

Some (statistical) considerations on the use of inscriptions for the study of “sociolinguistic” variation in Latin: The problem of “frequency”

ALESSANDRO PAPINI^{1,2*} 

¹ Ca' Foscari University of Venice, Italy

² ELTE Hungarian Research Centre for Linguistics, Hungary

Received: January 27, 2025 • Revised manuscript received: July 22, 2025 • Accepted: July 29, 2025

© 2025 The Author(s)



ABSTRACT

This article addresses the issue of defining “frequency” in quantitative/statistical sociolinguistic inquiries of the epigraphic sources. While the subject has thoroughly been discussed in the field of diatopic research, it has received limited scholarly attention in the analysis of diaphasic (and diastratic) variation. To bridge this gap in the literature, the present study will concentrate on the instances of the spelling variation with <E> vs. <AE> in a corpus of 1,650 “Christian” inscriptions from Rome (ca. 300 CE – ca. 600 CE). The chosen phenomenon is particularly well-suited to the problem under discussion, since the monophthongisation of /ae/ was most likely a completed linguistic change by the time of our inscriptions. In the course of the study, the epigraphic documents under scrutiny have been classified into three “sociolinguistic categories” (based on formality), i.e. formal, informal, and diaphasically low texts. The “frequency” of the cases of <E> for <AE> in these texts has been calculated according to two different methods, the “degree of error” and the “relative frequency of faults”. The results of the analysis (implemented by using *Conditional Inference Trees*) suggest that, in line with the most recent findings in the field of Latin (corpus) linguistics, the sociolinguistic study of inscriptions is better implemented by a cross-referenced use of both methods, even more so than in the case of diatopic inquiries.

KEYWORDS

corpus linguistics, Latin inscriptions, sociolinguistic variation, *Conditional Inference Trees*, monophthongisation of /ae/

* Corresponding author. E-mail: alessandro.papini@unive.it, ale.papini1@gmail.com

*There are three kinds of lies: lies,
damned lies, and statistics.*

Quotation attributed to B. Disraeli in
Mark Twain's *Autobiography*

1. FRAMING THE PROBLEM

In the last decade(s), scholars have increasingly recognised the value of inscriptions as a reliable source for the study of linguistic variation in Latin. In addition to their traditional use for the study of diatopic variation, a line of research that has historically utilised these sources since the beginning of the 20th century, the importance of inscriptions for the inquiry into sociolinguistic (especially diaphasic; see below) variation has more recently been acknowledged.¹ This is attributable to several key features of the inscriptional evidence that distinguish it from other kinds of testimonies available for sociolinguistic studies of Latin, including literary texts and metalinguistic evidence in Latin grammarians. To begin with, inscriptions can be defined as “contemporary documents”, i.e. documents that (for the most part) have not been transmitted through manuscript tradition and are consequently more likely to “reflect the language of the period in which they were composed”.² Moreover, the chronological, social, and typological range of these documents can hardly be paralleled by the extant literary records.³

In other words, inscriptions represent a distinct category of the “bad data” typically encountered in historical sociolinguistics, rendering these sources particularly well-suited for examining the complex dynamics that characterised the diverse varieties of the Latin diasystem.⁴ Not only do the “contemporary document” character of inscriptions, and the diversity of the extant epigraphic evidence (in terms of textual typologies), provide a more “direct access” to these varieties than literary texts, but the number of the inscriptional records that are known to date

¹Sociolinguistic variation is used here to refer to diaphasic (and diastratic) variation. This distinction reflects the fact that, albeit diatopic variation is part of sociolinguistic research, the two lines of inquiry have traditionally been investigated independently in Latin studies. This is evidenced, *inter alia*, by the separate treatment of the two topics by J.N. Adams (see Adams 2007 vs. Adams 2013, respectively). On the use of inscriptions for the study of diatopic variation in Latin, cf. the detailed accounts of Herman (2000b), Adams (2007, 624–683), and Adamik (2012). See also the discussion by Papini (2022b, 343–351) and references therein.

²Cuzzolin & Haverling (2009, 22). For this reason, Herman (2000a, 18) classifies inscriptions as “direct sources” for the study of variation in Latin. The statement that inscriptions are not transmitted through manuscript tradition “for the most part” reflects the fact that some epigraphic documents have been lost to date, so that information about their text is entirely dependent on later manuscript publications (see Cooley 2012, 360–370).

³Clackson (2011a, 30) and Marotta (2015, 48).

⁴Mancini (2019, 27). This is especially true for the ancient world, for, from the middle ages onwards, “disponiamo già di documenti più cospicui [...], che consentono di effettuare [...] una sorta di ‘intervista’ degli scriventi [we already have a larger number of documents that enable us to conduct a sort of ‘interview’ with the writers]” (Mancini 2014, 32). The well-known statement that historical sociolinguistics “can be [...] thought as the art of making best use of bad data” is found in Labov (1994, 11) and reflects the circumstance that sociolinguistic studies of the past can only rely on sources that are often incomplete and of uncertain interpretation. On this point, cf. among others, Winter (1999, 71–73), Janda & Joseph (2003, 14), Nevalainen & Raumolin-Brunberg (2012, 25), and Mancini (2014, 33–34).

also allows for more comprehensive studies at a quantitative and statistical level.⁵ As a consequence, contemporary scholarly debate is less concerned with the question of whether inscriptions are a valid source for investigating sociolinguistic variation in Latin, a fact that can hardly be denied, and more concerned with the correct methodology to follow in order to investigate such variation.⁶ In this regard, general consensus exists on two fundamental considerations. First, sociolinguistic studies based on the epigraphic material should ideally include a sufficient number of linguistic inscriptional forms and texts to allow statistically, or, at least, quantitatively significant observations to be made (provided that each form is only considered after careful qualitative evaluation). This follows from the intuitive notion that, in (historical) sociolinguistics (as in any other field of study), “più dati uguale maggiore finezza di analisi [more data means greater analytical precision]”.⁷

Second, and most importantly, these studies should always prioritise the analysis of diaphasic, as opposed to diastratic, variation. Considering the context of the epigraphic texts, as identified by their typology and communicative purpose, usually allows each inscription to be placed with satisfactory precision on the diaphasic axis. This is because, it can safely be assumed that texts of different typologies and purposes will represent varieties of the language that are marked differently with respect to the feature [$\pm$ formal] (that is, official writs will be expected to reflect more formal varieties of Latin than spontaneous writings such as graffiti or curse tablets). The same is not true of diastratic variation, for the identity of the writer(s) who produced most inscriptional records is usually impossible to establish, a circumstance stemming from the fact that the realisation of the extant epigraphic records could involve different personalities (the customer, the engraver, the so-called *ordinator*, etc.).⁸

Based on this reasoning, Mancini (2014, 38–39) posits that quantitative sociolinguistic studies of inscriptions should commence with the calculation of the frequency of the investigated phenomena, that is, epigraphic deviations from the “Classical Latin” (CL) norm, in a

⁵To provide a more precise order of magnitude, the EDCS database (the largest epigraphic corpus available online) has published to date more than 500,000 Latin inscriptional texts (cf. https://db.edcs.eu/epigr/epi.php?s_sprache=en). It can also be added that even those literary works that depict speech varieties usually considered substandard (such as Petronius’ *Satyricon*) are essentially artistic creations, meaning that they do not share the “contemporary document” character of inscriptions. Cf. Poccetti, Poli & Santini (1999, 25–26) and Mancini (2014, 33).

