Creating the First Digital *Handbook of Latin Phonetics*: Between Linguistics, Digital Humanities and Language Teaching

Tommaso Spinelli

Abstract: This article discusses the creation of an innovative e-learning resource that provides a unique breadth of frequency, grammatical, and phonetic information on both Classical and Ecclesiastical Latin. Designed to bridge teaching and research, this new digital toolkit, which is available as both an online program and an Android mobile app, provides a frequency list of the most common Latin lemmas, as well as phonetic and grammatical information, including their syllabication, accentuation, and Classical and Ecclesiastical phonetic transcription according to the standards of the International Phonetic Alphabet. After providing a concise overview of the different ways in which Latin was and still is pronounced, this article will discuss the methodological and practical issues faced by the creation of the toolkit from the choice of an effective lemmatizing technique for identifying and categorizing inflected word-forms, to the creation of algorithms to accentuate Latin lemmas and transcribe Latin sounds (potentially involving multiple characters of the Latin alphabet) into IPA characters. In so doing, it will offer insights into the technologies used to maximize the impact of this new e-learning resource on teaching and research.

Introduction

This article discusses the recent creation of the first online *Handbook of Latin Phonetics*, an innovative opensource digital toolkit that provides a unique breadth of frequency, grammatical, and phonetic information on both Classical and Ecclesiastical Latin. Originally conceived within the award-winning project *Latine Loquamur* (undertaken to support the reform of classical language teaching at the University of St Andrews), the digital toolkit described in this article was developed at the *Pontificium Institutum Altioris Latinitatis* (Pontifical Salesian University of Rome), to meet the needs of the ever-increasing number of scholars and students who study Latin in Latin, or who focus on late-antique texts.[1] Accordingly, the *Handbook of Latin Phonetics* toolkit currently provides, for the first time ever, a frequency

---

[1] The *Latine Loquamur Project* was designed by Tommaso Spinelli in collaboration with Alice König, Giuseppe Pezzini and Giacomo Fenzi, and was awarded funding by the Teaching Development Office of the University of St Andrews in 2018. This project involved the creation of an online *Dictionary of Latin Synonyms*, which was published by the University of St Andrews in December 2018 (https://doi.org/10.17630/3cf644e6-86b8-44d0-a50a-b33c7ca86072; last access 17.10.2020), and of other e-learning resources (e.g. Moodle presentations, exercises, and interactive games) for the study of Latin that will be discussed in another article for reasons of space. The *Handbook of Latin Phonetics* presented in this article is available as both an app and a program: the app was developed by Tommaso Spinelli during his Postdoc at the Pontifical Salesian University of Rome in collaboration with Cleto Pavanetto, Giacomo Fenzi, Kamil Kolosowski, and Jan Rybojad, and was published – thanks to the collaboration of Miran Sajovic – by the Pontifical Salesian University of Rome in 2020. (https://play.google.com/store/apps/details?id=com.kolosowski.latinhandbook; https://doi.org/10.17630/19ce37ba-2d35-4920-bd7f-6287977de369; last access 17.10.2020). The online version of the *Handbook of Latin Phonetics*, which was developed by Tommaso Spinelli with the informatic assistance of Giacomo Fenzi at the University of St Andrews, is currently hosted in the GitHub repository of the Latine Loquamur Project (https://github.com/latineloquamur?tab=repositories; last access 17.10.2020) and can be found in the folder titled Latineloquamur-toolkit-IPA-transcriber-and-App. In the same repository users can find also the Dictionary of Latin Synonyms, which is not discussed in this article, and the link to download its app (https://github.com/latineloquamur/dictionary-of-latin-near-synonyms; last access 17.10.2020).

list of the 6500 most common Latin lemmas as attested in the entire extant corpus of Latin literature, as well as unique phonetic and grammatical information, including their syllabication, accentuation, and Classical and Ecclesiastical phonetic transcription according to the standards of the International Phonetic Alphabet.

This toolkit, which is available as both a RUST program (referred to as *Latineloquamur-toolkit-IPA-transcriber-and-App* in GitHub) and an Android mobile app (titled Handbook of Latin Phonetics), faced significant methodological and practical issues during its creation and development, such as the choice of an effective lemmatizing technique for identifying and categorizing inflected word-forms, the creation of algorithms to accentuate Latin lemmas, and the development of an innovative program to transcribe Latin sounds (potentially involving multiple characters of the Latin alphabet) into IPA characters corresponding to different pronunciations of Latin.[2] After providing a concise overview of the different ways in which Latin was and still is pronounced, this article will discuss the complex interaction between linguistics, phonology, and digital humanities. It will explore the methodologies and principal technologies used within this digital project to offer rigorous frequency and phonological information on Latin lemmas, and to maximize its impact on teaching and research.

