SL

CLaSSES: Orthographic variation in non-literary Latin

Giovanna Marotta, Francesco Rovai, Irene De Felice, Lucia Tamponi

Abstract

CLaSSES (Corpus for Latin Sociolinguistic Studies on Epigraphic textS) is a digital resource which gathers non-literary Latin texts (epigraphs, writing tablets, letters) of different periods and provinces of the Roman Empire. This corpus has been tagged with linguistic and extra-linguistic information that allows quantitative and qualitative analysis of spelling variations in Latin sources. The resource is available on the web in open access and is structured in different sections: *Rome and Italy*, *Roman Britain*, *Egypt and Eastern Mediterranean*, and *Sardinia*.

Keywords: Latin digital resources, Latin non-literary texts, historical sociolinguistics.

1. *Introduction*

CLaSSES, i.e. Corpus for Latin Sociolinguistic Studies on Epigraphic textS, is a digital resource that contains non-literary Latin texts (epigraphs, writing tablets, documentary letters) of different periods and provinces of the Roman Empire. The database is available on the web in open access (*http://classes-latin-linguistics.fileli.unipi.it*) and has been developed at the Laboratory of Phonetics and Phonology of the Department of Philology, Literature and Linguistics at Pisa University[1].

This new resource joins a growing list of digital tools (epigraphic collections, lemmatizers, syntactic treebanks, etc.) suitable for academic research on the Latin language, of which a representative sample is provided in this issue of *Studi e Saggi Linguistici*. In-between the lemmatizers and

---

[1]   The construction of the corpus began during the PRIN project *Linguistic representations of identity. Sociolinguistic models and historical linguistics* (PRIN 2010, prot. 2010HXPFF2_001). The initial plan of the database included only the section *Rome and Italy*. Over the last few years, the sections *Roman Britain*, *Egypt and Eastern Mediterranean*, and *Sardinia* have been added, while maintaining the original structure and layout (see below, § 3).

treebanks, whose reference corpora are mostly based on the literary texts, and the available digital epigraphic collections, which are not specifically designed for linguistic research, *CLaSSES* is a representative corpus of epigraphic and other non-literary documents annotated with linguistic information.

The working hypothesis is that non-literary texts (inscriptions, ostraka, documentary papyri, private letters, ink tablets) can be a direct and reliable source in order to approach a picture of the sociolinguistic variation that characterized the Latin-speaking world. In particular it is the (ortho-)graphic variants, as testified by the misspellings (i.e. those spellings that are not congruent with the 'standard' language as exhibited in the literary texts of the Classical period) occurring in non-literary texts, which can be assumed as being clues for linguistic variation.

Of course the current debate on the reliability of the inscriptional evidence for the investigation of linguistic variation and change in the ancient languages, is polarized between more or less skeptical views (cf. e.g. Adams, 2007; 2013 vs Herman, 1985). On the one hand, inscriptions, ink tablets, ostraka, and papyri are the only direct, first-hand evidence left from antiquity, while in any other kind of written text the mediation of the later philological and manuscript tradition is present. On the other hand, their value has always to be checked against a fine-grained analysis of the philological, paleographic, archaeological, and historical aspects, in order to reduce the problem of data sparseness that originates from the fragmentary nature of non-literary texts, as well as the problem of the authorship of the text.

At present, several scholars believe that non-literary texts can be regarded as a fundamental source for studying language variation (e.g. Molinelli, 2006; Kruschwitz, 2015; Marotta, 2015; Rovai, 2015; Consani, 2016), so that studies on the sociolinguistic aspects of Latin in Rome and the Empire have recently flourished[2], although some seminal works date back to some decades ago (e.g. Campanile, 1971; Vineis, 1984; 1993).

Building on this hypothesis, *CLaSSES* has been specifically designed in order to collect non-literary documents which attest spelling variants that could be indicative of phenomena occurring in the phonological or morpho-phonological realms. Such orthographic variants have been labelled as

---

[2]   See for instance Adams (2003; 2007; 2013); Rochette (1997); Biville *et al.* (2008, *éds.*); Dickey and Chahoud (2010, eds.).

'non-classical' forms, with reference to the standard spelling forms of Classical Latin. For every non-classical form, the corresponding classical one is also presented. For instance, a form like <MENOS> has been considered non-classical, since its corresponding form in the classical orthographic norm is <MINVS>.

Before illustrating the structure of this paper, an introductory methodological consideration is necessary here. Since the database is intended as an instrument for future research, the linguistic annotation of the misspellings is always kept as descriptive as possible and makes reference exclusively to the (ortho-)graphic level. Thus, the phenomena annotated in the case of <MENOS> for <MINUS> are labelled as *Vowel alternation - Classical* <I>, /ĭ/ = <E> and *Vowel alternation - Classical* <U>, /ŭ/ = <O>. This is tantamount to saying that a short *i*-sound of the Classical Latin is represented here through the letter <E> and that a short *u*-sound of the Classical Latin is represented through the letter <O> – and both spellings are inconsistent with the standard Classical orthography, where <I> and <U> occur. No preliminary assumption is therefore made about a possible relative chronology of the two variants, neither in the light of the etymological criterion nor in view of otherwise well-attested patterns of phonetic change. Whether these phenomena can be regarded as the relics of older spellings, or as an early anticipation of a Proto-Romance development, is left to the researcher's conclusive subsequent interpretation.

This paper is structured as follows: Section 2 provides a short review of the digital resources currently available for Latin epigraphy; in Section 3 the documents contained in *CLaSSES* are described with reference to the kind of material and the area of provenance; Section 4 presents the criteria of annotation by which textual data have been implemented with linguistic, meta-linguistic, and extra-linguistic information; Section 5 illustrates the technical aspects for the use of the search interface; finally, in Section 6 we summarize our conclusions.

2. *Digital resources for Latin inscriptions and other non-literary texts: An overview*

In this section, we shortly present the main digital resources available for the study of Latin epigraphy, with reference to database organization, data structure, and the user interface.

2.1. *Epigraphik Datenbank Clauss-Slaby*

Several open-access databases are available online for the study of Latin epigraphy (Feraudi-Gruénais, 2010; Elliott, 2015; cf. also the section *Inschriften in der digitalen Welt* in Eck and Funke, 2014, *Hrsg.*: 501-517) of which the Epigraphik Datenbank Clauss-Slaby (*EDCS*)[3] is at present the most complete digital collection of searchable Latin inscriptions. It records 520,061 texts from 22,232 findspots that cover the entire area of the Roman provinces. Each text is identified with an *EDCS-ID* number and annotated with information containing relevant bibliography, province and findspot. In many cases, further extra-linguistic and meta-linguistic data are provided: dating (for 179,365 inscriptions), material specification (for 187,543 inscriptions), the social status of the people mentioned in the text, and the textual typology (the classification for personal status and inscription genus is conflated under a single heading and is available for 210,844 inscriptions). There are also links (847,661) to other 36 databases, frequently with photos (for 191,027 inscriptions). In order to keep the presentation of the texts as simple as possible, the texts are presented without abbreviations and completed (where possible). The search engine allows for simple and combined word queries also by using Boolean operators and regular expressions, and searches can be limited using various entries of the metadata: records, province, place, dating, material, text type, personal status. Though not specifically designed for linguistic studies, *EDCS* is one of the most valuable sources for the investigation of language variation, since it is possible to search for misspellings (such as CONSENTIONT and TEMPESTATEBUS for *consentiunt* "agree.IND.PRES.3PL" and *Tempestatibus* "Goddess of Storm.DAT.PL.F") through the 'Search entries: wrong spelling' function.

However, it has to be noted that single linguistic forms (either words or groups of letters) rather than linguistic phenomena can be searched and browsed through this function, so that the researcher must already know which form to search for. In this way, as with the other databases described below, there is a risk of sliding into those limitations highlighted by Cordell (2015: 421): «most digital archives hide more than they reveal, as keyword searches require prior knowledge of the texts to be discovered and can lead to evidentiary excess».