⁶The importance of inscriptions for the study of sociolinguistic variation in Latin is well clarified by (among others) Mancini (2014, 33–35), Marotta (2015, 42–44, 2016, 147–150), and Donati (2015, 21–26). See also the discussion and references in Papini (2024, 90–101; on which the present introduction is based). A positive judgement of inscriptions for the inquiry of sociolinguistic variation in Latin is also shared by Adams (2013, 33–34), although this scholar posits that linguistic data in the epigraphic material should in principle be complemented by “additional evidence, such as remarks by grammarians or other speakers”.

⁷Mancini (2014, 33). On this point, see especially Marotta (2015, 47) and Donati (2015, 23). The significance of statistical (as opposed to merely quantitative) analysis for the study of linguistic variation by using inscriptions is discussed in Papini (2022b). A partial exception to this rule are “coherent corpora”, that is, bodies of texts “about which we know something [...] and which belong together in one or more senses”. In such corpora, even a single linguistic form can be significant “because it has a context” (Adams 2007, 633–634). Nevertheless, it must be noted that “coherent corpora” are not a sufficiently comprehensive source for the study of Latin variation (whether diatopic or sociolinguistic). See Papini (2022b, 346–347, 2024, 97–99).

⁸Mancini (2014, 38). See also the discussion by Donati (2015, 29). For the linguistic problems connected with the production of the epigraphic records, see especially Adams (2003, 84). On the notion of context, see Clarkson (2011a, 30) and Mancini (2002, 40–42).

corpus (or, various subcorpora) of texts that share similar local and chronological coordinates, whilst diverging on the diaphasic (as opposed to the diastratic) dimension. This calculation allows to formulate sociolinguistic judgements on the deviations under study: if certain deviations are considerably less frequent in [+formal] texts than in [–formal] documents, (only then) can it be hypothesised that the underlying linguistic changes were also features of socio-linguistically-lower registers of the language and, possibly, lower-class speakers.

Mancini's formulation addresses one of the primary concerns of sociolinguistic studies of historical documents (particularly inscriptions), namely, the fact that, as just mentioned, in the case of the epigraphic sources, the diaphasic and diastratic dimensions “non sono [...] accessibili alla stessa maniera [are not equally accessible]”.⁹ Nevertheless, contemporary research in this field has not yet satisfactorily clarified how quantitative/statistical accounts of historical sociolinguistics (based on inscriptions) should define frequency (a question that, on the other hand, has thoroughly been investigated in the field of diatopic research).¹⁰ Leaving aside outdated (albeit still utilised) methods, such as the calculation of the “absolute figures”, i.e. “the mere counting of instances of an error without reference” to any “other comparandum”,¹¹ the approach most commonly used in sociolinguistic studies to define the frequency of one (or more) error(s) in an epigraphic corpus is to calculate the percentage value of its occurrences with respect to the corresponding correct spellings (“degree of error”, as defined by Adams 2007, 6).

This approach has traditionally been considered to have limited applicability to the study of regional variation, but it can serve the purposes of sociolinguistic inquiries well, especially when diachronic variation is also considered.¹² This is clearly evidenced by Mancini's (2019) epigraphic account of the digraph <XS>. By comparing the occurrences of this digraph with those of the customary orthography <X> in a large corpus of documents spanning several centuries, and differing with respect to the formality parameter (see above), Mancini was able to describe the “changing fate” of the spelling <XS> (instead of <X>) in Latin history. He demonstrated that this spelling, “[f]enomeno colto [...] e appartenente ai registri istituzionali in epoca arcaica, [...] finì con lo scomparire dalle varietà solenni ed ufficiali [a learned phenomenon belonging to the institutional register of archaic times, eventually disappeared from formal and official varieties]” from the Augustan period onwards, to only appear “in ambiti informali fino a epoca tardissima [in utterly informal contexts until very late]”.¹³

The results obtained from the calculation of the “degree of error”, however, might not always be so clear to interpret (especially when the analysis of phonetic/phonological variation is addressed and diachronic variation cannot be taken into account). This is due to the fact that such a calculation is profoundly influenced by the educational level of the writer(s) who produced the epigraphic documents under investigation, as identified by their ability to adhere to

⁹Mancini (2014, 38).

¹⁰On this point, see especially Adamik (2012).

¹¹Adams (2007, 629).

¹²On the former point, see *infra* nn. 15 and 19. On the latter, cf. Mancini (2014, 39): “[l]e regole del ‘tempo visibile’, e soprattutto, la comparazione con la diacronia linguistica [...] permetteranno di accertare se una determinata variabile sia stata [...] selezionata dalla porzione residuale [...] dei parlanti [the rules of ‘apparent time’ and, above all, comparison with linguistic diachrony will determine whether a given variable was selected by the residual portion of speakers]”.

¹³Mancini (2019, 47).

the CL orthographic norm, regardless of the pronunciation that they adopted in speech.¹⁴ This can be expected to yield significant implications for sociolinguistic (quantitative-based) inquiries of the past, similarly to how it affects the study of regional variation.¹⁵ As anticipated earlier in this section, the typology of each inscriptional document (and, by extension, its formality) can reasonably be assumed to influence the writing attitudes of the author(s) who produced it, so that it can be hypothesised that official texts will exhibit a higher degree of spelling correctness than private (= less formal) documents.¹⁶

This is particularly important for the problem of how to define frequency in (quantitative) sociolinguistic studies of the epigraphic sources, since the fact that certain inscriptional deviations are less frequent in [+formal] inscriptions than in [–formal] texts (with respect to the corresponding correct spellings) is in principle not sufficient evidence that the underlying linguistic changes had a different diffusion in the varieties of Latin represented by such diaphasically-different documents. Conversely, this circumstance may simply reflect the fact that the author(s) of [+formal] inscriptions made a greater effort to reproduce CL orthography than did the authors of [–formal] texts.¹⁷

For this reason, this article will advocate the hypothesis that quantitative/statistical sociolinguistic accounts based on inscriptions should define the frequency of one (or more) error(s) in diaphasically-different texts, not only by determining the “degree” of these errors, but also by considering the number of other linguistically-relevant phenomena that occur in the same texts, i.e. by also considering the “relative frequency” of the “faults” in the corpus (as defined by Adamik 2012, 131). This is because, unlike calculations based on the “degree of error”, calculations based on the “relative frequency of faults” are less influenced by extra-linguistic factors, such as the ability of the author(s) who produced the texts under scrutiny to reproduce the CL orthographic norm, meaning that they can be of great use not only in the case of diatopic research (the field for which they were conceived; cf. Herman 2000b), but also in the case of sociolinguistic studies.¹⁸

To substantiate this claim, the quantitative/statistical analysis in section 3 will calculate the frequency of the instances of <E> for <AE> in a corpus of syntopic, synchronic, but diaphasically-different inscriptions from Rome (as proposed by Mancini 2014) according to both types of calculations. The purpose of this analysis will be to illustrate how the results of the former

¹⁴Adams (2013, XVI). On this point (i.e. the complex relationship between speech and writing in historical sociolinguistics), see also Mancini (2014, 38).

¹⁵For a detailed discussion of this problem in diatopic studies, see Herman (2000b, 125–126), Adams (2007, 630–631), and Adamik (2012, 128–131).

¹⁶Kruschwitz (2015, 738).

