were originally spelled standardly or nonstandardly. Edit distance is the minimum number of single-character changes required to transform one text string into another. I utilise the Levenshtein edit distance where these changes can be insertions, deletions or substitutions. For example, the Levenshtein distance for the attested form *abi* and the respective standard Latin form *habes* 'you have' is three because the two forms differ from each other by three single-character changes (*h* > Ø, *e* > *i*, *s* > Ø). Since the standard form has five characters, the spelling correctness percentage of the attested form *abi* is 40 (two in five). This percentage of correct spelling is then averaged for all the characters found in the documents written by each of the 176 LLCT scribes.

The normalised word forms were acquired by the lemmatizing and morphological tagging of the Open Office Latin lexicon using Whitaker's WORDS tagger (Words 1.97FC). This procedure yielded a two-million-item list of classical Latin word form/lemma/morphological tag entries. I then matched the lemma/morphological tag pair of each LLCT word form with this entry list and obtained the respective classical word forms as output. When calculating the edit distance, the letters *u* and *v*, *i* and *j*, as well as *k* and *c,* were counted

as equivalents because they are merely alternative graphical representations of the same graphemes. Fragmentary and abbreviated words were excluded from the analysis, provided that their inflectional endings could not be restored and expanded with certainty. I explain the procedure in detail in Korkiakangas (2017).

From a linguistic point of view, the spelling deviations of documentary Latin are caused by various motivations. Apart from purely typographical errors, most nonstandard spellings are related to the substantial phonological change that had taken place. However, this phonological change was often tightly connected to morphological (and even syntactic) evolution. For example, the prepositional phrase *ad uscio,* instead of the standard *ad ostium* 'by the door', involves phonologically motivated changes, such as the closure of the stressed /*o*/ as /*u*/, the palatalization of *sti* as *sci* [ʃ:i], and the reduction of the original inflectional morpheme *-um* as *-o*. This reduction can be interpreted as resulting from the concomitant phonological merger of the syllables in question and from the levelling in prepositional complementation. The latter means that no distinction was made between the originally accusative- and ablative-form complements of prepositions (*-um* and *-o*, respectively) in the spoken language of the time, although the written language usually still cherished that difference. Thus, even spelling indirectly reflects the scribes' knowledge of standard Latin grammar. Since spelling was bound to reflect the drastic linguistic changes, correct spelling was one of the main concerns of school instruction from late antiquity onwards. Everything suggests that classical spelling was considered an integral part of a good competence in Latin.

4 Examined features

In Section 5, the association between spelling variation and a selection of linguistic features will be studied. Those features were selected so that they are indicative of language change and possibly represent different areas of language. In this Section, I present this selection: nine linguistic features that are either conservative or innovative, plus the date. The features are listed below in a tentative order of increasing grammatical complexity: first, lexicon, then, morphology and, lastly, syntax. The complexity of these grammatical domains will be discussed in Section 6.2. Five of the features that will be examined here were already discussed in my Chapter about formulaicity in documentary Latin (Korkiakangas 2018), albeit with different emphasis and operationalization. In the following, I explain the linguistic motivation for each feature and how the feature is operationalised as a variable.

Date. The date variable is included as a reminder that, although the LLCT is analysed here as an essentially synchronic data set, diachronic variation is also present, as I show in Korkiakangas (2017), in which I discuss the historical reasons for this. The continuous variable indicates the middle of each scribe's period of activity based on the LLCT evidence. If only one document survives from a scribe, the scribe's activity period consists of that year. If two or more documents were written by the same scribe in different years, the scribe's position on the timeline is the mean of the first and last dates. Only a few scribes wrote documents over several decades.

Innovative lexicon. The percentage of non-classical lemmas in the total number of words is expected to measure how readily the scribe utilised innovative lexicon. Seventy-nine lemmas were found that are indisputably non-classical: *aldia*, *aldiaricus*, *aldio*, *aldionalis*, *arimannus*, *arra*, *banda*, *barba*, *barbanus*, *batto*, *bluto*, *bullitanus*, *calderarius*, *cambiator*, *cambium*, *caminata*, *canavarius*, *cavallarius*, *cavallicatura*, *cergiolitum*, *caesa*, *concambiatio*, *concambio*, *concambium*, *debluto*, *fiuwadia*, *focacia*, *fossata*, *fossatum*, *fumarius*, *gahagium*, *gasindus*, *gastaldus*, *grunda*, *launechild*, *mallo*, *marepas*, *marscalcus*, *monto*, *morgingabum*, *mustariolum*, *ornile*, *patrinius*, *paupertacula*, *petia*, *petiola*, *petiolum*, *petium*, *rasula*, *scafilus*, *scherpha*, *scufia*, *sculdahis*, *spanga*, *sporus*, *staffilus*, *stanclio*, *stantarium*, *strata*, *summarra*, *sundrialis*, *sundrium*, *tessero*, *tia*, *tingatio*, *torta*, *ubiscarius*, *usitile*, *wadia*, *wadio*, *waldemannus*, *waldus*, *vassus*, *weregeldum*, *viccia*, *viganatio*, *viganio*, *vitellata*, *zapa*. The chronology of the lemmas was checked using several dictionaries. Lexicon is, of course, closely related to the propositional contents of text; therefore, the percentage of the Innovative lexicon is likely to vary from document to document depending on the matter discussed. Due to the skewness of its distribution, the variable was not encoded as continuous. Instead, I divided it in three categories that are roughly equal in size (no innovative lemmas, 0.02% to 0.40% in all words and 0.41% to 3.75% in all words). The variable varies from 0% to 3.75%.

Genitive plural form. The genitive plural form (*-orum*/*arum*/*um*) is likely to have decreased in use at a relatively early stage, unlike the genitive singular morphology, as is witnessed by its hypercorrect and crystallised attestations in late nonstandard texts. The form is assumed to have been utilised as a prestige form (Sornicola 2012: 59; Politzer and Politzer 1953: 26, 28). A prestige form is a variant to which language users assign social prestige (Sairio and Palander-Collin 2012). Within the LLCT, prestige variants tend to be classical Latin forms and constructions. The continuous variable indicates the proportion of genitive plural forms in the total number of words in the LLCT (range 0% to 2.4%).