Pronouncing Latin: between teaching and research

One of the aims of the *Latin Phonetics* digital toolkit is to further the creation of a shared rigorous methodology for the pronunciation of Latin lemmas, and for the identification of the words most used by the Latin authors that a given student or researcher might want to prioritize in their studies. Both 'frequency' and 'pronunciation' have played a key role in language teaching and rhetorical studies since antiquity. Latin authors such as Cicero, Varro, and Quintilian often referred to the *usus* (use) of a word or to its frequency in their literary, grammatical, and stylistic discussions.[3] Similarly, the pseudo-Cicero's *Rhetorica ad Herennium* devotes an entire section to the role of pronunciation in 'delivering' a speech (3.19.1–2), which is also discussed by Quintilian in his *Institutio Oratoria* (1.4; 1.7), while, in the third century CE, the grammarian Probus encourages his students to pronounce correctly the words *speculum* (mirror) and *columna* (column), avoiding the wrong forms *speclum* and *colomna*.[4] And yet, despite the importance of such themes, not enough attention has been paid to them by modern digital scholarship. While the last couple of decades have seen the publication of many frequency dictionaries for modern languages, no comprehensive frequency dictionary yet exists for Latin, as the few modern attempts to provide rigorous lemmatization and counts of Latin words have treated very limited textual corpora, and have adopted remarkably different methodologies, as we shall see better in the following analysis.[5]

Even more problematic is the situation concerning the pronunciation of Latin. Ancient literary and documentary sources indicate that Latin was spoken differently synchronically at different stages of

---

2   The two different names of the program and the app are due to the different stages of the development of the toolkit and to the different institutions that published those tools, the University of St Andrews and the Pontifical Salesian University of Rome respectively. However, to avoid confusion in this article I will refer to these tools as the *Latin Phonetics* app/program.

3   Joseph Denooz (2010), 1–2 has shown that the word *usus* ('use') is used to explain linguistic facts 45 times by Varro in his *De lingua Latina*, 163 times by Cicero in the *De Oratore* and the *Orator*, and 163 times in Quintilian's *Institutio Oratoria*. Moreover, Quintilian uses the adjective *frequens* ('frequent') and the adverb *frequenter* ('frequently') some 223 times in his linguistic and stylistic considerations. Cf. also Cic. *De Inv*. 1.9.4; 1.9.10; *De Or*. 3.140.4.

4   The so-called list of the '*appendix Probi*' has been variously dated to the third or the fifth century CE. See Barnett (2006), 257–278.

5   See, for instance Diederich (1939), Delatte/Evrard/Govaerts/Denooz (1981), Denooz (2010).

Roman history and in different regions of the empire by different social classes. For instance, Lucilius jokes about the rustic pronunciation of a certain *Caecilius*, who was *praetor urbanus* (urban pretor), by saying, in a phonetic spelling, '*Cecilius pretor ne rusticus fiat*' (Let Cecilius not be a rustic pretor; Lucil. 1130, M.), remarking on the fact that, as we know from Varro (L. 5.97), the diphthong ae was already pronounced /e/ in the countryside in Classical times.[6] Epigraphs show the existence of different pronunciations of Latin throughout the history of Rome, and the *Historia Augusta* (*Hadr*. 3.1) recounts that the emperor Hadrian (117–38 CE) was mocked for his Hispanic accent.[7] This ancient diversity has been only partially reduced in modern times; it has therefore been an urgent and challenging necessity to create a tool able to provide a standardized pronunciation of Latin.[8]