---

[3]   Cf. *http://www.manfredclauss.de/* [accessed on 20.02.2020].

2.2. *The* EAGLE *network*

In addition to *EDCS*, reference is to be made to the *EAGLE* Project (Orlandi, 2017; Orlandi *et al.*, 2017, eds.; Prandoni *et al.*, 2017)[4], which began in 2003 as a network of four epigraphic digital archives (Epigraphische Datenbank Heidelberg-*EDH*, Epigraphic Database Roma-*EDR*, Epigraphic Database Bari-*EDB*, and Hispania Epigraphica Online-*HE*) with the aim of assembling the epigraphic collections held by the *EAGLE* partners, in order to provide scholars with a single portal to the searchable inscriptions of the Ancient World. The original four major databases remain pillars of the *EAGLE* network, but an up-to-date overview of the collections represented is available on the website (*https://www.eagle-network.eu/eagle-project/collections/*)[5].

The Epigraphische Datenbank Heidelberg (*EDH*)[6] contains the texts of Latin and Latin-Greek bilingual inscriptions from the Roman provinces, excluding Italy with Sicily, Sardinia and Corsica (for which see Epigraphic Database Roma-*EDR* below), and Spain (for which see Hispania Epigraphica Online below). *EDH* is made up of four constituent parts: Epigraphic Text Database (80,870 inscriptions), Photographic Database (39,031 photos), Bibliographic Database (16,481 records concerning monographs, articles in journals, and other specialist literature), Geographic Database (the geographical details of the 30,272 findspots of the inscriptions included in *EDH*). Users can perform simple full-text searches of words or groups of letters as well as more advanced queries while taking into account the metadata that enrich every single text: findspot, present location, dating (when available), type of inscription (e.g. honorific inscription, epitaph, votive inscription, etc.), language, material (e.g. marble, copper, amber), size and type of the monument (e.g. altar, *cippus*, *sarcophagus*, etc.), writing technique (e.g. engraved, painted, scratched, etc.), and historical relevant data (e.g. religion to which the monument belongs, troop names, people mentioned and rel-

---

[4]   In origin, the acronym was for Electronic Archive of Greek and Latin Epigrahy, but it is now expanded Europeana network of Ancient Greek and Latin Epigraphy.

[5]   An important role in the integration of different databases is played by Trismegistos (*TM*; cf. *https://www.trismegistos.org*), which is a central database of metadata (not texts) for papyrological and epigraphic documents from the Greco-Roman world, with a particular focus on prosopographical (*TM* People) and place (*TM* Places) identifications. *TM* currently includes more than 720,000 entries. Since networks of databases such as *EAGLE* and papyri.info (see below, § 2.4) inevitably show duplicate entries for some documents, by using the unique catalog numbers from *TM* (the so-called 'stable identifiers') as identification numbers, users can collate duplicate entries.

[6]   Cf. *https://edh-www.adw.uni-heidelberg.de/home* [accessed on 20.02.2020].

ative status, when available, etc.). Each text is also annotated with relevant bibliography and commentary.

The systematic gathering of the inscriptions from Italy and its islands, excluding Christian texts (for which see Epigraphic Database Bari-*EDB* below), is the domain of the Epigraphic Database Rome (*EDR*; Panciera, 2013; Caldelli *et al.*, 2014)[7]. Up to date, the *EDR* collection includes 91,336 inscriptions and 59,097 photos. Every text in this database is richly annotated with metadata concerning its dating, findspot and storage place, type of object, material, state and dimension of the support, writing technique of the inscription, language, and text type; when available, personal status of those mentioned in the text is specified. Finally, information concerning relevant bibliography is included. The online query interface allows words or groups of letters to be entered (possibly with Boolean operators) and simple and advanced queries can be made in combination with the following fields: record number, place of provenance (ancient region, current region, ancient city, modern city), current location, object type, material, measurements, state of textual preservation, writing technique, language, religion, verse, inscription type, type of persons mentioned, apparatus, and dating.

Epigraphic Database Bari (*EDB*; Rocco, 2017)[8] is specialized in Christian epigraphic documents from Late Ancient Rome (3rd-7th century AD) and includes 41,602 items and 7,891 images. In addition to the text, for each inscription the following metadata are recorded and featured for the interrogation of the database: graphical (reuse / opisthographic inscription, Greek alphabet), meta-linguistic (metrical text, function), and linguistic (Latin or Greek language) elements, material and executing technique, findspot and current location, dating, and figurative apparatus (*signa Christi*, symbols, various representations). It is of particular interest for linguistic analysis that various options for textual research are featured, including a thesaurus that is intended to search also for misspellings and aberrant forms.

Hispania Epigraphica Online (*HE*)[9] focuses on the epigraphic documents of Portugal and Spain, in large part written in Latin, but with a few examples of Greek, Semitic, and Iberian inscriptions. The corpus includes 30,809 inscriptions, most of which include photos. However, metadata sets of the texts, their degree of elaboration, and search options are less accurate

---

[7]   Cf. *http://www.edr-edr.it/default/index.php* [accessed on 20.02.2020].
[8]   Cf. *http://www.edb.uniba.it/* [accessed on 20.02.2020].
[9]   Cf. *http://eda-bea.es/* [accessed on 20.02.2020].

than in the above-mentioned databases, so that the search interface holds the following fields: record number, title, object type, inscription type, keyword, inscription, place of finding, place of conservation, and museum.

On the whole, the *EAGLE* network profiles as a massive epigraphic digital resource, which is based on the Metadata Aggregation System (Mannocci *et al.*, 2017: 173-174), i.e. an Aggregative Data Infrastructure (Amato *et al.*, 2013) where all the information of the four major collections illustrated above is stored and indexed. Users can browse the content and interact with it by means of an interface (Prandoni *et al.*, 2014) allowing the searching and browsing of the rich set of data made available by *EAGLE* partners by using either a free text simple search or an advanced search where the user can specify the values of a number of fields. Images can be retrieved through an image recognition algorithm, and translations of the epigraphic texts are available. Finally, it is also possible to export the *EpiDoc* document describing the object[10].

2.3. *Towards a digital epigraphy designed for linguistic research*

As shown in §§ 2.1-2.2, a wide range of digital repositories of epigraphic content are currently accessible online, featuring a great variety of Latin inscriptions, and providing scholars with a cluster of extra-linguistic data, such as provenance place, dating, material, etc. by which to verify the reliability of historical reconstructions. An accurate reconstruction of the socio-historical context is – of course – of primary interest also for the study of language variation and change in the Latin epigraphic (and, more generally, non-literary) documents. In the last few decades, the widely acknowledged dimensions of sociolinguistic variation have proven to be a fertile field of investigation, giving rise to the field of historical sociolinguistics, whose aim is «the reconstruction of the history of a given language in its socio-cultural context» (Conde-Silvestre and Hernández-Campoy, 2012: 1). In particular, many scholars have shown that it is possible to identify different varieties of

---

[10]   Thanks to the collaboration of many different scholars working on Greek and Latin inscriptions, *EpiDoc* (Epigraphic Documents; *http://sourceforge.net/p/epidoc/wiki/Home/*) has been established as a robust system for what regards the representation and the encoding of epigraphic or papyrological texts in digital form (cf. BODARD, 2010). *EpiDoc* adopts a subset of the XML defined by the TEI standard for the digital representation of texts, which is now widely used in the humanities. This flexible system allows not only the transcribing of a Greek or Latin text, but also, for instance, the encoding of its translation, description, and other pieces of information such as dating, history of the inscription, bibliography, and the object on which the text is written.

Latin by combining the investigation of diastratic (Clackson, 2011a; Adams, 2013), diatopic (Herman, 1990; Adams, 2007), diaphasic (Kruschwitz and Halla-aho, 2007; Kruschwitz, 2015; Ferri and Probert, 2010), and diamesic variation.