Neuter in *-um* instead of *-o*/*-u*/*-us*. The neuter gender had mainly merged with the masculine by the late Latin period (Adams 2013, 383ff.). However, some scribes had memorised a greater or lesser set of neuters, apparently due to instruction in school. Therefore, the degree to which the *o*-stem neuters in the LLCT maintained their conservative nominative/accusative singular *-um* instead of *-o, -u* or *-us* can be used as a hint about a scribe's learned language skills (Korkiakangas 2010: 139; Löfstedt 1961: 226). The continuous variable represents the share of the *-um* ending of *o*-stem neuters in all singular *o*-stem neuters in the LLCT and varies from 0% to 100%.

***De* PP instead of the genitive**. By the early Middle Ages, the prepositional phrase with *de* had gradually replaced the genitive case as the main marker of adnominal possession (Valentini 2018; Fiorentino 1994: 45–6; Adams 2013: 267–74). The genitive morphology still survived, at least in the singular, where it probably merged with the dative singular morphology (Sornicola 2012: 73–4; de Dardel and Gaeng 1992: 99, 103–4; Zamboni 2000: 110). The continuous variable measures the share of the *de* PP in all cases of adnominal possession in the LLCT and varies from 0% to 66.7%.

Complementiser clauses instead of accusative and infinitive construction (ACI). Classical Latin encoded complement clauses of semantic classes, such as those reporting indirect speech, with ACI. The competing complementation strategy, i.e. finite clauses headed by complementisers, such as *ut* or *quod* 'that', became increasingly frequent beginning from the later Empire (Greco 2017; Zamboni 2000: 119–20; Ledgeway 2012: 244ff.). ACI seems to have been considered a conservative prestige variant and continued to be far more common (86%) in the LLCT than is the innovative complementiser clause (14%). The dichotomous variable indicates whether a scribe utilised complementiser clauses in which ACI could have been utilised or not.

Subjects/direct objects in *-bus*. In the LLCT, the original dative/ablative plural ending *-bus* appears to have been used in several hypercorrect ways with a restricted set of lemmas. For example, the dative/ablative *heredibus* acts as the subject of a finite verb in the example (1) (Fiorentino 1994: 39; Sornicola 2012: 57–8; Politzer and Politzer 1953: 28–9).

(1) *ChLA* 74.26 (817 CE)
nos et nostris heredibus tibi et successoribus tuis exinde iustitia per singulos annos reddere debeamus
we and our heirs must pay you and your successors the lease thereof every year

The origin of these hypercorrect uses as subject and direct object arguments in the LLCT is likely to be found in formulaic prepositional phrases of the type *cum nostris heredibus* 'with our heirs' (Korkiakangas 2016: 48–52). The dichotomous variable indicates whether a scribe utilised *-bus* ending in the subject or direct object function or not.

Accusative subjects. The morphosyntactic alignment of late Latin underwent a partial reorganization, which finally led to the disappearance of the case system as the marker of subject and object arguments in most Romance vernaculars. The accusative, originally the marker of the direct object, is likely to have extended first to those subjects of intransitive finite verbs that underwent verbal action, i.e. were Patients by their semantic role. These, like all subjects, had been originally marked with the nominative. The accusative was later extended to most other subjects as well, thus abolishing the case opposition. In fact, the vast majority of Romance nouns are derived etymologically from the Latin accusative forms (Rovai 2012: 103ff.; Cennamo 2009: 315ff.; Ledgeway 2012: 328ff.). LLCT is practically the only Latin corpus with copious occurrences of the accusative in the subject function (Korkiakangas 2016). In the early medieval Latin of Italy, several word-final sounds had weakened, which led to ambiguity regarding the nominative and accusative inflections in almost all the declensional classes. Therefore, only those non-pronominal subjects in which the nominative and accusative forms remained clearly distinguishable due to the different number of syllables are included in the analysis (e.g. *vir-tus*.NOM as opposed to *vir-tu-te*(*m*).ACC 'virtue, church'; Korkiakangas 2016: 111). The continuous variable measures the proportion of accusative subjects in all the subjects of finite verbs in the LLCT and varies from 0% to 100%.

OV word order. The transformation of the pragmatically conditioned word order in classical Latin to the syntactically conditioned one of the Romance languages is known to have been slow. The most typical verb/direct object order in classical Latin was OV, while VO dominates in Romance languages, albeit with considerable variation (Spevak 2010; Ledgeway 2012: 225–35). The transformation was likely to have been in progress at the time of the LLCT, with OV still being the most frequent order (64.7%). The OV order is included in the present analysis because it has sometimes been argued that it was considered a conservative prestige feature from a very early stage (see the references in Ledgeway 2012: 229). I decided to examine only the most prototypical and unambiguous verb/object combinations with one non-coordinated finite verb and non-pronominal direct object (Korkiakangas 2016: 196; Korkiakangas 2018: 437). The continuous variable measures the

proportion of OV-ordered clauses in all the LLCT clauses described above and varies from 0% to 100%.

Past participle/auxiliary verb order. The preferred classical order for the auxiliary verb *sum* 'to be' and the past participle in the perfective forms of passive and deponent verbs is past participle/auxiliary. This classical order shows a decline in the history of Latin in favour of the auxiliary/past participle order, which is typical of modern Romance languages (Bauer 2006: 294; Danckaert 2016; Ledgeway 2012: 223–4). The conservative past participle/auxiliary order is still the prevailing order in the LLCT (90.0% of occurrences), but the innovative order is seen at least once for 36% of the scribes. The dichotomous variable indicates whether a scribe utilised the past participle/auxiliary order in all past participle/auxiliary combinations or not.