Although the first *Congrès International Pour le Latin Vivant* (the first international conference for living Latin), held in Avignon in 1956, tried to foster a shared Classical pronunciation of Latin in modern times, at least three different ways of reading Latin are still commonly – and often unthinkingly – used by different institutions.[9] The first way is the so-called 'national' because of its proximity to the phonetic system of the modern languages of the countries in which Latin is read.[10] According to this pronunciation, for example, the lemma *Caesar*, which was pronounced /ˈkae̯.sar/ in Classical Latin and /ˈtʃɛ.sar/ in Ecclesiastical, is read as /ˈtʃɛ.ˈsar/ in Italy, /ʃɛ.ˈsar/ in France, /ˈsɪ.sar/ in Britain, and /ˈtsɛ.sar/ in Germany. The second way is the so-called 'Ecclesiastical' because it is officially used by the Catholic Church. Although it looks similar to the Italian pronunciation of Latin, this pronunciation is supranational and reflects the diction of Latin used in Rome during the fourth and fifth centuries CE.[11] The third way is the so-called 'Classical' pronunciation or '*restituta*'. Starting from the Renaissance period, this system of pronunciation used the phonetic clues provided by ancient grammatical texts and epigraphs to reconstruct the language arguably spoken by cultured Romans in the first century BCE and the first century CE.[12] A further complicating factor is that, while an ever-increasing number of institutions worldwide has started to teach Latin in Latin, using the Ørberg's and Cambridge's textbooks that encourage a more active use of the language in its 'Classical' pronunciation, other world-leading institutions (such as the Salesian University of Rome and the *Pontificium Institutum Altioris Latinitatis*) have continued to use the Ecclesiastical pronunciation that is also used to read late-antique and early-medieval texts, to which Classicists have increasingly shifted their attention in the last two decades.[13]

At this critical juncture, my new toolkit builds upon recent developments in the fields of digital humanities and Latin linguistics to provide students worldwide with a rigorous guide to the pronunciation of both Classical and Ecclesiastical Latin. In particular, while the Latin dictionaries currently available in many countries tend to provide only the quantity (or length) of the penultimate syllable of lemmas, the *Latin Phonetics* program and app provide more complete information on the accentuation, prosody, syllabication, and IPA phonetic transcription of Latin lemmas. The following analysis will explore the

---

6   See Ramage (1963), 390–414.

7   See, for example, the commonly attested form coss. for *consules*, or the names *Crescentsianus* and *Vincentza* respectively attested in *CIL* XIV, 246; VII, 216. On dialectal pronunciations of Latin see Oniga (2003), 39–62.

8   On the Church's use of Latin see the epistle *Romani Sermonis* by Paulus VI (1976). On the use of Latin in modern academia see Short/George (2013). On the bidirectional influence of national languages on Latin, see Serianni (1998), 27–45.

9   See Allen (1966), 102; Pavanetto (2009), 9–10; Traina/Bernardi-Perini (1998), 22–29.

10   See Collins (2012).

11   An overview of the most important features of Ecclesiastical Latin is provided by Collins (1988).

12   See, for instance, Erasmus' *De recta Latini Graecique sermonis pronuntiatione* (1528). On this theme see also Allen (1966); Oniga (2014).

13   See the overview provided by Chiesa (2012) and Spinazzè (2014), but also the seminal work on stylometry of the 'Quantitative Criticism Lab' (https://www.qcrit.org/researchdetail/kHXib8DissfMp53Yx; last access 17.10.2020). See also Harrington/Pucci (1997), Avitus (2018), and Norberg (1999).

methodology and innovative technologies used to create this tool, discussing the potential and the limits of the digital technologies that can be currently deployed to process the Latin language.

Existing technologies

The *Latin Phonetics* toolkit builds on and bridges together several different technologies developed in recent years to meet the new need for more complete phonetical information which can be used not only to speak and write in Latin, but also to study rhythmic prose texts and the style of authors. Leaving aside the work-in-progress Latin dictionary on Wikipedia that unsystematically offers some information on the pronunciation of Latin words, the best-equipped tool currently available is that offered by the Classical Language Toolkit Project (CLTK).[14] This international opensource project offers both a 'macronizer' and a 'phonetic transcriber'. Based on an original algorithm developed by Johan Winge in 2015, the macronizer can mark Latin vowels according to their length, using a POS tagger which matches words with the lexical entries of Morpheus.[15] Although this tool does not provide accentuation of lemmas and has an accuracy of around 86.3% (depending on which of the three available POS is used), it allows a more complete prosodic mark-up than that usually offered by traditional dictionaries.[16] Moreover, the 'phonetic transcriber' represents the first attempt to provide a rigorous phonetic transcription of the Latin language according to the IPA standards.[17] This tool transliterates Latin lemmas into their phonetic forms using a list of replacements based on Allen's reconstruction of the phonetics of Classical Latin (1966). However, while the source codes of both these tools are available on Github, they do not offer a user-friendly interface, so that only expert users, with a good knowledge of programming, can actually use them to process Latin words. Moreover, the CLTK phonetic transcriber only provides information on the Classical pronunciation of Latin. Similarly, the project *LatinWordnet2.0*, which is being developed at the University of Exeter by William Short, provides the Classical phonetic transcription for Latin lemmas, but this data is currently accessible only to expert users.[18] Although they do not provide a phonetic transcription of Latin, it is worth mentioning other programs which have tried to address similar issues. The first is *Google Translate*, which now offers the Ecclesiastical pronunciation (but not accentuation and IPA transcription) of Latin words.[19] The second is *Collatinus*, which was developed within the project *Biblissima* for the study of medieval and modern texts, and can divide by syllable and accentuate Latin lemmas or small texts.[20] The third is the *Quantitative Criticism Lab* that, while not providing the pronunciation of Latin lemmas, offers detailed information on the prosody of single words and entire texts, using quantitative metrics to support both linguistic and stylistic analysis.[21] Similar are the projects *Cursus in Clausula*, developed at the University of Udine, which detects the quantitative and tonic