¹⁷The epigraphic survey by Donati (2015) is also based on the premise that [+formal] texts should theoretically exhibit a higher degree of orthographic correctness than [–formal] documents. Donati analysed the distribution of CL vs. non-CL phonological variables in a corpus of 386 inscriptions displaying different degrees of formality from the 3rd and 2nd centuries BCE. The aim was to ascertain whether the distribution of these variables correlated with the formality of the texts under investigation. The survey yielded negative results, which the author attributes to the chronology of the inscriptions under investigation (whose production preceded the standardisation of Latin; see section 2 below). However, this is not the case for the inscriptions in our corpus, as section 3 will discuss in more detail.

¹⁸A similar discussion of calculations based on the “degree of error” in sociolinguistic studies can be found in Papini (2024, 99–101).

calculation (“degree of error”) can be complemented by the results of the latter (“relative frequency of faults”).¹⁹ Before moving on to this analysis, the epigraphic corpus considered will be described in section 2. Finally, section 4 will draw some conclusions.

2. MATERIALS AND METHODS

In order to address the problem discussed in the previous pages (i.e. the definition of frequency in sociolinguistic inquiries of the epigraphic sources), this paper will analyse the instances of the spelling variation with <E> vs. <AE> in a corpus of 1,650 (synchronic and syntopic) inscriptions written in Rome between approximately the year 300 CE and approximately the year 600 CE. On the basis of their dating, these “Christian” inscriptions can be regarded as representing a specific phase in the “sociolinguistic history” of Latin, i.e. the phase contemporary to the affirmation of so-called “neo-standard” Latin, the (new) communicative register “fatto proprio dai ceti dominanti della società tardoimperiale [adopted by the dominant classes in late imperial society]”²⁰ (after the 3rd cent. CE), but prior to the communicative breakthrough that would eventually lead to the (gradual) emergence of the Romance languages from about the 7th cent. CE onwards.²¹ This is particularly important for the purposes of the present study. First, the language of the epigraphic records examined here can still be seen as reflecting a unified linguistic system. Second, and most importantly, the existence of a neo-standard variety of Latin, coexisting with the “traditional standard variety” (CL) and incorporating a number of previously substandard features, can reasonably be assumed for this chronological phase.²²

With regard to the phenomenon studied, it can safely be assumed that the monophthongised pronunciation of /ae/ (= [e:]) was one of these features. As various scholars have emphasised, not only was /ae/ monophthongised in Latin by the 1st cent. CE, but (ɛ:) (< /ae/) also lost its

¹⁹In this case, it can be emphasised that recent scholarly accounts have shown that (*contra* Herman 2000b, 126 and Adamik 2012, 131–135), performing both types of calculations can also be advantageous in diatopic (as opposed to merely sociolinguistic) studies (at least in restricted circumstances). On this point, see Papini (2022b) and Tamponi (2022, 33–35). However, the idea that calculations based on the “degree of error” might be applicable to sociolinguistic inquiries is shared even by Adamik (2012, 131).

²⁰Mancini (2005, 143).

²¹Cf. Adamik (2015, 649–650). In this paper, “Christian” inscriptions are defined as epigraphic records composed after approximately the year 300 CE, in line with traditional definitions in Latin epigraphic studies (see, among others, Herman 2000b, 127–128).

²²This is in line with the recent findings in contemporary sociolinguistics, for these findings have shown that in re-standardisation processes “the traditional standard variety” can come to coexist with the “newly-emerged standard variety which includes many informal spoken features” (Cerruti & Vietti 2022, 261). In the case of Latin, the “traditional standard variety” can be identified with the form of the language codified during the standardisation of the 1st cent. BCE – 1st cent. CE (cf. Clackson 2011b), while the “newly-emerged standard variety” can be identified with so-called “neo-standard” Latin (cf. *supra*). The assumption of the existence of a neo-standard variety (coexisting with so-called CL) within the chronological period under investigation is of particular relevance here, as it addresses the issue raised by Donati (2015). See n. 17 above. On the applicability of the findings of Cerruti and Vietti to the present analysis, cf. *infra* n. 34.

markedness of stigmatised variable in the following centuries (unlike <E> for <AE>).²³ This is also crucial for the analysis in the following pages. Since (ε:) was no longer stigmatised in the 4th–6th cent. CE (the dating of our inscriptions), cases of the spelling variation with <E> vs. <AE> in our corpus can reasonably be expected to be primarily a reflection of the educational level of the writers who produced this corpus (i.e. their ability to write “correct” Latin regardless of their speech habits).²⁴ This renders such cases particularly suitable for interpreting the results of the quantitative methods described in the previous section (namely, the calculation of the “degree of error” and that of the “relative frequency of faults”), for these cases can be used to illustrate the implications of the complex dynamics involving “speech” and writing in sociolinguistic studies of the epigraphic sources (at least as far as “Christian” inscriptions are concerned; see section 3 below).

Regarding formality, following the classification already applied in Papini (2022a, 241–243, 2024, 112–115), all the inscriptions considered in this article have been classified into three “sociolinguistic categories” that differ with respect to the diaphasic dimension, namely, formal, informal, and diaphasically low texts (Table 1). By formal, I mean epigraphic records that were intended for public display and were specifically realised for a public purpose (e.g. building inscriptions and honorary texts = 578 records). The term informal refers to those inscriptions that were still intended for public display, but were made for a private purpose, such as

Table 1. Classification of the 1,650 inscriptions from Rome considered (based on formality)

Formal (578)	No.	% corpus (35)
Building inscriptions/Captions	273	16.5
Honorary inscriptions	190	11.5
Legal texts	19	1.2
Public sacred inscriptions	96	5.8
Informal (807)	No.	% corpus (48.9)
Funerary inscriptions	769	46.6
Private sacred inscriptions	38	2.3
Diaphasically low (265)	No.	% corpus (16.1)
Curse tablets	16	1.0
Graffiti	203	12.3
Objects of everyday use/brick stamps	46	2.8

²³See Adams (2013, 80): “by the fourth century the monophthong (*scil.* [ɛ:] < /ae/) was so established that it was acceptable even to grammarians”. However, regarding the use of <E> for <AE>, the same scholar points out that “[f]or centuries most writers aspiring to correctness attempted to go on using the digraph *ae*, long after it had ceased to be representative of speech” (Adams 2013, 79).

²⁴Cf. Herman (2000b, 126) and Adamik (2012, 135).

“ordinary” funerary texts (807 records in total).²⁵ Finally, the category diaphasically low includes all those epigraphic documents that had neither a public purpose nor a specific (or only limited) intention for public display (e.g. graffiti or curse tablets = 265 records).²⁶

The “sociolinguistic” classification in Table 1 (based on formality) will serve as the starting point for the quantitative/statistical study in the next section.²⁷ However, before turning to this study, the issues associated with the possible non-independence of the epigraphic data analysed and their qualitative evaluation should be addressed. Regarding the former issue (non-independence), the discussion in the first part of this section has clarified that the data considered in this study were found in an extensive corpus of 1,650 inscriptions realised in Rome over the course of about three centuries (ca. 300 CE – ca. 600 CE). This makes it highly unlikely that the data represent the linguistic habits of just a limited number of speakers, despite the attestation of various linguistic forms within the same inscription(s).