Five of the features above are encoded as continuous variables (Date, Genitive plural form, Neuter in *-um*, *De* PP, OV word order) and five as categorical variables (Innovative lexicon, Complementiser clauses, Subjects/objects in *-bus*, Accusative subjects, Past participle/auxiliary verb order), the latter four of which are dichotomous. The features that are here encoded as categorical variables with two or three categories would have particularly skewed distributions (over 30% of occurrences with the same value, which is either 0% or 100%) if they were treated as continuous variables.

5 Statistical analysis

The statistical association of correct spelling and the features presented in the previous Section is examined below using the general linear model procedure (Carey 2013: 129ff.). The general linear model was chosen because it can measure the effects of both continuous and categorical independent (explaining) variables on the means of various groupings of a continuous dependent variable (the variable to be explained); here, the dependent variable is the correctness of the spelling. It is customary to first present the bivariate analysis of the associations between the dependent variable and each independent variable individually. The F-statistic indicates the probability that the group means are equal (Carey 2013: 115–16). The higher the value of the F-statistic, the more likely the association in question is statistically significant. The core of the statistical analysis is the full model, which analyses the effect of all the independent variables on the dependent variable. The coefficient β indicates how the value of the dependent variable changes when an independent variable varies, while the other independent variables remain

fixed. A positive β predicts an increase in the dependent variable, and a negative β predicts a decrease. The greater the absolute value of the β, the stronger the prediction. N is the number of occurrences (here scribes) and s.e. is the standard error. Note that the β coefficients of categorical variables are compared to a conventional baseline category marked 'a'. With continuous variables, the change in β is compared to a one-unit change on the variable's scale (percentages).

Table 1. The effect of ten variables (see Section 4) on spelling correctness.

Independent variables		**Bivariate analysis**		**Full model**	
	N	β	s.e.	β	s.e.
Date (F)		59.38***			
year, scribe's average	176	0.0003***	$3.8*10^{-5}$	-0.0003***	$3.6*10^{-5}$
Innovative lexicon (F)		3.97*			
0%	57	-0.02 (n.s.)	0.004	0.001 (n.s.)	0.003
0.02% to 0.40% in all words	60	0.009*	0.004	0.007*	0.003
0.41% to 3.75% in all words	59	a		a	
Genitive plural form (F)		7.72**			
% proportion of gen. pl. in all words	176	0.873**	0.314	0.944***	0.266
Neuter in *-um* instead of *-o*/*-u*/*-us* (F)		6.62*			
% share of *-um*	159	0.013*	0.005	0.012**	0.004
***De* PPs instead of genitive (F)**		11.38***			
% share of *de* PP	176	-0.042	0.012	-0.051***	0.01
Compl. clause instead of ACI (F)		16.44***			
no	88	-0.013***	0.003	-0.007**	0.003
yes	88	a		a	

Subjects/objects in *-bus* (F)		9.31**			
no	106	-0.010**	0.003	-0.001 (n.s.)	0.003
yes	70	a		a	
Accusative subjects (F)		0.344 (n.s.)			
no	129	0.002 (n.s.)	0.004	0.002 (n.s.)	0.003
yes	47	a		a	
OV word order (F)		2.26 (n.s.)			
% share of OV order	173	0.009 (n.s.)	0.006	0.007 (n.s.)	0.005
Past participle/auxiliary verb order (F)		0.432 (n.s.)			
not all PPart/auxiliary	64	0.002 (n.s.)	0.003	-0.001 (n.s.)	0.003
all PPart/auxiliary	112	a		a	
Adjusted R^2 (full model)				0.458	
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, n.s. = not significant, s.e. = standard error, a = baseline category					

All associations that are statistically significant (see asterisks) in the bivariate analysis, except for the Subjects/objects in *-bus*, are also statistically significant in the full model. Accusative subjects, OV word order and Past participle/auxiliary verb order did not show statistically significant associations in either analysis. The following list summarises the interpretations of the statistically significant associations in the full model:

1) The more advanced the **Date** within the time frame under examination, the smaller the degree of correct spelling. The decreasing effect is, however, slight, and the bivariate analysis suggests an opposite trend: spelling correctness increases slightly by time if the effect of the other independent variables is not taken into account. The fact that this kind of diachronic change is involved is a reminder

that all the variations present in the LLCT data are by no means explained by synchronic linguistic features alone.

2) The category of 0.02% to 0.40% of **Innovative lexicon** is associated with greater spelling correctness than is the category of 0.41% to 3.75%. Even though the small coefficient β of the category 'no Innovative lexicon' is positive, it is not statistically significant. If that coefficient were statistically significant, it could be stated that the fewer innovative lemmas a scribe used, the higher his spelling correctness tended to be. With regard to the category 0.02% to 0.40%, the pattern is as expected, whereas it remains unclear why the complete absence of (the seventy-nine) innovative lemmas does not lead to a statistically significant association with spelling correctness when compared to the baseline category. This may be related to document types, as innovative lemmas are more frequent in certain types of documents than they are in others. It might be beneficial to include the document type as a variable in a future study.

3) The greater the number of **Genitive plural forms**, the greater the level of correct spelling. The pattern is as expected: the genitive plural was likely to have been considered a conservative prestige form associated with high levels of correct spelling (the largest coefficient).

4) The greater the number of **Neuters in *-um* instead of *-o*/*-u*/*-us***, the higher the level of correct spelling. The pattern is as expected: the *-um* ending of the neuter was likely to have been considered a conservative prestige form associated with high levels of correct spelling.

5) The greater the number of ***De* PPs instead of genitives**, the lower the level of correct spelling. The pattern is as expected: the *de* PP was likely to have been considered an innovative low-prestige construction when compared to the genitive case. Consequently, *de* PPs predict low levels of correct spelling.

6) The greater the number of **Complementiser clauses instead of ACI**, the lower the level of correct spelling. The pattern is as expected: ACI was likely to have been considered a conservative prestige construction when compared to the complementiser clause. Consequently, complementiser clauses seem to predict low levels of correct spelling.