---

14   On the dictionary offered by Wikipedia see: https://en.wiktionary.org/wiki/Wiktionary:Main_Page (last access 26.10.2020). The CLTK project is a Python library containing tools for the natural language processing (NLP) of ancient Eurasian languages: http://cltk.org/ (last access 02.09.2020).

15   The algorithm and its explanation are available at https://cl.lingfil.uu.se/exarb/arch/winge2015.pdf (last access 02.09.2020); Morpheus is a morphological parsing and lemmatizing tool integrated into the Perseus Project http://www.perseus.tufts.edu/hopper/ (last access 02.09.2020).

16   The Python macronizer is available at https://github.com/cltk/cltk/blob/master/cltk/prosody/latin/macronizer.py (last access 02.09.2020).

17   The CLTK transcriber can be accessed at https://github.com/cltk/cltk/blob/master/cltk/phonology/latin/transcription.py (last access 02.09.2020).

18   Cf. https://github.com/wmshort/latinwordnet-archive; https://latinwordnet.exeter.ac.uk/ (last access 02.09.2020).

19   See https://translate.google.com/?sl=la#view=home&op=translate&sl=la&tl=en&text (last access 02.09.2020).

20   The codes are available at https://github.com/biblissima/collatinus (last access 02.09.2020). See also https://projet.biblissima.fr/ (last access 02.09.2020).

21   See https://www.qcrit.org/ (last access 02.09.2020).

rhythm of prose *clausulae*, and the toolkit of *Pedecerto* that, developed within the project *FIRB Traditio Partum* by the University Ca' Foscari of Venice, can perform automatic scansion of Latin verses.[22]