However, none of the corpora illustrated above allows researchers to directly access specific information about relevant linguistic variation phenomena, and they do not satisfactorily meet the needs of the linguist to study Latin epigraphic texts from a variationist perspective. In particular, as already stated above (§ 2.1), in all of them queries can be performed using a token-level keyword search by entering single words or set of words or letters, and this requires prior knowledge of what to search for. In addition, one cannot always be sure that the digital editions of the texts are free from emendations and standardizations of those aberrant forms, misspellings, and spelling variants that are of primary relevance for the linguist.

Thus, in order to systematically address the massive (ortho)graphic and linguistic variation observable in Latin inscriptions, differently designed tools are necessary. This is the reason why *CLaSSES*, while providing annotation for both extra- and meta-linguistic data (§ 4.1), also provides fine-grained linguistic information about specific spelling variants that can be regarded as clues for phonetic-phonological and morpho-phonological variation (cf. § 4.2). Another database that is designed to be a helpful tool in the study of linguistic (diatopic) variation is the Computerized Historical Linguistic Database of the Latin Inscriptions of the Imperial Age (*LLDB*)[11], a comprehensive digital resource for the Vulgar Latin inscriptions from the Roman provinces (Adamik, 2012). More than 87,800 spellings that diverge from the Classical norm are collected in *LLDB* and they are accurately classified according to a wide range of phonetico-phonological, morphological, and syntactic phenomena. Moreover, each form is richly annotated with extra- and meta-linguistic information including findspot, dating, type of inscription (e.g. Christian or non-Christian, prose or verse, private or official), type of object, comments on issues concerning the reading of the texts (e.g. presence of fractures on the object, etc.), and relevant bibliography. The search interface makes it possible to perform simple and advanced queries by combining an unlimited number of search criteria and by using Boolean operators. However, it has to be noted that this resource has been

---

[11]   Cf. *http://lldb.elte.hu/* [accessed on 20.02.2020]. The database is a revised and upgraded version of József Herman's Late Latin Data Base, hence the acronym.

designed in order to meet the requirements of Herman's (2000, for the last version) approach to the investigation of language variation. According to him, divergent spellings can be assumed as representative of diatopic variation only if their relative frequency is expressed as a percentage against the total number of other linguistically relevant divergent spellings (for an updated discussion of the methodological issues, see Tamponi, 2020: 24-26). As a consequence, in *LLDB* it is possible to elicit lists of misspelled forms, but they cannot be checked against the total amount of the corresponding Classical spellings.

2.4. *Other non-literary texts; papyri, wooden tablets, and Medieval charters*

A few last words are due for the digital editions of other non-literary texts (such as papyri and letters of correspondence) that can be a valuable source for variationist analysis. Papyri.info, is an extensive digital text collection of Greek and Latin documentary papyri dating from the 4th century BC to the 8th century AD, in large part from Egypt. The resource is based on the Papyrological Navigator (*PN*), a tool that aggregates three major databases of documentary papyri: the Duke Databank of Documentary Papyri (*DDbDP*), the Heidelberger Gesamtverzeichnis der griechischen Papyruskunden Ägyptens (*HGV*), and the University of Michigan Advanced Papyrological Information System (*APIS*)[12]. The main bibliographical database for papyrological research, the Bibliographie Papyrologique (*BP*), is also integrated. The texts, coming from the *DDbDP*, have been converted in *EpiDoc* and are now integrated and merged with metadata and images drawn from the *HGV* and the *APIS* databases. The archive currently includes 56,779 texts (in addition, there are 29,867 records with metadata only). The Navigator allows both simple and complex string-searching and the search can be refined by adding further criteria (series and collection, provenance, dating, language, etc.). Annotations of linguistic phenomena are lacking, so that the texts cannot be queried in this way, but it is worth mentioning that a corpus of Greek texts exported from papyri.info has been enriched with linguistic information as part of the SEMATIA Project (Linguistic Annotation of the Greek

---

[12]   Cf. *http://papyri.info/* [accessed on 20.02.2020]. Another important tool that is available in papyri.info, is the Papyrological Editor (*PE*), which enables users to contribute to the collection by entering new texts and metadata, or editing those already existing.

Documentary Papyri - Detecting and Determining Contact-Induced, Dialectal and Stylistic Variation) of the University of Helsinki[13]. The result is an extensively annotated corpus that enables the comparison between the misspellings and spelling variants of the scribes' original text and the standard Greek, as well as the analysis of the morpho-syntactic structures of the texts. For a corpus of linguistically annotated Latin papyri, see below (§ 3.3) the section of *CLaSSeS Egypt and Eastern Mediterranean*.

Another important collection of Latin first-hand texts is the digital publication of the ink-written wooden tablets from the Roman garrison of Vindolanda, dating between the 1st and the 3rd century AD. The documents include private correspondence, military reports, accounts, and other informal or non-literary writings. The online edition is hosted by two separate websites: *http://vindolanda.csad.ox.ac.uk* for the tablets published in Bowman and Thomas (1983) and Bowman and Thomas (1994), *http://vto2. classics.ox.ac.uk* for the tablets published in Bowman and Thomas (2003) and the earlier publications. Simple word queries can be performed by means of the 'Latin text search', while other information (subjects, categories and types of documents, people, places, military terms, archaeological context) can be accessed through the 'General text search' facility or through browsing. Every text is transcribed, translated, provided with a photo and an accurate description with particular focus on the palaeographic aspects. Specific linguistic annotation is missing also in this case, but for its implementation for 762 ink-written tablets as part of *CLaSSeS*, see § 3.2 *Roman Britain* below.

While none of the corpora illustrated in this section is specifically designed for linguistic analysis, a notable exception is the Late Latin Charter Treebanks (*LLCT*), which is developed for the research of the non-literary Latin of the Early Middle Ages (Korkiakangas, 2020, and references therein). The *LLCT* treebank is a set of three morphologically and syntactically annotated corpora (*LLCT1*, *LLCT2*, *LLCT3*), which also feature a textual annotation layer that indicates abbreviated and restored words. *LLCT1* and *LLCT2* are now completely accessible online[14]: the former includes 225,834 tokens distributed within 519 charters written in Tuscany between 714 and 869 AD; the latter includes 257,819 tokens in 521 Tuscan charters between 774 and 897 AD. *LLCT3*, under construction, is going to contain

---

    13   Cf. *https://sematia.hum.helsinki.fi*.
    14   Cf. *https://zenodo.org/record/3633607#.XjU4lSNS9EY* (for *LLCT1*) and *https://zenodo.org/record/3633614#.XjU6zCN7lEY* (for *LLCT2*).

ca. 110,400 tokens in 221 charters written in Tuscany as well as in several locations in northern and southern Italy between 721 and 1000 AD. As the lemmatization and grammatical parsing of traditional treebanks[15] is mainly based on texts of Classical authors (for an overview of Latin lemmatizers and morphological analyzers, see Celano, 2020), in *LLCT* particular attention is paid to the lemmatization and additional annotation of all those non-classical and late forms that are typical of non-literary Early Medieval Latin.

3. *Materials*

CLaSSES is structured in four different sections, whose contents are hereafter described with reference to the kind of material, dating, and area of provenance: *Rome and Italy* (§ 3.1), *Roman Britain* (§ 3.2), *Egypt and Eastern Mediterranean* (§ 3.3), and *Sardinia* (§ 3.4). These sections can be also accessed from an interactive map, which shows the number and the geographic distribution of the inscriptions included in the database. The criteria of tokenization, lemmatization, as well as those of linguistic, meta-linguistic and extra-linguistic annotation are illustrated in § 4 below.