The first conclusion that can be drawn from the full model is that the statistical analysis partly confirms the working hypothesis presented in Section 1: good spelling is actually associated to some features that I defined as indications of the scribes' L2 competence in Section 4. This is shown by the meaningful and statistically significant associations between correct spelling and five of the linguistic variables that represent the features the scribes had to learn to use or to avoid in order to write in standard Latin. On the other hand, correct spelling

showed no statistically significant association with the other four features. The next step is to find an answer to the following questions: Do the statistically significant associations have something in common? Equally importantly, do the statistically non-significant associations have something in common?

A preliminary interpretation is that the features seem to be grouped according to their grammatical domain: standard Latin spelling appears to be associated in a meaningful and statistically significant way mainly with lexical (Innovative lexicon) and predominantly morphology-related features (Genitive plural, Neuter *-um*, *De* PP and Complementiser clauses), whereas predominantly syntactic features (Subjects/objects in *-bus*, Accusative subjects, OV order and Past participle/auxiliary order) have no statistically significant association with correct spelling. At this point, it must be emphasised that Latin uses inflections to mark the relationship between the word-level grammatical structure and the sentence-level structure. Therefore, it is sometimes difficult to draw a line between morphology and syntax (see the discussion on *-bus* in Section 6.2). In this study, 'morphological features' or 'morphology-related features' refer to those features that contain a functional morpheme but are not related to syntactic functions.

It appears that L2 syntax was not acquired as successfully as were the less complex L2 grammar domains of lexicon and morphology. Consequently, even good spellers used classical Latin syntactic features haphazardly. In the following Sections, I seek to show that the way in which scribal education dealt with language (Section 6.1) and the complexity of the grammatical domain (Section 6.2) are likely to have determined why the non-syntactic features were attributed a prestige status and thus appear to be associated with spelling correctness, which is considered to be an indicator of language competence in the present study.

6 Discussion

6.1 Scribal education

In Section 4, I proposed in passing that instruction in schools may have kept the neuter category alive artificially by reminding students of its existence and by making pupils memorise lists of neuters. Donatus' *Ars minor*, the basic pedagogical grammar of the time, defined neuter nouns by the sentence quoted in excerpt (2). Therefore, at least some future scribes may have learnt to consider the neuter *-um* ending to be a conservative feature. I (Korkiakangas 2010: 139) showed for the papyri of Ravenna (legal documents mainly from the sixth and

seventh centuries) and Löfstedt (1961: 226) for the juridical Latin of the *Edictum Rothari* (seventh century) that the neuter *-um* was largely maintained and more often than was the accusative *-um* of the masculine. This, and the fact that *-um* was also used hypercorrectly for the dative/ablative *-o* with neuters but not as often with masculines, supports my hypothesis.

(2) Donatus, *Ars minor* (*GL* 4.356.10–11)
scamnum nomen appellatiuum generis neutri numeri singularis figurae simplicis, casus nominatiui accusatiui et uocatiui
scamnum [bench] is a neuter-gender singular-number non-compound common name, the form of the nominative, accusative and vocative case

Relatively little is known about early medieval Tuscan scribes' education (Mencacci 2012: 75–90; Petrucci 1973: 639; Schiaparelli 1924: 56ff.). An LLCT document from 767 CE (*ChLA* 997) mentions a school in the portico of the cathedral of Lucca. This school, which was obviously run by the church, seems to have provided clerics and some laymen with education in writing Latin for the purposes of ecclesiastical administration and/or documentary production (Coturri 1973: 695ff.). Initially, most documentary scribes were clerics. In the ninth century, when documentary production passed into lay hands (Keller 1973: 120–24), it is likely that the education of the scribes changed accordingly. This administrative reform seems to be reflected in the increasing uniformity of the documents, even in terms of improved spelling (Korkiakangas 2017). Schiaparelli (1924: 57) identified a few documentary scribes whose texts are included in the LLCT documents among the copyists who compiled the famous Manuscript 490, which is the only surviving book copied with certainty in Lucca during the eighth century. Thus, at least some Luccan scribes also knew book scripts, arguably a sign of higher literary education.

What the actual school curriculum in early medieval Tuscany included is unknown but, considering the tremendous linguistic variety in the LLCT documents, not much was required from all those who wrote documents. The grammatical treatises of the time can be used to approximate the maximum level that a researcher can expect from scribal education in early medieval Tuscan schools. The treatises that were available before the Carolingians began to promote Priscian's *Institutiones grammaticae* in the late eighth century only dealt with spelling, the parts of speech and the figures of speech (Black 2001: 64; Luhtala 1993: 145–6). The available grammatical treatises had been written in late antiquity when the distance between the syntax of written and spoken varieties of Latin was still

relatively small. Therefore, the treatises were concerned with syntactic issues only in passing, for example, when they discussed rhetorical figures.

Coturri (1973: 696) maintained that, in addition to the above-mentioned school, Lucca also had an elementary writing school, which would explain the high number of literate, but apparently not very experienced, people in Lucca, as evidenced by the large number of autograph subscriptions in the LLCT documents (cf. Petrucci 1973: 640). Technically, any higher literary education, such as the adoption of documentary formulae, must have been based on immersion in extant authoritative documents and on their subsequent imitation, either independently or in the classroom. This method had been in use since ancient times and seems to have worked without many difficulties in the culture of early medieval Europe (Black 2001: 64–70, cf. Cribiore 2001: 132–4). This was possible because the gap between the phonology and grammar of the spoken language and of classical Latin was still narrow enough not to compromise communication. Apparently, a superficial memorization of the classical spellings of frequent words and some understanding of parts of speech and their inflections occasionally passed for sufficient Latin literacy. Similarly, the imitation of extant texts was evidently sufficient to prevent the scribes from lapsing too often into non-classical syntactic choices, as shown in the relatively high percentages of the conservative variants of the syntactic variables in Section 4.