The first app of Latin phonetics: outline and features

Distinct from the previous contributions described above, the *Handbook of Latin Phonetics* has been designed to bring together teaching and research. Accordingly, it aims to advance the automated processing of the Latin language through the creation of original algorithms for a rigorous phonetic transcription of both Classical and Ecclesiastical Latin. It also aims to provide students, teachers, and researchers across the world with a compact, freely accessible, and easy-to-use toolkit to study Latin in Latin. For this reason, the *Handbook of Latin Phonetics* has been made available both as an online opensource program for expert users (discussed in detail in the following section) and as a user-friendly Android app (discussed in this section) which, developed in collaboration with Kamil Kolosowski, displays and provides the most important features of the toolkit in an accessible format.
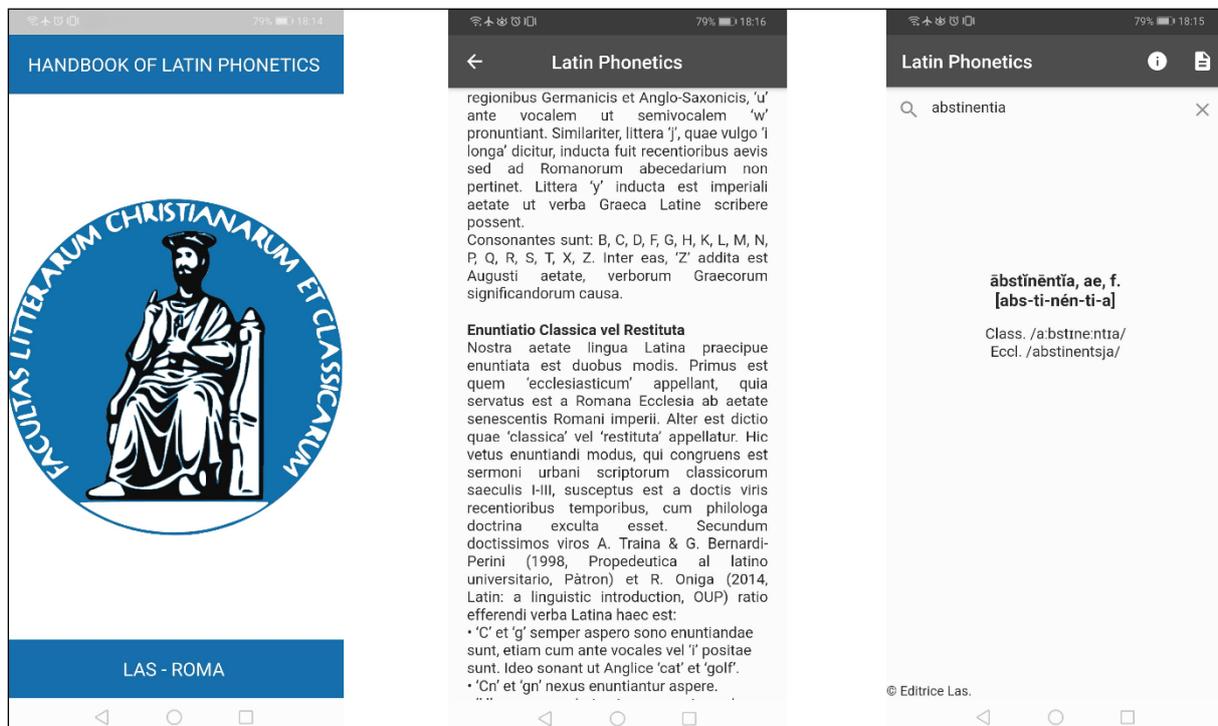

Fig. 1 Latin Phonetics App.

---

22   See respectively http://cursusinclausula.uniud.it/public/ (last access 02.09.2020) and http://www.pedecerto.eu/public/ (last access 02.09.2020).

The Android app, freely available on Google Play, is organized as follows: 1) a learning section, containing an introduction to Latin phonetics and a list of the most frequently attested Latin lemmas, 2) a search tool offering a wide range of prosodic and phonetic information on Latin lemmas, and 3) an 'info' section providing details on the app and its related programs.[23]

Latin Phonetics

DE RATIONE EFFERENDI VERBA LATINA
[Tommaso Spinelli]

Sermo Latinus est lingua flexiva Indoeuropaea, qua pars incolarum paeninsulae Italicae usi sunt circa ex sexto saeculo ante Christum natum. Etiamsi lingua mortua appelari solet, usus huius sermonis usque ad tempora hodierna continuo ab antiquitate extenditur. Italia Langobardis capta et Romano Imperio Occidentali everso, sermo Latinus mansit ut lingua publica academicaque multarum civitatum Europaearum et lingua quotidiana liturgicaque Ecclesiae Catholicae Romanae. Propter hunc diuturnum usum per totum orbem terrarum, lingua Latina pronuntiata est varie, secundum usus nationum et diversarum institutionum.

Abecedarium
Litterae apud Romanos viginti tres sunt. Littera quae per se vocem edit vocalis dicitur. Consonans est littera quae non per se sed una cum vocali sonat.
Vocales sunt: A, E, I, O, U, (Y). Inter eas distinguuntur vocales breves (ă, ĕ, ĭ, ŏ, ŭ) et longae (ā, ē, ī, ō, ū). Prisca aetate abecedarium Latinum utebatur littera grandi 'V' et littera minuta 'u' ad eundem rendendum sonum (exempli gratia 'VINUM' et 'uinum'). Recentiore distinctione inter 'u' et 'v' inducta, nunc litteram 'v' ut consonantem et 'u' ut

Fig. 2 The app's introductory section.