3.1. *Rome and Italy*

The first section, *Rome and Italy*, is a collection of 1,250 Latin inscriptions (for a total number of 11,804 tokens), dating between the 6th century BC and the 1st century AD, mainly from Rome and Central Italy. The inscriptions belong to five different textual typologies (*tituli honorarii*, *tituli sepulcrales*, *instrumenta domestica*, *tituli sacri publici*, and *tituli sacri privati*; cf. § 4.1 for the criteria of classification), and their texts have been retrieved from the following editions: Lommatzsch (1918, *Hrsg.*; 1931, *Hrsg.*; 1943, *Hrsg.*), Degrassi and Krummrey (1986, eds.), Dressel (1899 [1969]), Gordon and Gordon (1958), Panciera *et al.* (1991), Degrassi (1957-1963), Wachter (1987), and Warmington (1940)[16].

---

[15]   Cf. the Latin Dependency Treebanks (*LDT*, *https://perseusdl.github.io/treebank_data/*), the *PROIEL* treebanks (*https://proiel.github.io*), and the Index Thomisticus Treebank (*IT-TB*, *https://itreebank.marginalia.it*).

[16]   Note that, among the available material, not every inscription is significant for linguistic studies. As a consequence, the following texts have been excluded: (i) legal texts, since they are generally prone to archaisms; (ii) too short (single letters, initials) or fragmentary inscriptions; (iii) inscriptions from the necropolis of Praeneste, as they contain only anthroponyms in the nominative form.

The study of spelling variants in archaic and early epigraphy is of particular relevance for the investigation of the long-lasting process of formal codification of the language that led to what is currently labelled 'Classical Latin'. The contrastive analysis between the language of these inscriptions and that which became an established standard with fixed rules and forms, is representative of the fundamental process of selection, regularization, and reduction of variation underlying the ideology of *Latinitas* "correct Latin", which was progressively elaborated by grammarians, rhetoricians, poets, and prose writers between the final decades of the Republic and the early Empire (Poli, 1999; Clackson and Horrocks, 2007: 130-182; Clackson, 2011a; 2011b; Cuzzolin and Haverling, 2009; Mancini, 2005; 2006).

In the absence of an established standard and as a consequence of specific and particular issues of single inscriptions, the texts of this period may raise problems with their reading and with the linguistic interpretation of their forms. In such cases, the numerous readings that have been proposed so far by scholars have been compared in order to guarantee the most reliable and updated philological accuracy.

3.2. *Roman Britain*

The section *Roman Britain* has, so far, an assemblage of 762 ink-written tablets (for a total number of 11,446 tokens) from the auxiliary fort Vindolanda just south of Hadrian's Wall, dating between the 1st and the 3rd century AD. The inscriptions belong to ten different textual typologies: military reports, *commeatus*, *numera*, *memorandum*, *commendatio*, male / female correspondence, *literaria*, miscellany, and *descripta* (cf. § 4.1 for the criteria of classification). For this section, the inscriptions were collected from the following corpora and online resources: Bowman and Thomas (1983; 1994; 2003), Bowman, Thomas and Tomlin (2010), Bowman, Thomas and Tomlin (2011), *http://vindolanda.csad.ox.ac.uk/*, *http://vto2.classics.ox.ac.uk/* (cf. above, § 2.4).

Since Adams (1995) the language of the Vindolanda writing-tablets has attracted the attention of scholars working on language variation and contact. On the one hand it is possible to identify different degrees of literacy between the texts written by the prefects and their scribes, and those written by other people with poorer competence, whose misspellings allow linguistic considerations (Cotugno, 2015; Cotugno and Marotta,

2017). On the other hand, in this military post Latin was used by auxiliary troops coming mainly from Gallia Belgica, i.e. Celto-Germanic people whose Latin writings may bear tell-tale signs of second-language learning (Cotugno, 2018).

3.3. *Egypt and Eastern Mediterranean*

This section has a collection of 220 documentary letters (for a total number of 9,224 tokens) written on papyri and ostraka from Africa Proconsularis, Aegyptus, Palestine, and Syria, dating between the 1st and the 6th century AD. Two different textual typologies have been distinguished on the basis of the epistolary genre: formal (i.e. public) and informal (i.e. private) letters. The documents from these areas were retrieved from the following editions: Cugusi (1992a; 1992b; 2002) and Marichal (1992).

Greek remained the lingua franca of all the eastern regions of the Empire and it was used as such also by the Romans, and the Latin-speaking population in these areas largely consisted of not locally born Latin speakers, but «mobile personnel, who would no doubt adopt 'regional' usages as they came and went» (Adams, 2003: 525). As a consequence, this corpus of documentary letters, which was the work of a variety of bilingual (and possibly bi-literate) scribes[17], is of particular interest both for the study of regional variation and for the study of linguistic and graphemic interference between Latin and Greek (Barchi, 2019).

3.4. *Sardinia*

The last section contains 1,184 inscriptions (for a total number of 14,413 tokens) from Sardinia, dating between the 1st century BC and the beginning of the 7th century AD. In line with the criteria adopted for the section *Rome and Italy*, the following textual typologies have been identified: *tituli honorarii*, *tituli sepulcrales*, *tituli sacri publici*, *tituli sacri privati*, *instrumenta domestica*; the supplementary category *military diplomas* has been added. The reference editions for the texts are Mommsen (1883, *Hrsg.*), Ihm (1899), Sotgiu (1961; 1968; 1988), Corda (1999), Floris (2005).

---

[17] Cf. the well-known case of Claudius Terentianus, illustrated (among others) in ADAMS (2003: 527-637, 741-750 and *passim*).

As Roman Sardinia was a multi-faceted community of speakers, a quantitative analysis of the surviving Latin inscriptions can provide insights into the dynamics of diatopic variation and interference. In particular, it is likely it will sustain quantitative evidence backing the traditional hypothesis that acknowledges a number of common features between African Latin and the Latin of Sardinia (Fanciullo, 1992; Lupinu, 2003; Lorenzetti and Schirru, 2010; Loporcaro, 2015: 48 ff.), as well as casting some light on the specific evolution of the Sardo-Romance varieties among the Romance languages (Tamponi, 2020).

4. *Corpus annotation*

As just described, *CLaSSES* includes 3,416 Latin documents in digitized form. As a preliminary operation for the creation of the database, all texts have been automatically tokenized, i.e. broken into a sequence of words and units of punctuation (for a total number of 46,887 tokens).

Each token of the corpus is univocally associated with a token-ID, i.e. a short string of alphanumeric characters that provide basic information: the source of the text, the number of the inscription, and the position in which the token occurs within the inscription (e.g. BTT-118-1 refers to Bowman, Thomas and Tomlin's edition of the Vindolanda writing-tablets, publication number 118, and first word of the text).

After tokenization, all words of the corpus (also abbreviated and incomplete forms that could be fully understood) have been lemmatized. This operation was conducted manually, due to the high frequency in letters and inscriptions of abbreviated, incomplete, and misspelled words that could not be easily processed by automatic tools.

Once tokenized and lemmatized, a rich linguistic, meta-linguistic, and extra-linguistic annotation has been added to the texts, as described in the following paragraphs (cf. also De Felice *et al.*, 2015). Data were recorded in a tabular form in Excel worksheets by four expert annotators (cf. Section *Acknowledgments*), who worked separately on the different subsections of *CLaSSES*. All the data collected were carefully cross-checked by other annotators and researchers involved in the project (disagreements were collaboratively discussed to reach consensus), before being converted into a database that can now be freely accessed from the *CLaSSES* website (cf. § 5).

4.1. *Extra-linguistic and meta-linguistic annotation*

Place of provenance and dating. Extra-linguistic information related to the place of provenance and dating of each document included in the database has been annotated (these data were derived from the sources from which the texts were retrieved). Places of provenance can be grouped into four main areas: Rome and peninsular Italy, Sardinia, Egypt and Eastern Mediterranean, and Roman Britain. The dating of the collected documents spans from the 6th-5th century BC of some inscriptions from Central Italy to the 5th-7th century AD of some Egyptian papyri and Sardinian texts (cf. § 3).