Although the learning method based on memorization and imitation still worked in early medieval Tuscany, it is likely that Latin L2 acquisition became increasingly challenging over the centuries, which widened the gap between the written standard Latin and the vernacular. By the early Middle Ages, an Italian Latin learner would have needed a basic understanding of syntactic functions to be able to cope with classical Latin word order and subject case marking, but school instruction did not foster such an understanding. The recognition of this syntactic gap may have been a stimulus that interested Carolingian scholars in Priscian's work at the end of the eighth century. Luhtala (1993: 161ff.) showed that the Carolingian grammarians elaborated on Priscian's (infant) syntactic analysis of subject and predicate. Black (2001), however, maintained that this scholarly interest had little or no effect on the everyday pedagogical grammar of Latin before the twelfth century, when it became necessary to adapt Latin teaching practices to the needs of larger audiences and when the advances in logic introduced the Aristotelian notions of subject and predicate into real grammatical praxis (Black 2001: 70).

Thus, the observations made in the statistical analysis in Section 5 confirm the picture that arises from the close reading of ancient and early medieval

grammatical treatises: the pedagogical grammar was concerned with spelling and morphology-related issues. There had not been a need for syntactic explanations in the centuries during which the grammatical tradition took shape and, when the need arose, it was not recognised for centuries. In the meanwhile, schoolroom instruction was limited to easily understandable features, depending on the teacher's skills.

On the other hand, it would be naïve to claim that the early medieval Latin educational system transmitted theoretical grammatical knowledge in such an efficient and uniform manner that it determined the scribes' writing performances in the way they are seen in the statistical analysis above. Studies of modern L2 acquisition show that, despite effective and theoretically grounded school instruction, acquisition is dependent on several other factors outside of the curriculum content, including students' motivation, learning skills, exposure time, the distance between the L1 and the L2 and, importantly, the properties of the linguistic features to be learnt (e.g. Lightbown and Spada 2006: 53ff., 82ff.; Matras 2009: 69–70; Jarvis and Odlin 2000). In fact, learning depends ultimately on the learner's personal *intake*, not the *input* provided, however adequate the latter may be. As stated above, early medieval L2 learners were supposed to immerse themselves in extant texts and to imitate them in order to reach the higher levels of literacy; in other words, recognised syntax and style.

It is also apparent that much of what was taught of morphology was left to be internalised by this immersion. No basic grammar guided L2 Latin learners to use the genitive case instead of the *de* PP in certain semantic contexts, or not to use a complementiser clause if one could use an ACI; that is, mainly with speech verbs. Grammatical treatises and, assumedly, teachers made pupils memorise lists of prepositions and complementisers (and probably neuters), but the understanding of the contexts in which these were to be used had to be adopted largely by immersion in texts. The following Section discusses how the perceptual properties of the linguistic features to be learnt may have affected the learning processes and, consequently, the patterning of the features examined in the present study.

6.2 Grammatical domains

As observed in Section 5, the statistically significant and non-significant associations of the full model seem to be grouped according to each feature's grammatical domain. I suggest here that this split between syntax and other grammatical domains is closely related not only to medieval scribes' education,

but also to the linguistic properties of the features themselves. Thus, teaching and linguistic properties are not to be considered mutually exclusive, but complementary, explanations: what was taught was likely to reflect the relative order of the complexity of linguistic domains, albeit unintentionally (see Table 2). Similarly, prescriptive instruction in school is bound to interfere with the linguistic motivation of the pattern observed in the statistical analysis.

Cognitive linguistics offers a useful framework for systematizing the split mentioned above. Cognitive theories do not assume clear-cut boundaries between what are traditionally called lexicon, morphology and syntax, but consider them to form a continuum (Broccias 2012). According to Croft and Cruse (2004: 247ff.), language is a repository of constructions of varying degrees of generality. This is illustrated by the syntax-lexicon continuum, a scale that locates language domains on a continuum according to their atomicity/complexity and specificity/schematicity (Table 2). Atomicity refers to the indivisibility into meaningful parts and is the opposite of complexity, while schematicity means the generality of category and is the opposite of specificity.

Table 2. Syntax-lexicon continuum (after Broccias 2012: 738 and Croft and Cruse 2004: 255).

Grammar domain	**Traditional name**	**Example**
Complex and (mostly) schematic	Syntax	noun verb noun (= transitive construction)
Complex and (mostly) specific	Idiom	*pull one's leg*
Complex but bound	Morphology	noun-*s*
Atomic and schematic	Word class	pronoun, adjective
Atomic and specific	Word/lexicon	*this*, *green*

The linguistic features were presented in Section 4 and examined in Section 5 in a tentative order of increasing grammatical complexity in terms of the syntax-lexicon continuum. The features belong to the domains Word/lexicon, Morphology and Syntax. As seen previously, the statistically significant associations are related to the domains Word/lexicon and Morphology, and the statistically non-significant associations to the complex and schematic domain of Syntax.

I argue elsewhere (Korkiakangas 2018: 433–4) that the order of the domains in the syntax-lexicon continuum reflects, among other things, the cognitive effort involved in recognizing those domains during L2 acquisition. The more complex domains (such as Syntax) are not recognised and, consequently, are not adopted as easily as are the less complex domains (such as Word/lexicon). However,

the evidence of the statistical analysis, as well as that gathered in Korkiakangas (2018), suggests that the complexity, as understood in terms of the syntax-lexicon continuum, is caused by various underlying motivations that can be identified and explained. This view implies that the complexity hierarchy inherent in the syntax-lexicon continuum should possibly be considered a useful but epiphenomenal simplification. In the following passages, I discuss one underlying motivation that is likely to be relevant for L2 acquisition of documentary Latin. This motivation is salience.

The complexity dimension of the syntax-lexicon continuum seems to be partly related to the perceptual properties of each domain's typical representatives (auditory properties in the oral context and visual properties in the written context). Lexical items and longer morphological inflections tend to be easily perceptible pieces of phonetic/graphic substance that a language learner recognises and adopts simply by memorization. This prominence or noticeability of linguistic units, such as words and (other) morphemes, is termed 'perceptual salience' (Cintrón-Valentín and Ellis 2016: 3; MacLeod 2015). Note that several other motivations, such as token frequency (e.g. Bybee 2007), cause linguistic units to be salient. However, the present research setting only applies to perceptual salience.