When it comes to the qualitative evaluation, section 1 has anticipated that quantitative/statistical accounts of the inscriptional sources should always be preceded by a comprehensive qualitative analysis of the epigraphic deviations under scrutiny (in the present case, instances of the spelling variation with <E> vs. <AE>). The purpose of this analysis should be to ensure that only those deviations that are more likely to be a genuine reflection of the phenomenon (or phenomena) of interest (in this essay, the monophthongisation of /ae/) are included in the statistics, while all those deviations for which an alternative explanation (either linguistic or non-linguistic) is possible are excluded.²⁸

This principle has been followed in the present article. To begin with, the quantitative/statistical study in section 3 did not take into account cases of <E> for <AE> that occurred in the dative singular of first declension Greek feminine personal names (e.g. <Selene> instead of *Selenae* ‘Selene(F).DAT.SG’, cf. Gk. Σελήνη = *ICUR*, IV 10271; 4th cent. CE). This is because, in

²⁵The 1,650 “Christian” inscriptions analysed here are part of the larger epigraphic corpus (6,787 records) defined in Papini (2024). In this case, it should be emphasised that public/private sacred inscriptions do not belong to the same “sociolinguistic category” (the former having been classified as formal documents rather than as informal texts; see Papini 2024, 114–115). A similar distinction is made in the CLASSES database (<https://classes-latin-linguistics.fileli.unipi.it/en>), where *tituli sacri privati* vs. *publici* are classified as independent textual typologies.

²⁶The most represented textual typology among diaphasically low text in our corpus is catacomb graffiti (203 out of 265 records). These types or documents can in all probability be considered to represent the Latin of Rome, since, before the transformation of the catacombs into sanctuaries associated with the cult of the martyrs in the 7th cent. CE, the “visitors” to these sacred places were mostly Roman citizens (cf. Felle 2021, 59). Moreover, it can be assumed that ancient graffiti had a much less “public” purpose than modern ones (see Cooley 2012, 111 and 239).

²⁷In this case, the grouping of our texts into three “sociolinguistic categories” (instead of analysing each textual typology separately) is also a partial solution to the problem that the analysis of certain typologies (e.g. the *instrumenta inscripta*) is problematic with quantitative approaches due to their brevity (cf. Tamponi 2022, 120, 156, and 172).

²⁸On this point, see Adams (2013, 37): “[m]ere statistics showing the incidence of misspellings cannot be accepted on their own [...]. The individual tokens must be examined, and those that might be explained in other ways eliminated”.

such cases, “-e is not a monophthongisation of -ae, but represents the Greek feminine dative -ῆ”²⁹ (i.e. these cases might well have an alternative linguistic explanation).³⁰ As far as alternative explanations of a non-linguistic nature are concerned, the *e*-spellings (instead of /ae/) occurring in our corpus whose interpretation was problematic were also excluded from the quantitative/statistical analysis in the next section. Examples of this are those forms whose reading was uncertain due to severe damage to the epigraphic support (e.g. <[—]ime>, probably for <[dulciss]ime> instead of *dulcissimae* ‘sweet.F.DAT.SG’ = *ICUR*, IV 9556; 342 CE).

Having excluded the cases of <E> for <AE> in our corpus that could have an alternative explanation (either linguistic or non-linguistic), 328 *e*-spellings (instead of /ae/) remained that are more likely to reflect the monophthongised pronunciation [e:] ≠ [ae]. The calculation of the frequency of these *e*-spellings in the texts of the three “sociolinguistic categories” defined above (either in relation to the corresponding correct spellings, or in relation to the number of other phonological errors that occur in the same texts) will be the subject of the quantitative/statistical study in the following pages. The results of this study will allow some considerations on the calculation of frequency in sociolinguistic inquiries of the epigraphic sources.

3. QUANTITATIVE AND STATISTICAL ANALYSIS

To discuss the implications of the quantitative/statistical approaches to the study of sociolinguistic variation in Latin introduced in section 1 (i.e. the calculation of the “degree of error” and that of the “relative frequency of faults”), the proportion of the cases of <E> for <AE> occurring in the 1,650 diaphasically-different inscriptions included in our corpus was first calculated in relation to the number of the cases of correctly written <AE> that are found in the same inscriptions (“degree of error”). This first analysis is summarised in Table 2 (and in the accompanying Figure 1).

The findings of the quantitative study based on the calculation of the “degree of error” suggest that the distribution of the investigated instances of the spelling variation with <E> vs. <AE> is strongly influenced by the formality parameter. This is evidenced by the fact that

Table 2. <E> for <AE> in Rome (“degree of error”)

Formality	<AE>	<E>
Formal	988 (93.8%)	65 (6.2%)
Informal	275 (53.8%)	236 (46.2%)
Diaphasically low	14 (34.1%)	27 (65.9%)

²⁹Adams (2003, 475). See also Galdi (2004, 29) and Papini (2019, 106).

³⁰*ICUR*, IV 10271 (ll. 1–3) reads: *Victoriae Selene* / [a]nimae / [in]nocentissimae. All Latin lexemes in the text spell <AE> correctly, which further suggests that <Selene> in the inscription (a funerary text showing the “Christian” symbol of the dove; cf. *ICUR*, IV 10271, p. 132) is an attempt to reproduce the Gk. form Σελήνη. This hypothesis is even more likely if one considers that Greek features were often used by Christians in Rome as a sign of group identity (see Felle 2018).

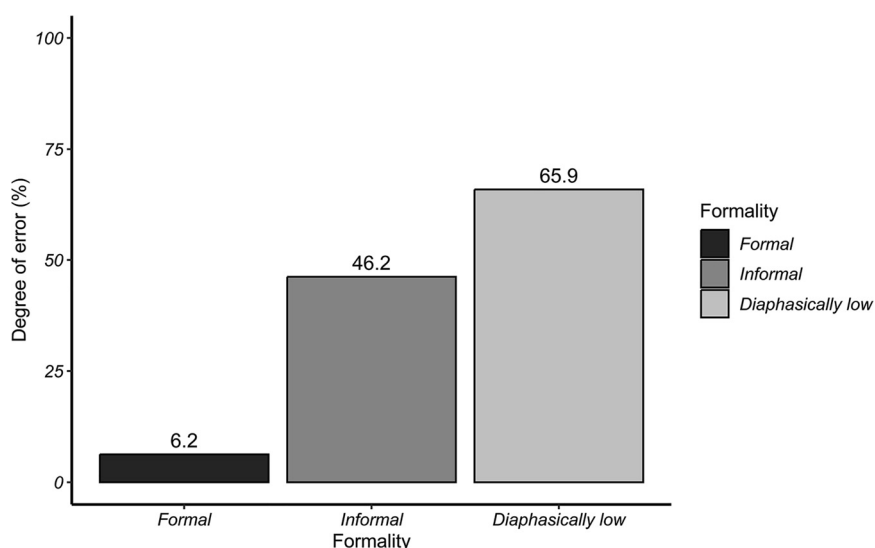


Figure 1. <E> for <AE> in Rome (“degree of error”)

such instances are less frequent in formal inscriptions than in either informal or diaphasically low texts (when compared to the corresponding correct spellings). Formal inscriptions have 65 occurrences of <E> for <AE> compared to 988 occurrences of correctly written <AE>, representing a “degree of error” of only about 6.2%. On the contrary, such low figures could not be observed in the texts of the other two “sociolinguistic categories”. In informal inscriptions, <AE> is correctly spelled 275 times and incorrectly represented by <E> 236 times (“degree of error” = ca. 46.2%). Finally, the diaphasically low texts included in our corpus (e.g. graffiti or *tabellae defixionum*) spell <AE> correctly in just approximately 34% of the possible cases (14 out of 41), thus exhibiting a “degree of error” of about 65.9%.³¹