Text type. Each text has also been classified according to its typology. Among the epigraphic texts collected in the sections *Rome and Italy* and *Sardinia* we find *tituli honorarii* (honorary inscriptions dedicated by public figures and monumental inscriptions), *tituli sepulcrales* (commemorative inscriptions and epitaphs), *instrumenta domestica* (inscriptions on everyday objects), *tituli sacri publici* (votive inscriptions dedicated by public figures), *tituli sacri privati* (votive inscriptions dedicated by private customers), and *military diplomas* (this last category is used only for Sardinian texts, to classify personal legal documents on bronze tablets that contain a copy of imperial constitutions by which Roman citizenship and *conubium* were granted to veterans of the auxiliary army units, the fleet and the Praetorian Guard).

Vindolanda's tablets (section *Roman Britain*) may be classified as *military reports* (communications between officers regarding the activity of the garrison), *commeatus* (applications of leave to the prefect of the cohort), *memoranda* (short communications left by one garrison to the other), *commendationes* (letters of recommendations), *numera* (accounts of various types), *literaria* (writing exercises), *male/female correspondence*, *miscellany* (tablets of uncertain attribution), and *descripta* (tablets with a very faded text, for which there are doubts about their reconstruction).

Finally, the letters collected in the section *Egypt and Eastern Mediterranean* have been classified as either *formal* (i.e. public) or *informal* (i.e. private letters of information).

Most of the categories adopted for classifying the text types were derived from the original sources of the digitized texts, but, in many cases, annotators created specific labels to provide a more fine-grained classification (for instance, making a distinction, within the group of the inscriptions tra-

ditionally classified in the *CIL* as *tituli sacri*, between *tituli sacri privati* and *tituli sacri publici*; cf. also Donati, 2015).

Graphic form. The epigraphic texts, tablets, and letters collected in *CLaSSES* rarely consist of well-written and fully readable words; rather, they often present faint or missing letters betrayed by the conservation status of the support, or incomplete forms (initials, abbreviations). Therefore, each token of the corpus has been also classified according to its graphic form. For this level of annotation, we distinguish the following categories: *complete words*; *abbreviations*, for every kind of shortening (e.g. BTT-135-5 COH for COHORS), including personal name initials; *incomplete words*, for words partly integrated by editors (e.g. ILSARD-I-388-37 AURE[LIO]) or impossible to integrate (e.g. CEL-I-5-2 GLAU[); *words completely integrated by editors* (e.g. BTT-257-2 [CERIALI]); *presumed misspellings* (e.g. CEL-I-1-416 SITULUS for TITULUS); *uncertain words*, for words that cannot be interpreted, not even in their graphical form (e.g. CIL-I²-59-9 STRIANDO); *numbers*; *symbols*, only in the sections *Roman Britain* and *Sardinia*, for non-alphabetical signs (that are presented in the database not as graphic signs, but with an indication of their meaning between brackets, e.g. BTT-138-6 and EE-VIII-710-11 SYMBOL(CENTURIAE)); *lacunae*, i.e. gaps in the inscription (*lacunae* are identified by the string [...] and they are considered to be tokens, since they occupy a specific position within the texts, and they actually exist in their critical editions).

Language. Even if the documents which compose the corpus *CLaSSES* are primarily written in Latin, they sometimes include foreign words. Therefore, we distinguished Latin forms from words belonging to other languages, manually annotated as *Greek*, *Oscan*, *Umbrian*, *Etruscan*, *Iberian*, *Neo-Punic*, *Semitic*, *Coptic*, *Hebrew*, *Egyptian*, and *Persian*. Moreover, mixed forms are marked as *hybrid* (e.g. CIL-I²-553-2 ALIXENTROM, a Greek loanword in a Latin form with Etruscan phono-morphological interferences), whereas those of unknown language are marked as *unknown* (e.g. CEL-I-150-39 ATESTAS).

Author/addressee. Only for the section *Roman Britain*, containing letters from Vindolanda, did we choose to also annotate the author of the texts and his/her addressees when the identity of these persons is known. For instance, the tablet BTT-233 is written by *Cerialis* and addressed to *Aelius Brocchus*.

4.2. *Linguistic annotation of non-classical variants*

The most relevant part of the annotation process, which provides the corpus with a rich set of qualitative data, is the result of an accurate and in-depth linguistic analysis of the collected documents. The purpose of this annotation is (i) to identify non-classical variants, i.e. all words that deviate from Classical Latin from a purely (ortho-)graphic point of view (as described in § 1; see also Marotta, 2015; 2016), and (ii) to classify non-classical variants according to the kind of variation phenomenon involved. Therefore, first annotators manually identified all words that clearly do not belong to the classical literary language (e.g. DEDE instead of classical DEDIT; MENERVAI instead of classical MINERVAE) and marked them as *non-classical* (tot. 3,838, i.e. 8,2% of tokens in the four sub-sections of *CLaSSES*). Then, they associated each non-classical form with its corresponding classical form (e.g. nom. sg. CORNELIO, non-classical - CORNELIUS, classical). Finally, all non-classical variants were classified according to the type of variation phenomena that distinguish them from the corresponding classical equivalents. More precisely, such variation phenomena may regard the vowel or consonant system, as well as morpho-phonology (when variation occurs in morphological endings of words). The most relevant phenomena annotated for vowels are the following:

– vowel alternations (CIL-I²-2909-4 MENERVA for MINERVA; BTT-206-34 SENICIO for SENECIO);

– phenomena related to the notation of vowel length, such as vowel doubling (CIL-I²-365-11 VOOTUM for VOTUM), *apex* (CEL-I-8-33 SUÓ for SUO), and *I longa* (BTT-297-9 FECI for FECI);

– omission of vowel (CIL-I²-37-10 VICESMA for VICESIMA; CIL-X-7756-28 OCLOS for OCULOS) and insertion of vowel (BTT-187-15 CRISPIA for CRISPA);

– phenomena related to diphthongs (such as <E> for Classical <AE> in CEL-I-157-17 ETATIS).

The main phenomena related to consonants can be summarized as follows:

– omission of final consonant (CIL-I²-8-2 CORNELIO for CORNELIUS; CIL-X-7809-15 ANNU for ANNUM);

– omission of nasal before consonants (CEL-I-177-8 PRAESES for PRAESENS; BTT-609-39 SACTIUS for SANCTIUS);

–   assimilation (CEL-I-77-47 MASSIPIUM for MARSIPIUM);
–   double *pro* single consonant (CIL-I²-16-1 [P]AULLA for [P]AULA) and
    single *pro* double consonant (CEL-I-234-37 QUATUOR for QUATTUOR);
–   <B>/<V> confusion (CIL-X-7990-16 BIXIT for VIXIT; CIL-X-7619-11
    VENE for BENE).

Most of the categories just presented are further articulated into sub-cat-
egories, in order to allow a more fine-grained classification of variation phe-
nomena; for instance, for vowel alternations we annotated as two separate
phenomena (i) <I>, /ĭ/ = <E> and (ii) <I>, /ī/ = <E>.

If non-classical variants occur in morpho-phonological position (gen-
erally, in word endings), we also annotated the special ending attested, such
as the *-e* ending of the dative singular of the first declension (CEL-I-146-57
MEE for MEAE), the *-os* and *-o* endings of the nominative singular of the sec-
ond declension (CIL-I²-406b-2 CANOLEIOS and CIL-I²-408-2 CANOLEIO
for CANOLEIUS), the *-om* ending of the accusative singular of the second
declension (CIL-I²-403-8 LOCOM for LOCUM), or the *-et* ending of the 3rd
person of the perfect (CIL-I²-365-12 DEDET for DEDIT; CIL-X-7632-12
FECET for FECIT).