Contrary to lexicon and morphology, syntax pertains to the rules that operate behind the material façade of words and morphemes. The rules do not often have manifestations that are noticed easily, or the rationale for the manifestations is not easily deduced from the use contexts without specific linguistic intuition or without the help of an experienced teacher. A learner of Latin syntax must be capable of mapping information conveyed by the morphological inflections of words with the functions those words carry at the sentence level. This is much more difficult than is adopting individual non-contextualised words or morphemes. Consequently, syntactic features are at risk of passing unnoticed in superficial language acquisition. For example, an L2 learner of English learns lexical items and free functional morphemes, such as prepositions, quickly, but has to invest considerably more cognitive effort in learning to put *-s* at the end of those present tense verbs that are in the third person singular (cf. Zobl and Liceras 1994: 169–71).

The interpretation presented above is supported by several experimental studies on the L2 acquisition order of grammatical categories in modern languages, such as Zobl and Liceras (1994) and Goldschneider and DeKeyser (2001). Processability-informed theories assume that learners first acquire the relationships between lemmas, then those within words (lexical morphology),

within phrases, between phrases and, finally, between clauses (Pienemann 1999: 7–9). The studies by Goldschneider and DeKeyser (2001) and Cintrón-Valentín and Ellis (2016) confirm that perceptual salience plays an important role in L2 morpheme acquisition.

Free morphemes, such as words, the preposition *de* and complementisers, as well as the phonetically/graphically substantial genitive plural ending (e.g. *-orum*), are readily considered perceptually salient. It is also easy to agree that word order issues (OV order and Past participle/auxiliary verb order) are not determined by perceptually salient units, but by the linearization of syntactic functions. Even though Subject case encoding (Accusative subjects) does rely on case inflections, nominative and accusative, and the morphological difference between the nominative and accusative inflections of type *virtus – virtute*(*m*) may be perceptual as such, it is hard to decide which form is the salient one (perhaps *virtus*; cf. Korkiakangas 2018: 444).

There is, however, no need to speculate on the salience of the nominative and accusative forms, given that OV order, Past participle/auxiliary verb order and Subject case encoding are all based on syntactic rules, and syntactic rules require an understanding of abstract categories, such as syntactic function, which cannot be tackled in terms of salience. In order to assign the nominative case classically to all the subjects of finite verbs contrary to the late Latin morphosyntactic alignment, a late Latin speaker had to understand what the subject was. The knowledge of syntactic functions was also necessary in order to avoid the ablative inflection *-bus* with subjects and objects. The relatively salient ending *-bus* was possibly erroneously considered to be a prestige feature and was misused hypercorrectly in any syntactic context, even by good spellers.

However, not all the features examined in the statistical analysis seem to fit in with this explanation. The neuters in *-um* instead of *-o*/*-u*/*-us* can hardly be described as perceptually salient. The pronunciation of the ending is likely to have been /*o*/ while, in writing, <*um*> does not stand out due to its brevity, at least in comparison to the genitive plural ending, *-orum,* as four characters are considerably more than are two characters. Moreover, the neuter is not the only habitat of *-um*, which is also the classical accusative singular form of the *o*-stem masculine as well as the genitive plural form of the third declension. Thus, the perceptual properties of the *-um* ending are unlikely to have guaranteed it the position of being recognised as a prestige feature and the ensuing association with correct spelling. Nonetheless, the *-um* neuters are strongly associated with correct spelling; thus, the prestige attribution must derive from elsewhere, namely from prescriptive school instruction.

To summarise, both scribal education and linguistic properties are likely to have contributed to the diverse associations between correct spelling and the linguistic features observed in Section 5. On one hand, instruction in schools tends to sanction some linguistic features instead of others. On the other hand, an L2 learner naturally tends to pay more attention to easily perceivable, salient, linguistic features. Nor is it an accident that the split arises between syntax and other domains for both criteria. It is only to be expected that the concerns of the grammatical tradition are primarily and originally related to the perceptually most salient, that is, the most easily manageable, domains, whereas syntax receives attention only as a result of the theoretical development of the field.

7 Conclusion

This study examined whether the variations in spelling and the variations in the use of certain linguistic features were associated with each other in early medieval documentary Latin as documented in LLCT. This was accomplished by quantifying spelling variations and selected linguistic features that are known to be indicative of prestige (conservative, classical Latin features) or the lack thereof (innovative, Romance-type features) and by the subsequent analysis of their statistical associations using a general linear model.

The statistical analysis confirmed the working hypothesis that correct spelling and learned grammatical competence are associated. However, statistically significant associations are limited to lexical and morphology-related features. This suggests that L2 syntax was not acquired as successfully as were the less complex L2 grammar domains. I explained this pattern via the complexity of each grammatical domain in question and via the prestige assessments derived from prescriptive school instruction. The pedagogical grammatical tradition available at the time was not concerned with syntax, but concentrated on lexicon, spelling and morphology instead. The correct contexts of use were expected to be learnt by imitating authoritative models, and it is here that the grammatical domains seem to play a role.

The syntax-lexicon continuum, a framework adopted from cognitive linguistics, ranks different grammatical domains according to their complexity, which seems to determine the ease with which L2 learners acquire them. I have suggested, however, that the fundamental reason behind this ranking, as far as L2 acquisition is concerned, might be the fact that the domains tend to be represented differently in perceptual terms: perceptually salient lexical items and

morphological morphemes are easy to notice and acquire, while syntax is much more complex in the sense that it requires an understanding of the relationship between words and their inflections. Consequently, an L2 learner succeeded in associating prestige (such as with the genitive plural form) or the absence thereof (such as with the *de* PP) with perceptual features. For syntactic features, the prestige/no prestige attribution often failed, hence the statistically non-significant associations with correct spelling. The salience-based explanation obtains support from several studies on the L2 acquisition order of grammatical categories in modern languages. To achieve a more detailed picture of the role of salience, a further study of the association between correct spelling and lowly salient morphological features other than *-um* would be required. In fact, Korkiakangas (2018) suggested that morphological features that are particularly non-salient easily escaped the scribes' attention.