According to the *Conditional Inference Tree* (CIT) in Figure 2, the distribution of the cases of <E> vs. <AE> in the formal, informal, and diaphasically low texts included in the inscriptional sample investigated in this article is statistically significant. CIT’s are classification (and regression) models that operate on the principle of recursive partitioning. In essence, the model tests the statistical association between the predictor(s) under consideration (or their values) and the selected outcome (in this case, “formality” and “degree of error”, respectively). This results in the original dataset being “partitioned” into two subsets based on the “strongest” association found, and the process being repeated until there is no significant association left to justify further

³¹The instances of correctly written <AE> in the 1,650 inscriptions considered were counted after removing all the characters occurring between brackets in the (digital) edition of the texts in the corpus using regular expressions (REGEX). This allowed us to consider only those cases of <AE> that actually appeared on the stone. For the same reason, epigraphic forms involving the hypercorrect use of this digraph (e.g. <diae> instead of *die* ‘day’_{(M).ABL.SG} = *ICUR*, VIII 23422; first quarter of the 5th cent. CE) were also removed from the count.

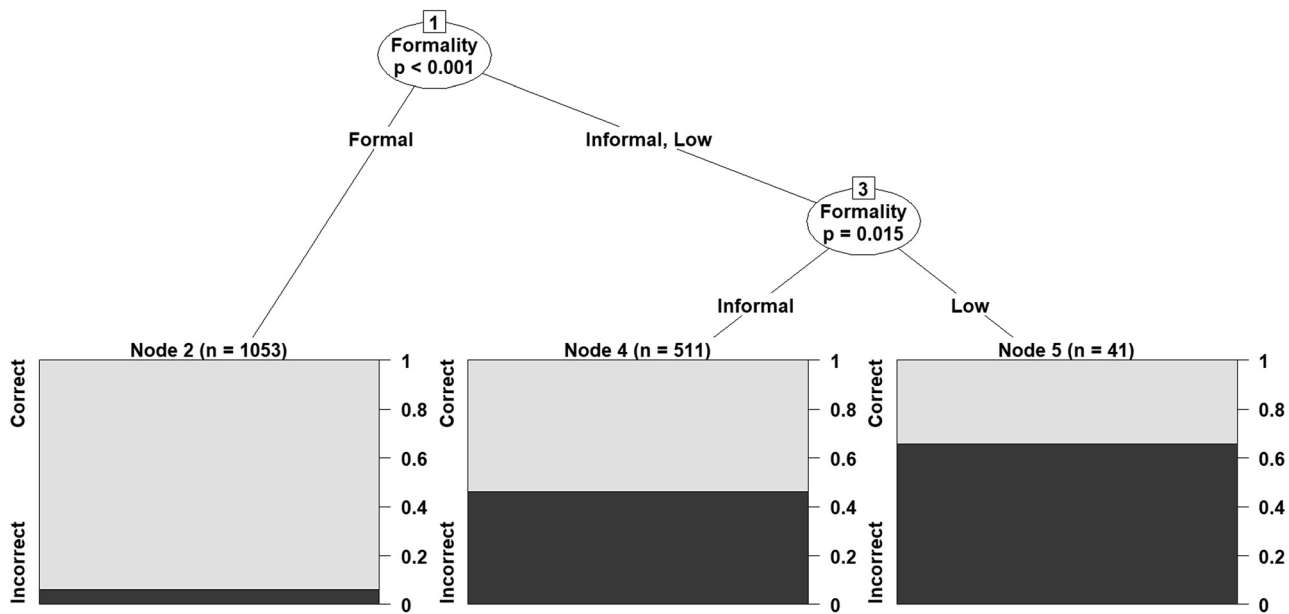


Figure 2. <E> for <AE> in inscriptions from Rome (Conditional Inference Tree: “Degree of error” ~ “Formality”)

partitioning (at a pre-determined level of statistical significance).³² The results of the analysis can then be visualised as a “tree-structure” (as in Figures 2 and 4). This structure represents each “partition” made by the algorithm as a different node of the “tree”, as well as indicating the significance of that partition at a statistical level (as represented by the associated *P*-value).³³

In the case of our data, the CIT in Figure 2 made an initial partition (based on “formality”) separating formal texts from informal and diaphasically low inscriptions (node 1). This represents the fact that cases of <E> for <AE> are significantly less attested in formal inscriptions than in the epigraphic records of the other two “sociolinguistic categories” (*P*-value <0.001). Furthermore, an additional partition between informal and diaphasically low texts was made at node 3, indicating that, according to our model, the *e*-spellings (instead of /ae/) under investigation occur significantly more often in diaphasically low inscriptions than in informal texts (when compared to the number of the corresponding correct alternatives) (*P*-value = 0.015). In other words, the statistical analysis presented by the CIT in Figure 2 validated the results of the quantitative study in the initial part of this section. It confirmed that “formality” is a relevant predictor of whether <E> vs. <AE> would be used in texts belonging to different “sociolinguistic categories”, by showing that the *e*-spellings (instead of /ae/) found in our corpus are less frequent in informal inscriptions than in diaphasically low texts, and least frequent in formal epigraphic records.

This state of affairs, which (apparently) contradicts the view that (ε:) (</ae/) was no longer a stigmatised sociolinguistic variable in (Roman) Latin by the 4th–6th cent. CE (see section 2), requires further commentary. The distribution of the cases of <E> vs. <AE> in the diaphasically-different inscriptions included in our corpus (with <AE> mostly attested in formal inscriptions and <E> occurring with the highest frequency in diaphasically low texts) could be justified by the hypothesis that the diphthong ([ae]) was still maintained in the highest varieties of Latin that are reflected by the orthography of formal texts. This hypothesis would be in line with the findings of contemporary sociolinguistics, which have shown that, during processes of re-standardisation (such as the one that Latin underwent in the period analysed here; cf. *supra*), the speakers can still choose to use “traditionally standard” (as opposed to neo-standard) variables, in the present case, (ae) vs. (ε:), in [+formal] communicative contexts.³⁴

However, the main reason for the fact that our formal inscriptions correctly spell <AE> significantly more often than informal and diaphasically low texts is most likely that, as discussed in section 2, these inscriptions are a more faithful representation of CL orthography

³²In this research, *P*-values <0.05 were considered significant, in line with the traditional use in statistical studies.

³³The description of CITs provided here is based on Levshina (2015, 291–292). These types of models are generally applied to analyse datasets with small sample size and large number of predictors (cf. *ibidem*). In the present case, the decision to utilise CITs on a dataset with a single predictor (“formality”) is motivated by the fact that such models allow the association between different values of the predictor(s) to be explored and visualised more efficiently than using traditional statistical techniques. On the use of CITs in Latin studies (particularly inscriptions), see Szlovicsák (2024) and, especially, Papini (2024). Both the CIT in Figure 2 and that in Figure 4 (see below) were fitted using the R package “partykit” (cf. <https://cran.r-project.org/web/packages/partykit/index.html>).

³⁴On this point, see the sociolinguistic study conducted by Cerruti & Vietti (2022, 271–276) on a sample of 36 university students. The study indicated that the student’s choice of whether to use features associated with the traditional vs. neo-standard variety of contemporary Italian correlated with the formality of the communicative context. However, it should be emphasised that these scholars did not take into account the complex relationship between speech and writing, which is central in this study.