5. *Search interface*

The open-access search interface currently available on the *CLaSSES*
website (*http://classes-latin-linguistics.fileli.unipi.it*) has been specifically de-
veloped to explore the corpus, to perform queries on it, and to access the
fine-grained linguistic annotation conducted on texts.

Basic queries can be made by clicking the *Search* button from the top
menu of the website and by selecting the sub-corpus of interest: *Rome and
Italy*, *Sardinia*, *Roman Britain*, or *Egypt and Eastern Mediterranean* (docu-
ments can be also selected from the map in the *Homepage*, which shows the
geographic distribution and the number of the texts included in the data-
base). It is also possible to query the whole corpus, by selecting *Cross-corpora*.
Once the section of interest is selected and the search interface accessed, the
entire (sub-)corpus is displayed in a vertical column, with one token per row.
Most data annotated for each token are reported in multiple columns in a
tabular format: its ID (containing information about the publication num-
ber of the inscription or letter and the source from which the text is derived);

its lemma, language, and graphic form; its classification as either a classical word or a non-classical variant; the typology of the inscription or the letter which the token belongs to, its place of provenance, its dating; the author and addressee of the letters (only for the section *Roman Britain*); and the support material of the document (only for the section *Egypt and Eastern Mediterranean*).

It is possible to perform simple queries on the corpus, either by searching for a specific form (the use of 'wildcard' characters is supported), or by using and combining the filters at the top of each column (for instance, to visualize only classical or non-classical forms, to filter results per publication number, lemma, language, graphic form, and place of provenance, etc.). With the *Advanced search* functionality, users can select more than one option for a search filter (e.g. for language: Latin AND Greek AND Hybrid); most importantly, it is also possible to search for specific linguistic phenomena annotated for vowels, consonants, or morphophonology. Finally, the export options in the *Search* page allow exporting the data in different formats (CSV, Text, Excel 1995+, Excel 2007+), at any moment.

The two columns on the rightmost part of the *Search* page (*Text* and *More Info*) allow access to further information. By clicking on the two symbols present in the *Text* column, it is possible to visualize the immediate linguistic context of each form of the corpus (5 words before/after) and to read the entire text of the document. By clicking on the symbol present in the *More Info* column, a new page will open containing all data annotated for a given form: token ID, language, graphic form, lemma, classical/non-classical classification, text typology, place of provenance, dating, author, addressee, support material, linguistic context, and entire inscription; in case of non-classical form, the equivalent classical form is reported (for instance, CONSUL for non-classical COSOL). At the end of this page, the variation linguistic phenomena individuated for non-classical forms are reported.

6. *Conclusions*

In conclusion, *CLaSSES* aims at being an additional digital resource for academic scholarship which is interested in carrying out variationist studies on the non-literary documentation of the Latin language.

Of course, the database is built on a reference corpus of texts which is not on a par either with other available, extensive digital epigraphic collections, or with the existing treebanks and lemmatizers that are based on large repertoires of literary texts (many of which are described in the other papers of this special issue of *Studi e Saggi Linguistici*). A full coverage of the non-literary documents is not its purpose, after all.

Rather, the corpus is designed for the investigation of orthographic variants in non-literary Latin texts of various ages and provenance. Due to their nature, these sources allow us to draw relevant data on the phonological and morpho-phonological domains, which other available digital tools do not provide with such fine-grained annotation.

CLaSSES relies on single and coherent corpora of texts, in which the annotation of orthographic variation is systematically cross-referenced with the meta-linguistic information. Such a correlation between linguistic data and extra-linguistic variables can provide reliable clues in order to perform diachronic, diatopic, and diaphasic analyses, which may hopefully cast some further light on the sociolinguistic variation within the Latin language.

References

Adamik, B. (2012), *In search of the regional diversification of Latin: Some methodological considerations in employing the inscriptional evidence*, in Biville, F., Lhomme, M. and Vallat, D. (2012, *éds.*), *Latin vulgaire - Latin tardif IX. Actes du IXe colloque international sur le latin vulgaire et tardif, Lyon, 6-9 septembre 2009*, Maison de l'Orient et de la Méditerranée 'Jean Pouilloux', Lyon, pp. 123-139.

Adams, J.N. (1995), *The Language of the Vindolanda Writing Tablets: An Interim Report*, Cambridge University Press, Cambridge.

Adams, J.N. (2003), *Bilingualism and the Latin Language*, Cambridge University Press, Cambridge.

Adams, J.N. (2007), *The Regional Diversification of Latin 200 BC-AD 600*, Cambridge University Press, Cambridge.

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

Amato, G., Bollettieri, P., Gennaro, C., Manghi, P., Mannocci, A., Zoppi, F., Casarosa, V. and Falchi, F. (2013), *AIM infrastructure specification* [available online at *https://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D4.1_AIM_Infrastructure_Specification_update.pdf*].

Barchi, S. (2019), *On vowel prosthesis before sC in Substandard Latin and Koine Greek: a synoptic review*, in «Studi e Saggi Linguistici», 57, 2, pp. 45-82.

Biville, F., Decourt, J.C. and Rougemont, G. (2008, *éds.*), *Bilinguisme gréco-latin et épigraphie,* Maison de l'Orient et de la Méditerranée 'J. Pouilloux', Lyon.

Bodard, G. (2010), *EpiDoc: Epigraphic Documents in XML for publication and interchange*, in Feraudi-Gruénais, F. (2010, ed.), *Latin on Stone: Epigraphic Research and Electronic Archives*, Lexington Books, Lanham, pp. 101-118.

Bowman, A.K. and Thomas, J.D. (1983), *Vindolanda: The Latin Writing Tablets*, Society for the Promotion of Roman Studies, London.

Bowman, A.K. and Thomas, J.D. (1994), *The Vindolanda Writing Tablets (*Tabulae Vindolandenses *II)*, British Museum Press, London.

Bowman, A.K. and Thomas, J.D. (2003), *The Vindolanda Writing Tablets (*Tabulae Vindolandenses *III)*, British Museum Press, London.

Bowman, A.K., Thomas, J.D. and Tomlin, R.S.O. (2010), *The Vindolanda writing tablets (*Tabulae Vindolandenses *IV part 1)*, in «Britannia», 41, pp. 187-224.

Bowman, A.K., Thomas, J.D. and Tomlin, R.S.O. (2011), *The Vindolanda writing tablets (*Tabulae Vindolandenses *IV part 2)*, in «Britannia», 42, pp. 113-144.

Caldelli, M.L., Orlandi, S., Blandino, V., Chiaraluce, V., Pulcinelli, L. and Vella, A. (2014), EDR - *Effetti collaterali*, in «Scienze dell'Antichità», 20, 1, pp. 267-289.

Campanile, E. (1971), *Due studi sul latino volgare*, in «L'Italia Dialettale», 34, pp. 1-64.

Celano, G. (2020), *Lemmatization and morphological analysis for the Latin Dependency Treebank*, in «Studi e Saggi Linguistici», 58, 1, pp. 21-38.

Clackson, J. (2011a), *Latin inscriptions and documents*, in Clackson, J. (2011, ed.), *A Companion to the Latin Language*, Wiley / Blackwell, Chichester / Malden, pp. 29-39.

Clackson, J. (2011b), *The social dialects of Latin*, in Clackson, J. (2011, ed.), *A Companion to the Latin Language*, Wiley / Blackwell, Chichester / Malden, pp. 505-526.

Clackson, J. and Horrocks, G. (2007), *The Blackwell History of the Latin Language*, Blackwell, Malden.

Conde-Silvestre, N. and Hernández-Campoy, J.M. (2012), *Introduction*, in Hernández-Campoy, J.M. and Conde-Silvestre, N. (2012, eds.), *The Handbook of Historical Sociolinguistics*, Oxford University Press, Oxford, pp. 1-7.