Thus, school instruction and linguistic properties are not to be considered mutually exclusive explanations, but are both likely to contribute to the differing associations between spelling and linguistic features in the LLCT. Although the exact relationship between domain complexity and salience cannot be fully established within the present research setting, they both seem to have an important role in how the LLCT scribes acquired their Latin. In summary, spelling correctness can duly be utilised as a rough measure of early medieval scribes' Latin competence.

Online resources

LLCT = Late Latin Charter Treebank (LLCT1, version 2) is available at https://zenodo.org/record/3633607#.YlUMr4VBxPY.

Open Office Latin lexicon is available at https://github.com/cisocrgroup/Resources/tree/master/lexica/latin.

Whitaker's WORDS is available at http://mk270.github.io/whitakers-words/operational.html.

References

Adams, J. N. 2013. *Social Variation and the Latin Language*. Cambridge University Press.

Amsler, Mark E. 1993. 'History of Linguistics, "Standard Latin", and Pedagogy', in V. A. Law (ed.), pp. 49–66.

Bamman, David, Passarotti, Marco, Crane, Gregory and Raynaud, Savina 2007. *Guidelines for the Syntactic Annotation of Latin Treebanks*, v. 1.3. (https://static.perseus.tufts.edu/docs/guidelines.pdf, accessed 12 April 2022)

Bartoli Langeli, Attilio 2006. *Notai: scrivere documenti nell'Italia medievale*. Rome: Viella.

Bauer, Brigitte L. M. 2006. '"Synthetic" vs. "Analytic" in Romance: the Importance of Varieties', in R. Gess and D. Arteaga (eds.), *Historical Romance Linguistics: Retrospective and Perspectives*, pp. 287–304. Amsterdam: John Benjamins.

Black, Robert 2001. *Humanism and Education in Medieval and Renaissance Italy*. Cambridge University Press.

Broccias, Cristiano 2012. 'The Syntax-Lexicon Continuum', in Terttu Nevalainen and Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, pp. 735–47. Oxford University Press.

Burt, Jennifer S. 2006. 'Spelling in Adults: The Combined Influences of Language Skills and Reading Experience', *Journal of Psycholinguistic Research* 35: 447–70.

Bybee, Joan 2007. *Frequency of Use and the Organization of Language*. Oxford University Press.

Carey, Gregory 2013. *Quantitative Methods in Neuroscience*. Online book draft. (http://psych.colorado.edu/~carey/qmin/QMIN_2013_03_17.pdf, accessed 12 April 2022)

Castellani, Arrigo 1976. *I più antichi testi italiani*. Bologna: Pàtron.

Cennamo, Michela 2009. 'Argument Structure and Alignment Variations and Changes in Late Latin', in Jóhanna Barðdal and Shobhana L. Chelliah (eds.), *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case*, pp. 307–46. Amsterdam: John Benjamins.

Cintrón-Valentín, Myrna C. and Ellis, Nick C. 2016. 'Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus', *Frontiers in Psychology* 7: 1284.

Coturri, Enrico. 1973. 'Fonti per uno studio della cultura a Lucca dall'VIII all'XI secolo', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, pp. 695–702. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Cribiore, Raffaella 2001. *Gymnastics of the Mind: Greek Education in Hellenistic and Roman Egypt*. Princeton University Press.

Croft, William and Cruse, D. Alan 2004. *Cognitive Linguistics*. Cambridge University Press.

Danckaert, Lieven 2016. 'Variation and Change in Latin BE-Periphrases: Empirical and Methodological Considerations', in J. N. Adams and Nigel Vincent (eds.), *Early and Late Latin: Continuity or Change?*, pp. 132–62. Cambridge University Press.

de Dardel, Robert and Gaeng, Paul A. 1992. 'La déclinaison nominale du latin non classique: essai d'une méthode de synthèse', *Probus* 4: 91–125.

de Paolis, Paolo 2010. 'L'insegnamento dell'ortografia latina fra Tardoantico e alto Medioevo: teorie e manuali', in Lucio Del Corso and Oronzo Pecere (eds.), *Libri di scuola e pratiche didattiche dall'antichità al rinascimento*, pp. 229–91. Università degli studi di Cassino.

Fiorentino, Giuliana 1994. 'Aspetti della morfosintassi nominale nelle carte cavensi del IX secolo', *Archivio Glottologico Italiano* 79: 23–50.

GL = *Grammatici Latini*, ed. Heinrich Keil, Leipzig: Teubner 1855–80.

Goldschneider, Jennifer M. and DeKeyser, Robert M. 2001. 'Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants', *Language Learning* 51: 1–50.

Greco, Paolo 2017. 'La complementazione frasale', in Rosanna Sornicola, Elisa D'Argenio and Paolo Greco (eds.), *Sistemi, norme, scritture: la lingua delle più antiche carte cavensi*, pp. 267–310. Naples: Giannini.

Herman, József 1991. 'Spoken and Written Latin in the Last Centuries of the Roman Empire: A Contribution to the Linguistic History of the Western Provinces', in Wright (ed.), pp. 29–43.

Herman, József 1992. 'Sur quelques aspects du latin mérovingien: langue écrite et langue parlée', in Maria Iliescu (ed.), *Latin vulgaire – latin tardif III: Actes*

du IIIème Colloque international sur le latin vulgaire et tardif, pp. 173–86. Tübingen: Max Niemeyer.

Jarvis, Scott and Odlin, Terence 2000. 'Morphological type, spatial reference, and language transfer', *Studies in Second Language Acquisition* 22: 535–55.