(i.e. their authors made a greater effort to write “correct” Latin than did the authors of the texts of the other two “sociolinguistic categories”, regardless of their speech habits). This view can be confirmed by calculating the “frequency” of the *e*-spellings (instead of /ae/) in the diaphasically-different texts under investigation in relation to the number of other phonologically-relevant deviations (related to changes in the vowel system) that occur in the same texts, i.e. by applying a calculation based on “the relative frequency of faults” (see Table 3 and Figure 3 below).

As shown in Table 3, the results obtained from the calculation of the “relative frequency of faults” are mirror images of those obtained from the calculation of the “degree of error”. This emerges from the fact that, compared to the number of other errors of a phonological nature attested in the corpus, cases of <E> for <AE> occur most frequently in formal texts (where they represent about 55.6% of the total number of deviations, namely, 65 out of 117) and least frequently in diaphasically low inscriptions (where they appear with a relative frequency of only about 33.8% = 27 out of 80). When it comes to the informal epigraphic documents, these documents have 236 cases of <E> for <AE> and 225 cases of other types of phonological “misspellings” reflecting various changes in the vowel system (other than /ae/ > [ɛ:]). The percentage of <E> vs. <AE> confusions in these documents is, therefore, about 51.2% (a value similar to the value observed in the case of formal inscriptions).³⁵

Following the same procedure applied in the case of the “degree of error”, the statistical significance of the distribution of the cases of <E> for <AE> in our corpus, as described by the calculation of the “relative frequency of faults”, was also assessed by means of a *Conditional Inference Tree* (CIT) (Figure 4). Similarly to the CIT in Figure 2, the CIT in Figure 4 also uses “formality” as a predictor. However, unlike the former CIT (whose outcome was “degree of error”; see above), the outcome used in the statistical analysis of the “relative frequency of faults” was “type of error” (i.e. cases of <E> for <AE> vs. “others”). In the latter case, the algorithm could not detect a statistically significant difference in the relative frequency of the cases of <E> for <AE> in formal and informal texts (node 2). Instead, the model

Table 3. <E> for <AE> in Rome (“relative frequency of faults”)

Formality	Others	<AE>/<E>
Formal	52 (44.4%)	65 (55.6%)
Informal	225 (48.8%)	236 (51.2%)
Diaphasically low	53 (66.2%)	27 (33.8%)

³⁵The list of “other phonological spellings” considered in this article includes those vowel confusions that are usually attested in “Christian” inscriptions. Examples are: cases of 1) syncope (e.g. <domnae> instead of *dominae* ‘lady_{(F).DAT.SG}’ = *ICUR*, IX 24853, 9; first quarter of the 6th cent. CE), 2) prosthesis (e.g. <iscola> for *schola* ‘school_{(F).ABL.SG}’ = *CIL*, VI 32965; 4th cent. CE), 3) <I>/<V> for /e/ or /o/ in hiatus position (e.g. <timiant> instead of *timeant* ‘be afraid_(PRS.SUBJ.3.PL)’ = *AE* 2010, 221; 4th–5th cent. CE), and 4) cases of <E> vs. <I> and <O> vs. <V> that point to the merger of /i, u/ and /e, o/ as /e/ and /o/, respectively (e.g. <vede> for *vīde(t)* ‘see_(PRS.IND.3.SG)’ = *AE* 1941, 138; 4th–5th cent. CE). Similarly to the count of the instances of <E> vs. <AE> that are the primary focus of this article, the count of the “other phonological spellings” does not include those cases for which an alternative explanation (either linguistic or non-linguistic) is possible (see section 2 above).

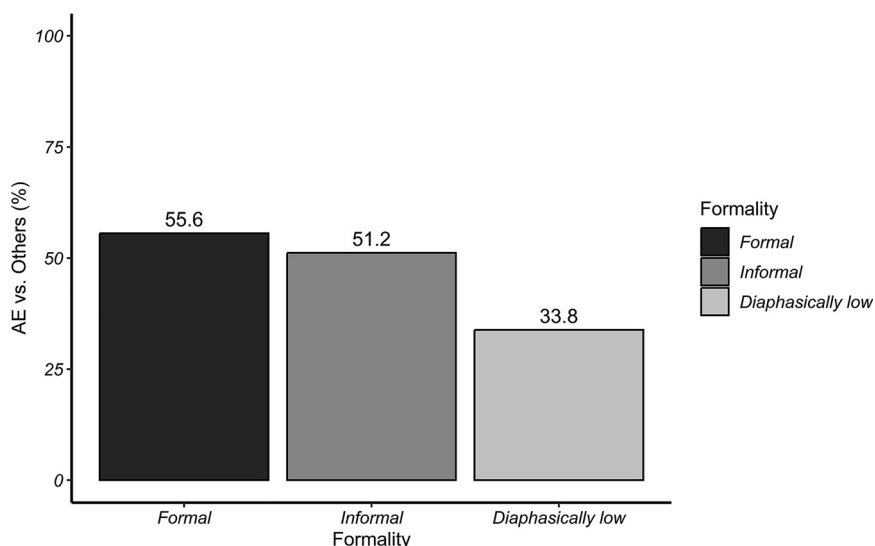


Figure 3. <E> for <AE> in Rome ("relative frequency of faults")

distinguished diaphasically low inscriptions and the epigraphic records of the other two "sociolinguistic categories" (node 1), thus confirming that, when compared to the number of other phonologically-relevant "misspellings" occurring in the corpus, cases of <E> for <AE> occur significantly less frequently in diaphasically low inscriptions than in both formal and informal texts (P -value = 0.006), a completely specular image from the one revealed by the calculation of the "degree of error" (cf. *supra* this same section).

The results of the two quantitative/statistical approaches applied in the previous pages (i.e. the calculation of the "degree of error" and that of the "relative frequency of faults") can now be cross-referenced to provide a picture that is more in line with what has been said in section 2 about the status of (<ɛ>) in the Latin diasystem during the 4th–6th cent. CE. First, the fact that cases of <E> for <AE> are least frequent in diaphasically low texts in relation to the number of other "misspellings" of a phonological nature is very unlikely to indicate that the monophthongised pronunciation of /ae/ was not widespread in the (low) registers of Latin that these types of texts represent. For this hypothesis is strongly contradicted by the fact the authors of diaphasically low inscriptions use <E> instead of <AE> much more often than the writers of both formal and informal epigraphic records (as shown by the CRT in Figure 2). Rather, this circumstance is most likely a reflection of the fact that diaphasically low inscriptions attest to a number of "other misspellings" that are not represented (to the same extent) in the language of the other two "sociolinguistic categories" (such as instances of the spelling variation with <E> vs. <I> and <O> vs. <V> that anticipate the vowel mergers characterising most Romance languages, i.e. /i, e:/ > /e/ and /u, o:/ > /o/).³⁶ Same, and conversely, the calculation of the "relative frequency of faults"

³⁶On this point, see the discussion by Papini (2024, 291–308), where /i, e:/ > /e/ and /u, o:/ > /o/ are studied using a similar methodology as the one used here (including the application of CRTs).