Consani, C. (2016), *Fenomeni di contatto a livello di discorso e di sistema nella Cipro ellenistica (Kafizin) e le tendenze di "lunga durata"*, in «Linguarum Varietas», 5, pp. 51-65.

Corda, A.M. (1999), *Le iscrizioni cristiane della Sardegna anteriori al VII secolo*, Pontificio Istituto di Archeologia Cristiana, Città del Vaticano.

Cordell, R. (2015), *Reprinting, circulation, and the network author in antebellum newspapers*, in «American Literary History», 27, 3, pp. 417-445.

Cotugno, F. (2015), I longa *in iato nel* Corpus Vindolandense, in «Studi e Saggi Linguistici», 53, 2, pp. 189-206.

Cotugno, F. (2018), *The Linguistic Variation of Latin in Roman Britain*, PhD thesis, University of Pisa.

Cotugno, F. and Marotta, G. (2017), *Geminated consonants in the Vindolanda tablets. Empirical data and sociolinguistics remarks*, in Molinelli, P. (2017, ed.), *Language and Identity in Multilingual Mediterranean Settings. Challenges for Historical Sociolinguistics*, Mouton de Gruyter, Berlin, pp. 269-288.

CUGUSI, P. (1992a), *Corpus Epistularum Latinarum Papyris Ostracis Tabulis servatarum*. Vol. 1: *Textus*, Gonnelli, Firenze.

CUGUSI, P. (1992b), *Corpus Epistularum Latinarum Papyris Ostracis Tabulis servatarum*. Vol. 2: *Commentarius*, Gonnelli, Firenze.

CUGUSI, P. (2002), *Corpus Epistularum Latinarum Papyris Ostracis Tabulis servatarum*. Vol. 3: *Addenda, Corrigenda, Indices rerum, Index verborum omnium*, Gonnelli, Firenze.

CUZZOLIN, P. and HAVERLING, G. (2009), *Syntax, sociolinguistics, and literary genres*, in BALDI, P. and CUZZOLIN, P. (2009, eds.), *New Perspectives on Historical Latin Syntax: Syntax of the Sentence*, De Gruyter, Berlin / New York, pp. 19-64.

DE FELICE, I., DONATI, M. and MAROTTA, G. (2015), *CLaSSES: A new digital resource for Latin epigraphy*, in «Italian Journal of Computational Linguistics», 1, 1, pp. 119-130.

DEGRASSI, A. (1957-1963), *Inscriptiones latinae liberae rei publicae*. 2 Voll., La Nuova Italia, Firenze.

DEGRASSI, A. and KRUMMREY, J. (1986, *Hrsg.*), *CIL I² 2,4: Addenda tertia. Addenda ad inscriptiones vetustissimas*, De Gruyter, Berlin.

DICKEY, E. and CHAHOUD, A. (2010, eds.), *Colloquial and Literary Latin*, Cambridge University Press, Cambridge.

DONATI, M. (2015), *Variazione e tipologia testuale nel corpus epigrafico* CLaSSES I, in «Studi e Saggi Linguistici», 53, 2, pp. 21-38.

DRESSEL, H. (1899 [1969]), *CIL XV, 2: Inscriptiones urbis Romae Latinae. Instrumentum domesticum*, Reimer, Berlin.

ECK, W. and FUNKE, P. (2014, *Hrsg.*), *Öffentlichkeit - Monument - Text. XIV Congressus Internationalis Epigraphiae Graecae et Latinae Akten*, De Gruyter, Berlin.

ELLIOTT, T. (2015), *Epigraphy and digital resources*, in BRUUN, C. and EDMONDSON, J. (2015, eds.), *The Oxford Handbook of Roman Epigraphy*, Oxford University Press, Oxford / New York, pp. 78-85.

FANCIULLO, F. (1992), *Un capitolo della Romania submersa: il latino africano*, in KREMER, D. (1992, *éd.*), *Actes du XVIIIᵉ Congrès International de Linguistique et de Philologie Romanes (Trier 1986)*. Vol. 1, Niemeyer, Tubingen, pp. 162-187.

FERAUDI-GRUÉNAIS, F. (2010), *An inventory of the main archives of Latin inscriptions*, in FERAUDI-GRUÉNAIS, F. (2010, ed.), *Latin on Stone: Epigraphic Research and Electronic Archives*, Lexington Books, Lanham, pp. 157-160.

Ferri, R. and Probert, P. (2010), *Roman authors on colloquial language*, in Dickey, E. and Chahoud, A. (2010, eds.), *Colloquial and Literary Latin*, Cambridge University Press, Cambridge, pp. 12-41.

Floris, P. (2005), *Le iscrizioni funerarie pagane di Karales*, Edizioni AV, Cagliari.

Gordon, J. and Gordon, A. (1958), *Album of Dated Latin Inscriptions*. Vol. 1, University of California Press, Berkeley / Los Angeles.

Herman, J. (1985), *Témoignage des inscriptions latines et préhistoire des langues romanes: le cas de la Sardaigne*, in Deanović, M. (1985, *éd.*), *Mélanges de linguistique dédiés à la mémoire de Petar Skok (1881-1956)*, Jugoslavenska Akademija Znanosti i Umjetnosti, Zagreb, pp. 207-216.

Herman, J. (1990), *Du latin aux langues romanes. Études de linguistique historique*, Niemeyer, Tübingen.

Herman, J. (2000), *Differenze territoriali nel latino parlato dell'Italia: un contributo preliminare*, in Herman, J. and Marinetti, A. (2000, *a cura di*), *La preistoria dell'italiano. Atti della Tavola Rotonda di Linguistica Storica. Università Ca' Foscari di Venezia, 11-13 giugno 1998*, Niemeyer, Tübingen, pp. 123-135.

Ihm, M. (1899), *Additamenta ad Corporis vol. IX et X*, in *Ephemeris epigraphica. Corporis inscriptionum Latinarum supplementum, edita iussu Instituti archaeologici Romani*. Vol. 8, Reimer, Berlin, pp. 1-22.

Korkiakangas, T. (2020), *Theoretical and pragmatic considerations on the lemmatization of non-standard Early Medieval Latin charters*, in «Studi e Saggi Linguistici», 58, 1, pp. 67-94.

Kruschwitz, P. and Halla-aho, H. (2007), *The Pompeian wall inscriptions and the Latin language: a critical reappraisal*, in «Arctos», 41, pp. 31-49.

Kruschwitz, P. (2015), *Linguistic variation, language change, and Latin inscriptions,* in Bruun, C. and Edmondson, J. (2015, eds.), *The Oxford Handbook of Roman Epigraphy*, Oxford University Press, Oxford / New York, pp. 721-743.

Lommatzsch, E. (1918, *Hrsg.*), *CIL I² 2,1: Inscriptiones vetustissimae*, De Gruyter, Berlin.

Lommatzsch, E. (1931, *Hrsg.*), *CIL I² 2,2: Addenda Nummi Indices. Addenda ad inscriptiones vetustissimas*, De Gruyter, Berlin.

Lommatzsch, E. (1943, *Hrsg.*), *CIL I² 2,3: Addenda altera Indices. Addenda ad inscriptiones vetustissimas*, De Gruyter, Berlin.

Loporcaro, M. (2015), *Vowel Length from Latin to Romance*, Oxford University Press, Oxford.

Lorenzetti, L. and Schirru, G. (2010), *Un indizio della conservazione di /k/ dinanzi a vocale anteriore nell'epigrafia cristiana di Tripolitania*, in Tantillo, I. and Bigi, F. (2010, *a cura di*), *Leptis Magna. Una città e le sue iscrizioni in epoca tardoromana*, Edizioni dell'Università degli studi di Cassino, Cassino, pp. 303-311.

Lupinu, G. (2003), *Tra latino epigrafico e sardoromanzo: sulla datazione di alcuni sviluppi fonetici*, in «Verbum», 5, 1, pp. 59-68.