Keller, Hagen 1973. 'La marca di Tuscia fino all'anno Mille', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, 117–36. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Korkiakangas, Timo 2010. 'Neutro plurale e femminile singolare: il fattore grafico nell'interpretazione delle peculiarità della declinazione tardolatina', *Arctos* 44: 133–45.

Korkiakangas, Timo 2016. *Subject Case in the Latin of Tuscan Charters of the 8th and 9th Centuries*. Helsinki: Societas Scientiarum Fennica.

Korkiakangas, Timo 2017. 'Spelling Variation in Historical Text Corpora: The Case of Early Medieval Documentary Latin', *Digital Scholarship in the Humanities* 33: 575–91.

Korkiakangas, Timo 2018. 'Spoken Latin Behind Written Texts: Formulaicity and Salience in Medieval Documentary Texts', *Diachronica* 35: 429–49.

Korkiakangas, Timo and Lassila, Matti 2013. 'Abbreviations, Fragmentary Words, Formulaic Language: Treebanking Medieval Charter Material', in Francesco Mambrini, Marco Passarotti and Caroline Sporleder (eds.), *Proceedings of the Third Workshop on Annotation of Corpora for Research in the Humanities*, pp. 61–72. Sofia: Bulgarian Academy of Sciences.

Korkiakangas, Timo and Passarotti, Marco 2011. 'Challenges in Annotating Medieval Latin Charters', *Journal of Language Technology and Computational Linguistics* 26: 103–14.

Larson, Pär 2012. 'Le carte alto-medievali come fonte di lingua: qualche esperienza personale', in Rosanna Sornicola and Paolo Greco (eds.), *La lingua dei documenti notarili alto-medievali dell'Italia meridionale: bilancio degli studi e prospettive di ricerca*, pp. 63–73. Naples: Tavolario.

Law, Vivien A. 1982. *The Insular Latin Grammarians*. Woolbridge: The Boydell Press.

Law, Vivien A. (ed.) 1993. *History of Linguistic Thought in the Early Middle Ages*. Amsterdam – Philadelphia: John Benjamins.

Ledgeway, Adam 2012. *From Latin to Romance: Morphosyntactic Typology and Change*. Oxford University Press.

Lightbown, Patsy M. and Spada, Nina 2006. *How Languages are Learned*. 2nd edition. Oxford University Press.

Löfstedt, Bengt 1961. *Studien über die Sprache der langobardischen Gesetze*. Stockholm – Göteborg – Uppsala: Almqvist & Wiksell.

Lüdtke, Helmut 1964. 'Die Entstehung romanischer Schriftsprachen', *Vox Romanica* 23: 3–21.

Luhtala, Anneli 1993. 'Syntax and Dialectic in Carolingian Commentaries on Priscian's *Institutiones grammaticae*', in V. A. Law (ed.), pp. 145–91.

MacLeod, Bethany 2015. 'A Critical Evaluation of Two Approaches to Defining Perceptual Salience', *Ampersand* 2: 83–92.

Matras, Yaron 2009. *Language Contact*. Cambridge University Press.

Mencacci, Paolo 2012. *Vita civile ed ecclesiastica a Lucca nell'alto Medioevo, secoli VI–IX.* Lucca: Pacini Fazzi.

Petrucci, Armando 1973. 'Scrittura e libro nella Tuscia altomedievale (secoli VIII–IX)', in *Atti del V Congresso internazionale di studi sull'alto medioevo*, pp. 627–43. Spoleto: Centro Italiano di Studi sull'Alto Medioevo.

Pienemann, Manfred 1999. *Language Processing and Second Language Development: Processability Theory*. Amsterdam: John Benjamins.

Politzer, Frieda N. and Politzer, Robert L. 1953. *Romance Trends in 7th and 8th Century Latin Documents*. Chapel Hill: The University of North Carolina Press.

Rovai, Francesco 2012. *Sistemi di codifica argomentale: tipologia ed evoluzione*. Pisa: Pacini.

Sabatini, Francesco 1965. 'Esigenze di realismo e dislocazione morfologica in testi preromanzi', *Rivista di Cultura Classica e Medievale* 7: 972–98.

Sairio, Anni and Palander-Collin, Minna 2012. 'The Reconstruction of Prestige Patterns in Language History', in Juan Manuel Hernández-Campoy and Juan Camilo Conde-Silvestre (eds.), *The Handbook of Historical Sociolinguistics*, pp. 626–38. Chichester: Blackwell.

Schiaparelli, Luigi 1924. *Il codice 490 della Biblioteca capitolare di Lucca*. Pompeo Sansaini Editore: Rome.

Schiaparelli, Luigi 1933. 'Note diplomatiche sulle carte longobarde II: Tracce di antichi formulari nelle carte longobarde', *Archivio storico italiano* 19: 3–34.

Sornicola, Rosanna 2012. 'Bilinguismo e diglossia dei territori bizantini e longobardi del Mezzogiorno: le testimonianze dei documenti del IX e X secolo', *Quaderni dell'Accademia Pontaniana* 59: 1–102.

Spevak, Olga 2010. *Constituent Order in Classical Latin Prose*. Amsterdam: John Benjamins.

Valentini, Cecilia 2018. *L'evoluzione della codifica del genitivo dal tipo sintetico al tipo analitico nelle carte del* Codice diplomatico longobardo. Università degli Studi di Firenze.

Walsh, Thomas J. 1991. 'Spelling Lapses in Early Medieval Latin Documents and the Reconstruction of Primitive Romance Phonology', in Wright (ed.), pp. 205–18.

Wright, Roger (ed.) 1991. *Latin and the Romance Languages in the Early Middle Ages*. The Pennsylvania State University Press.

Wright, Roger 1991. 'The Conceptual Distinction Between Latin and Romance: Invention or Evolution', in Wright (ed.), pp. 103–13.

Zamboni, Alberto 2000. *Alle origini dell'italiano: dinamiche e tipologie della transizione dal latino*. Rome: Carocci.

Zobl, Helmut and Liceras, Juana 1994. 'Functional Categories and Acquisition Orders', *Language Learning* 44: 169–80.