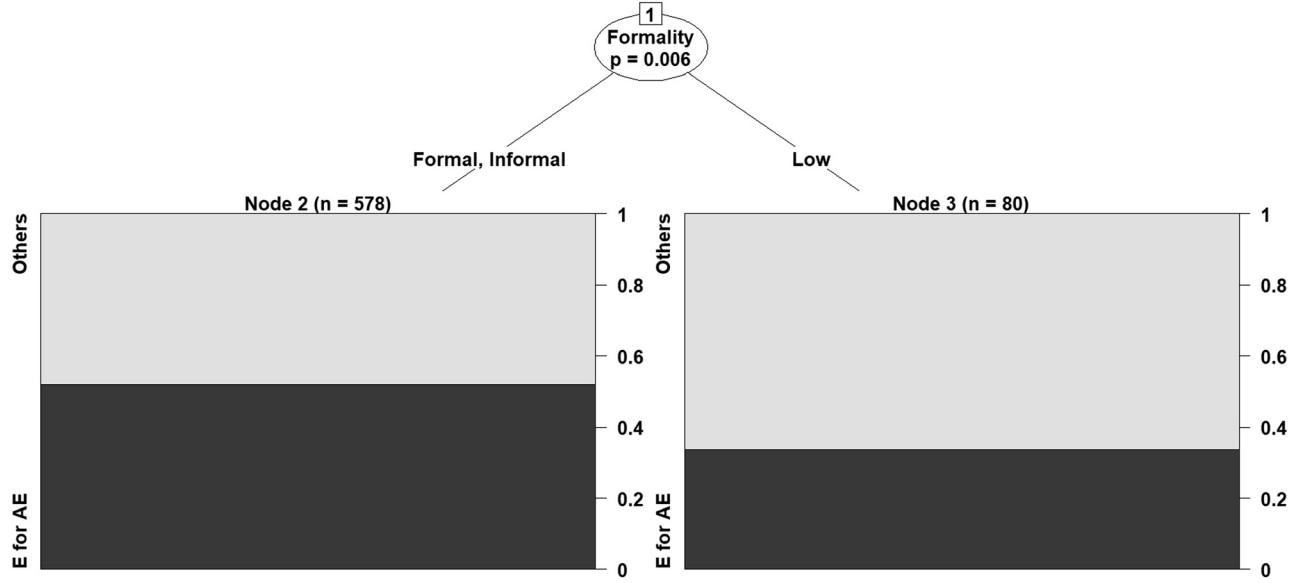


Figure 4. <E> for <AE> in inscriptions from Rome (*Conditional Inference Tree: “Type of error” ~ “Formality”*)

(cf. Figure 4) suggested that the low “degree of error” observed in formal inscriptions with respect to the use of <E> for <AE> is not sufficient evidence that (ɛ:) did not reach the (highest) varieties of Latin reflected by the language of such inscriptions in the (late) chronological period analysed in this article (4th–6th cent. CE), but that this circumstance is most likely to be an indication of the fact that authors of these types of inscriptions made a greater effort to “respect” CL orthography (i.e. “correctly” write <AE>) than did the authors of informal and diaphasically low epigraphic records.

In other words, the discussion in the previous pages provided significant evidence that both the quantitative/statistical approaches utilised here can offer relevant insights for sociolinguistic studies based on the epigraphic sources, thereby confirming the view expressed in section 1 that both approaches should be applied in these types of studies, even more so than in the case of diatopic inquiries of Latin.³⁷

4. CONCLUSIONS

This article addressed the problem of defining “frequency” in historical sociolinguistic inquiries of the inscriptional sources. Section 1 provided a (necessarily partial) literature review on the subject, highlighting the importance of the epigraphic evidence for the study of sociolinguistic variation in Latin. It was pointed out that such study should prioritise the analysis of diaphasic (as opposed to diastratic) variation (Mancini 2014), in addition to introducing the quantitative/statistical approaches traditionally employed in Latin inscriptional studies, namely, the calculation of the “degree of error” (Adams 2007) and that of the “relative frequency of faults” (Herman 2000b; Adamik 2012).

Section 2 described the corpus of 1,650 inscriptions (ca. 300 CE – ca. 600 CE) from Rome that formed the object of the present research (these inscriptions were classified into three “sociolinguistic categories”: formal, informal, and diaphasically low). The quantitative/statistical analysis in section 3 (considering the vowel alternations with <E> vs. <AE> occurring in the corpus) supported the claim that both quantitative/statistical methods just mentioned should be applied in sociolinguistic inquiries of inscriptions. This was illustrated by showing how the results of such methods (both implemented via the use of *Conditional Inference Trees*) can be cross-referenced to offer a more reliable picture of the complex dynamics that characterised the Latin diasystem.

In the case under consideration, the calculation of the “relative frequency of faults” allowed us to confirm that initial hypothesis that (ɛ:) (</ae/) was no longer a stigmatised sociolinguistic

³⁷For the application of both approaches in diatopic studies, see n. 19 above. A similar picture to the one highlighted here regarding the “degree of error” of <E> vs. <AE> in a corpus of diaphasically-different inscriptions from Rome (about 50 CE – about 250 CE) was found in Papini (2019, 108–109). However, it is not clear whether this actually indicates that (ɛ:) (</ae/) was still a substandard variable at that time. This requires further investigation, which will be offered in a future contribution. An attempt to use the “relative frequency of faults” for sociolinguistic studies of the epigraphic sources is also found in Tamponi (2022, 206). Before moving on to the conclusions section, it should be emphasised that, due to the limitations of the inscriptional sources (mainly, the linguistic interpretation of each deviation) the counts presented in this article (either deviations or correct spellings) cannot be regarded as perfect. Rather, these counts should be considered as a reliable approximation, that is sufficient for the purposes of the present analysis.

variable in (Roman) Latin between the 4th and 6th centuries CE (a fact that did not emerge from the calculation of the “degree of error”). Specifically, this could be demonstrated by observing that instances of <E> for <AE> in the corpus were well represented even (and especially) in formal inscriptions (when compared to the number of other deviations pertaining to changes in the vowel system). Same and conversely, the calculation of the “degree of error” made it possible to confirm that <AE> (instead of <E>) was still a sign of “good literacy” in the chronological period considered (regardless of the diffusion of the monophthong in all registers of Latin), as indicated by the use of the two graphemes in the orthography of formal texts, as well as allowing us to better frame the results obtained by the calculation of the “relative frequency of faults” with regard to the diffusion of (ε:) in the (low) sociolinguistic registers of Latin reflected by the language of diaphasically low texts.

ACKNOWLEDGMENTS/FUNDING INFORMATION

This article was prepared as part of the HORIZON-ERC-2022-ADG project (no. 101098102), entitled “Digital Latin Dialectology” (DILADI): Tracing Linguistic Variation in the Light of Ancient and Early Medieval Sources funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The data/corpus used in this article were retrieved/defined during a PhD fellowship at UGent (FWO – PhD Fellowship 11B0921N) (see [Papini 2024](#)). I would like to thank Professor Claudia Ciancaglini for reviewing an early version of this article, and the two anonymous reviewers for their insightful comments. I would also like to thank the editors, Professor Béla Adamik and Professor András Cser, for their invaluable suggestions. Any errors contained within the article are the sole responsibility of the author.