The first page of the app is an introductory section that, divided in two parts, contains material for the independent e-learning of Latin. The first part, titled *De Ratione Efferendi Verba Latina* offers a brief history of the Latin language, basic notions of linguistics and phonetics, and an up-to-date explanation (written in Latin) of the main differences between Classical and Ecclesiastical pronunciations. This explanation deals especially with the differences in sound between the diphthongs *ae* and *oe* (which are pronounced as monophthongs in Ecclesiastical Latin), and with the pronunciation of velar and voiced plosives (*c*, *g*), which are never soft in Classical Latin, and of the group '-*ti* + vowel', generally pronounced /tɪ/ in Classical Latin and /tsi/ in Ecclesiastical.[24] Different from many modern grammars and digital programs based on Allen's Vox Latina (1966) for Classical Latin, and Nunn's Introduction to Ecclesiastical Latin (1927), this explanation builds on more recent studies such as those on Classical Latin by Traina/Bernardi-Perini (1998) and Oniga (2014), and those on Ecclesiastical Latin by Collins (1998) and especially Pavanetto (2009), who was the head of the Pontifical Institute Latinitas governing the Catholic Church's official use of Latin. This approach governs the phonetic transcription performed by the program, which is summarized in the following table (table 1), especially concerning the sounding of diphthongs in Classical Latin. In this respect, the development of the so-called historical and generative grammar over the past century has revealed that the diphthongs *ae* and *oe* evolved from the older forms *ai* and *oi*, which left a mark on the spelling used on some epigraphs composed before the end of the second century CE. For instance, in the inscription adorning the tomb of Scipio Barbatus, dated around the 250 BCE, the term ***aedilis*** ('aedile', the censor aedilis was an elected officer responsible for the maintenance of public buildings) is spelled ***aidilis*** (*CIL* 06, 01287).

23 The app can be downloaded at https://play.google.com/store/apps/details?id=com.kolosowski.latinhandbook (last access 15.10.2020). The online program is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App (last access 02.09.2020). The constitutive elements of the app can be inspected through this link: https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/Mobile%20APP (last access 02.09.2020).

24 The introduction is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/Mobile%20APP/Intro_App.txt (last access 02.09.2020).

Fig. 3 The sarcophagus of Scipio Barbatus, currently displayed in the Vatican Museums.

In another epigraph composed some thirty years later, the word *praefectura* ('prefecture') is written *praifectura* (*CIL* 10, 06231).[25] Matching the phonetic spellings of these inscriptions with the general linguistic tendency of diphthongs to take as second element a semi-consonantal sound 'i' or 'u', a past generation of scholars suggested reading classical texts by pronouncing the diphthongs '*ae*' and '*oe*' as /aj/ and /oj/, like the English sounds of 'high' and 'boy.'[26] However, scholars have more recently pointed out that these pronunciations might simply reflect an archaic transition between the diphthong ai and ae, since the canonical form *aedem* is already attested in the famous text of the so-called *senatus consultum de bacchanalibus*, written in 186 BCE.[27] Thus, while maintaining that the second element of a diphthong is always an asyllabic vowel that cannot be stressed, modern scholarship has suggested that, in Classical Latin, "the pronunciation of the diphthongs *ae* and *oe* is [ae] and [oe] respectively."[28] The introductory section of the app presents the results of these studies in the form of simple Latin rules, often comparing the sounds of Latin with that of modern languages.

LATIN ALPHABET	CLASSICAL PRONOUNCIATION	ECCLESIASTICAL PRONOUNCIATION
ā	a:	a
ă	a	a
(ae)	ae̯	ɛ
b	b	b
c	k	k
c + e, i, y, ae, oe	k	tʃ
ch	kʰ	k
d	d	d
ē	e:	e

---

25   *CIL* is an acronym standing for *Corpus Inscriptionum Latinarum*: this work contains a comprehensive collection of ancient Latin inscriptions.

26   See Allen (1966), 131–32.

27   See Cupaiuolo (1991), 77–87.

28   Quotation from Oniga (2014), 22. See also Cupaiuolo (1991), 86–87 and Traina/Bernardi-Perini (1998), 66.

Spinelli: Creating the First Digital *Handbook of Latin Phonetics*   DCO 7 (2021)   **7**

ĕ	ɛ	e
f	f	f
g	g	g
g + e, i, y, ae, oe	g	dʒ
gn	ŋn	ɲ
h	h	-
ĭ	ɪ	i
ī	iː	i
i (semiconsonant)	j	j
k	k	k
l	l	l
m	m	m
n	n	n
ŏ	ɔ	o
ō	oː	o
(oe)	oe̯	e
p	p	p
ph	pʰ	f
q	kʷ	kw
r	r	r
s	s	s
t	t	t
th	tʰ	t
ŭ	ʊ	u
ū	uː	u
u (semiconsonant)	w	v
x	ks	ks
z	z	dz

Tab. 1 Latin phonetic transcription.