Mancini, M. (2005), *La formazione del neostandard latino: il caso delle* differentiae uerborum, in Kiss, S., Mondin, L. and Salvi, G. (2005, *éds.*), *Latin et langues romanes. Études linguistiques offertes à J. Herman à l'occasion de son 80ème anniversaire*, Niemeyer, Tübingen, pp. 137-155.

Mancini, M. (2006), Dilatandis litteris*: uno studio su Cicerone e la pronunzia 'rustica'*, in Bombi, R., Cifoletti, G., Fusco, F., Innocente, L. and Orioles, V. (2006, *a cura di*), *Studi linguistici in onore di Roberto Gusmani*, Edizioni dell'Orso, Alessandria, pp. 1023-1046.

Mannocci, A., Casarosa, V., Manghi, P. and Zoppi, P. (2017), *The* EAGLE *data aggregator: Data quality monitoring*, in Orlandi, S., Santucci, R., Mambrini, F. and Liuzzo, P.M. (2017, eds.), *Digital and Traditional Epigraphy in Context. Proceedings of the EAGLE 2016 International Conference*, Sapienza Università Editrice, Roma, pp. 161-172.

Marichal, R. (1992), *Les ostraca de Bu Njem*, Grande Jamahira Arabe, Libyenne, Populaire et Socialiste, Département des Antiquités, Tripoli.

Marotta, G. (2015), *Talking stones. Phonology in Latin inscriptions*, in «Studi e Saggi Linguistici», 53, 2, pp. 39-63.

Marotta, G. (2016), *Sociolinguistica storica ed epigrafia latina. Il corpus* CLaSSES I, in «Linguarum Varietas», 5, pp. 145-159.

Molinelli, P. (2006), *Per una sociolinguistica del latino*, in Arias Abellán, C. (2006, *éd.*), *Latin vulgaire - Latin tardif VII. Actes du VIIᵉ colloque international sur le latin vulgaire et tardif*, Secretariado de Publicaciones Universidad de Sevilla, Sevilla, pp. 463-474.

Mommsen, Th. (1883, *Hrsg.*), *CIL X 1: Inscriptiones Bruttiorum, Lucaniae, Campaniae, Siciliae, Sardiniae latinae. Pars posterior inscriptiones Siciliae et Sardiniae comprehendens*, Reimer, Berlin.

Orlandi, S. (2017), *Il progetto* EAGLE. *Europeana network of Ancient Greek and Latin Epigraphy e le sue molteplici sfide*, in Mastandrea, P. (2017, *a cura di*), *Strumenti digitali e collaborativi per le Scienze dell'Antichità*, Edizioni Ca' Foscari, Venezia, pp. 49-60.

ORLANDI, S., SANTUCCI, R., MAMBRINI, F. and LIUZZO, P.M. (2017, eds.), *Digital and Traditional Epigraphy in Context. Proceedings of the* EAGLE *2016 International Conference*, Sapienza Università Editrice, Roma.

PANCIERA, S. (2013), *Notizie da* EAGLE, in «Epigraphica», 75, pp. 502-506.

PANCIERA, S. *et al.* (1991), *Epigrafia. Actes du Colloque international d'épigraphie latine en mémoire de Attilio Degrassi pour le centenaire de sa naissance*, Università di Roma 'La Sapienza' / École Française de Rome, Roma.

POLI, D. (1999), *Il latino tra formalizzazione e pluralità*, in POCCETTI, P., POLI, D. and SANTINI, C. (1999, *a cura di*), *Una storia della lingua latina. Formazione, usi, comunicazione*, Carocci, Roma, pp. 377-431.

PRANDONI, C., CASAROSA, V. and ALFARANO, N. (2014), EAGLE *Portal* [available online at *https://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D5.2_EAGLE-Portal_v1.0.pdf*].

PRANDONI, C., FRESA, A., ALFARANO, N., AMATO, G., ZOPPI, F., MANNOCCI, A., LIUZZO, P.M., MAMBRINI, F., ORLANDI, S. and SANTUCCI, R. (2017), *Searching inscriptions through the* EAGLE *Portal*, in ORLANDI, S., SANTUCCI, R., MAMBRINI, F. and LIUZZO, P.M. (2017, eds.), *Digital and Traditional Epigraphy in Context. Proceedings of the* EAGLE *2016 International Conference*, Sapienza Università Editrice, Roma, pp. 173-185.

ROCCO, A. (2017), *EDB 2.0. How Eagle Europeana project improved the Epigraphic Database Bari*, in ORLANDI, S., SANTUCCI, R., MAMBRINI, F. and LIUZZO, P.M. (2017, eds.), *Digital and Traditional Epigraphy in Context. Proceedings of the* EAGLE *2016 International Conference*, Sapienza Università Editrice, Roma, pp. 115-130.

ROCHETTE, B. (1997), *Le latin dans le monde grec*, Latomus, Bruxelles.

ROVAI, F. (2015), *Notes on the inscriptions of Delos. The Greek transliteration of Latin names*, in «Studi e Saggi Linguistici», 53, 2, pp. 163-185.

SOTGIU, G. (1961), *Iscrizioni latine della Sardegna. Supplemento al Corpus Inscriptionum Latinarum, X e all'Ephemeris Epigraphica, VIII)*. Vol. 1, CEDAM, Padova.

SOTGIU, G. (1968), *Iscrizioni latine della Sardegna*. Vol. 2, 1: *Instrumentum domesticum. Lucerne*, CEDAM, Padova.

SOTGIU, G. (1988), *L'epigrafia latina in Sardegna dopo il C.I.L. X e l'E.E. VIII*, in TEMPORINI, H. and HAASE, W. (1988, *Hrsg.*), *Aufstieg und Niedergang der römischen Welt (*ANRW*)*. Vol. 2: *Principat*, 11.1, De Gruyter, Berlin / New York, pp. 552-739.

TAMPONI, L. (2020), *Sardinian Latin through Inscriptions: A Variationist Analysis*, PhD thesis, University of Pisa.

VINEIS, E. (1984), *Problemi di ricostruzione della fonologia del latino volgare*, in VINEIS, E. (1984, *a cura di*), *Latino volgare, latino medioevale, lingue romanze*, Giardini, Pisa, pp. 45-62.

VINEIS, E. (1993), *Preliminari per una storia (e una grammatica) del latino parlato*, in STOLZ, F., DEBRUNNER, A. and SCHMIDT, W.P. (1993, *a cura di*), *Storia della lingua latina*, Pàtron, Bologna, pp. xxxvii-lviii.

WACHTER, R. (1987), *Altlateinische Inschriften. Sprachliche und epigraphische Untersuchungen zu den Dokumenten bis etwa 150 v. Chr.*, Peter Lang, Bern / Frankfurt am Main / New York / Paris.

WARMINGTON, E.H. (1940), *Remains of Old Latin*. Vol. 4: *Archaic Inscriptions*, Harvard University Press / Heinemann, Cambridge, MA / London.

GIOVANNA MAROTTA
Dipartimento di Filologia, Letteratura e Linguistica
Università di Pisa
Via Santa Maria 36
56126 Pisa (Italy)
giovanna.marotta@unipi.it

FRANCESCO ROVAI
Dipartimento di Filologia, Letteratura e Linguistica
Università di Pisa
Via Santa Maria 36
56126 Pisa (Italy)
francesco.rovai@unipi.it

IRENE DE FELICE
Dipartimento di Lingue e Culture Moderne
Università di Genova
Piazza Santa Sabina 2
16124 Genova (Italy)
irene.defelice@edu.unige.it

LUCIA TAMPONI
Dipartimento di Filologia, Letteratura e Linguistica
Università di Pisa
Via Santa Maria 36
56126 Pisa (Italy)
lucia.tamponi@fileli.unipi.it