REFERENCES

- Adamik, Béla. 2012. In search of the regional diversification of Latin. Some methodological considerations in employing the inscriptional evidence. In F. Biville, M.-K. Lhommé and D. Vallat (eds.) *Latin vulgaire – latin tardif IX: Actes du IX^e colloque international sur le latin vulgaire et tardif*. Lyon: Maison de l’Orient et de la Méditerranée. 123–139.
- Adamik, Béla. 2015. The periodization of Latin: An old question revisited. In G. Haverling (ed.) *Latin linguistics in the early 21st century: Acts of the 16th International Colloquium on Latin Linguistics*. Uppsala: Uppsala Universitet. 638–652.
- Adams, James N. 2003. *Bilingualism and the Latin language*. Cambridge: University Press.
- Adams, James N. 2007. *The regional diversification of Latin, 200 BC – AD 600*. Cambridge: University Press.
- Adams, James N. 2013. *Social variation and the Latin language*. Cambridge: University Press.
- Cerruti, Massimo and Alessandro Vietti. 2022. Coexisting standards in spoken Italian. In K. V. Beaman and G. R. Guy (eds.) *The coherence of linguistic communities: Orderly heterogeneity and social meaning*. New York, NY: Routledge. 261–280.

- Clackson, James. 2011a. Latin inscriptions and documents. In J. Clackson (ed.) *A companion to the Latin language*. Chichester & Malden, MA: Wiley-Blackwell. 29–39.
- Clackson, James. 2011b. Classical Latin. In J. Clackson (ed.) *A companion to the Latin language*. Chichester & Malden, MA: Wiley-Blackwell. 236–256.
- Cooley, Alison E. 2012. *The Cambridge manual of Latin epigraphy*. Cambridge: University Press.
- Cuzzolin, Pierluigi and Gerd Haverling. 2009. Syntax, sociolinguistics, and literary genres. In P. Baldi and P. Cuzzolin (eds.) *New perspectives on historical Latin syntax: Syntax of the sentence*. Berlin & New York, NY: De Gruyter. 19–64.
- Donati, Margherita. 2015. Variazione e tipologia testuale nel corpus epigrafico CLASSES I. *Studi e Saggi Linguistici* 53(2). 21–38.
- Felle, Antonio E. 2018. The use of Greek in the early Christian inscriptions from Rome and Italy (3rd–4th cent.). In C. Breytenbach and J. M. Ogereau (eds.) *Authority and identity in emerging Christianities in Asia Minor and Greece: Ancient Judaism and early Christianity*. Leiden & Boston, MA: Brill. 303–325.
- Felle, Antonio E. 2021. Late antique Christian Graffiti: The case of Rome (third to fifth centuries CE). In A. Felle and B. Ward-Perkins (eds.) *Cultic graffiti across Mediterranean and beyond*. Turnhout: Brepols. 57–76.
- Galdi, Giovanbattista. 2004. *Grammatica delle iscrizioni latine dell'Impero (province orientali): Morfosintassi nominale*. Rome: Herder.
- Herman, József. 2000a. *Vulgar Latin*. University Park, PA: Pennsylvania State University Press.
- Herman, József. 2000b. Differenze territoriali nel latino parlato dell'Italia tardo-imperiale: Un contributo preliminare. In J. Herman, A. Marinetti and L. Mondin (eds.) *La preistoria dell'italiano: Atti della Tavola Rotonda di Linguistica Storica*. Tübingen: De Gruyter. 123–135.
- Janda, Richard and Brian Joseph. 2003. On language, change, and language change – or, of history, linguistics, and historical linguistics. In B. Joseph and R. Janda (eds.) *The handbook of historical linguistics*. Hoboken, NJ: Wiley-Blackwell. 3–180.
- Kruschwitz, Peter. 2015. Linguistic variation, language change, and Latin inscriptions. In C. Bruun and J. Edmondson (eds.) *The Oxford handbook of Roman epigraphy*. Oxford: University Press. 721–744.
- Labov, William. 1994. *Principles of linguistic change, Vol I: Internal factors*. Cambridge, MA & Oxford: Wiley-Blackwell.
- Levshina, Natalia. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam & Philadelphia, PA: John Benjamins Publishing.
- Mancini, Marco. 2002. Tra linguistica ed ermeneutica: Nuove acquisizioni falische. *Incontri Linguistici* 25. 23–46.
- Mancini, Marco. 2005. La formazione del neostandard latino: Il caso delle *differentiae verborum*. In S. Kiss, L. Mondin and G. Salvi (eds.) *Latin et langues romanes: Études de linguistique offertes à József Herman à l'occasion de son 80^{ème} anniversaire*. Tübingen: Niemeyer. 135–153.
- Mancini, Marco. 2014. Testi epigrafici e sociolinguistica storica. Le '*defixiones*' sannite. In R. Giacomelli and A. Robbiati-Bianchi (eds.) *Le lingue dell'Italia antica oltre il latino: Lasciamo parlare i testi*. Milan: Istituto Lombardo di Scienze e Lettere. 29–61.
- Mancini, Marco. 2019. Repertori grafici e regole d'uso: Il caso del latino <xs>. In L. Agostiniani and M. Marchese (eds.) *Lingua, testi, storia: Atti della giornata di studi in ricordo di Aldo Luigi Prosdocimi*. Rome: Giorgio Bretschneider editore. 13–53.
- Marotta, Giovanna. 2015. Talking stones: Phonology in Latin inscriptions? *Studi e Saggi Linguistici* 53(2). 39–64.

- Marotta, Giovanna. 2016. Sociolinguistica storica ed epigrafia latina: Il corpus CLaSSES I. *Linguarum Varietas* 5. 145–160.
- Nevalainen, Terttu and Helena Raumolin-Brunberg. 2012. Historical sociolinguistics: Origins, motivations, and paradigms. In J. M. Hernández-Campoy and J. C. Conde-Silvestre (eds.) *The handbook of historical sociolinguistics*. Chichester: Wiley-Blackwell. 22–40.
- Papini, Alessandro. 2019. Some preliminary remarks concerning sociolinguistic variation within the “vulgar” Latin vowel system: As evidenced by the inscriptional data. *Acta Classica Universitatis Scientiarum Debreceniensis* 55. 97–112.
- Papini, Alessandro. 2022a. The <e>/<i> spelling variation in Latin inscriptions from Rome (250 BC – AD 300): A preliminary (historical) sociolinguistic analysis based on the inscriptional data. *Revue Belge de Philologie et d’Histoire* 100. 231–274.
- Papini, Alessandro. 2022b. *Ipsa latinitas et regionibus cotidie mutetur et tempore*: Some methodological considerations on the use of Herman’s quantitative method. *Listy Filologické* 145(3–4). 343–378.
- Papini, Alessandro. 2024. Latin epigraphy and the “vulgar” Latin vowel system (ca. 250 BCE – ca. 600 CE): A qualitative, quantitative (and statistical) study based on the inscriptional sources of the city of Rome. Doctoral dissertation. Ghent University, Ghent.
- Pocetti, Paolo, Daniele Poli and Carlo Santini. 1999. *Una storia della lingua latina*. Rome: Carrocci.
- Szlovicsák, Béla. 2024. Gender confusions and other linguistic changes. *Acta Antiqua Academiae Scientiarum Hungaricae* 63(4). 575–594.
- Tamponi, Lucia. 2022. *Variation and change in Sardinian Latin: The epigraphic evidence*. Pisa: Pisa University Press.
- Winter, Walter. 1999. Sociolinguistics and dead languages. In E. J. Jahr (ed.) *Language change: Advances in historical sociolinguistics*. Berlin: De Gruyter Mouton. 67–84.