The second part of the app's introductory section contains a list of the most common Latin lemmas which are crucial for students in their vocabulary-learning. Ideally, it would have been nice to have a dedicated section for this frequency list. However, toolbars of mobile apps tend to be very limited in terms of space. Therefore, we decided to place the list of Latin lemmas in the introductory section, after the explanation of Latin phonetics. While the online program (in GitHub) can virtually scan every Latin lemma, the app offers a selection of the most common 6500 Latin words as attested across a wide corpus of Classical and Christian texts dating from the fourth century BCE to the sixth century CE.[29] Following a growing scholarly consensus that frequency information plays a key role not only in computational linguistics but also in literary and intertextual research, and in language teaching, the last two

---

29   The online Python transcriber can be accessed at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/CLASSICAL%26ECCLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER (last access 02.09.2020). The App frequency list is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/Mobile%20APP/INFO_LIST.txt (last access 02.09.2020).

decades have seen the publication of many frequency dictionaries for modern languages.[30] Yet, while Latin authors themselves often referred to the frequency or *usus* of Latin words in their commentaries, no comprehensive Latin frequency dictionary exists.[31] The few modern attempts to provide a rigorous lemmatization and count of Latin words have always adopted limited textual corpora based on 'highly representative' authors from the so-called 'golden literature.'[32] These dictionaries consequently struggle to meet the needs of contemporary students and researchers, who are increasingly shifting their attention to the 'less famous' literature of the early Republican, Christian, and Late Antique periods. By contrast, the frequency list provided by the app is based on a wide corpus of 307 Classical and Christian authors, which has been analyzed using original algorithms and the capabilities of the new lemmatizer *Lemlat* to provide a realistic picture of the most common terms used in Latin texts of different periods.[33] Since average cultured speakers of a language know around twenty thousand lemmas, and use only a few thousand of them in their daily life, our 6500-word list provides students not only with basic lemmas, but also with the most important technical and specific words most commonly attested in Latin literature.[34] At the same time, the reasonably small size of the corpus makes it possible for the information provided to be manually checked, and for the app to work even offline.

The section 'search' contains the most important contribution offered by the app: the phonetic transcription of Latin lemmas in both Classical and Ecclesiastical Latin.[35] Using this function, users can type a Latin lemma without diacritics, and access information on the quantities of its syllables, and its accentuation, pronunciation(s), and basic grammatical information, including the presence of homographs that have different meaning and prosody. For example, when one searches the word *praedico*, the program shows that two lemmas have the same spelling, one of them with the penultimate syllable short and being a verb of the first conjugation (*prāedĭco*, *prāedĭcas*, *praedĭcāre*, *prāedicavi*, *prāedicatum*; to announce), while the other has the penultimate syllable long and belongs to the third conjugation (*prāedīco*, *prāedīcis*, *prāedīcĕre*, *prāedīxi*, *prāedīctum*; to foretell). Although the verb *prāedĭco* is less common than the verb *prāedīco*, in these cases, the database displays both entries to help users note potentially ambiguous forms, showing eventual differences in their pronunciation.

To make the program more accessible to beginner students, the app provides not only the IPA transcription, but also the syllabication and accentuation of each lemma using the Latin alphabet. This information, displayed between squared brackets, can be used for both Classical and Ecclesiastical Latin.[36] However, when reading late-antique and medieval texts in Ecclesiastical pronunciation, users should be aware that, after the quantity of vowels was no longer perceived by Latin speakers, the accentuation of some words changed. For instance, while Classical Latin could not preserve the original accentuation of Greek words such as φιλοσοφία (philosophy), which was pronounced *philosóphĭa* according to the Latin

30   On the importance of frequency lists in the pedagogy of Latin, see Muccigrosso (2004). On the use of frequency data for language teaching in general, see Sinclair (1991), 30 and Davies (2005), vii.

31   See Folco Martinazzoli (1953) on the use of the concept of *hapax legomenon* by ancient commentators and Denooz (2010), 1–2.

32   Latin frequency dictionaries have been published by Diederich (1939); Delatte/Evrard/Govaerts/Denooz (1981); and Denooz (2010). The largest corpora used so far is that of Denooz (2010), which includes nineteen authors but does not include important texts such as Ovid's *Metamorphoses*.

33   The original source code is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/src (last access 02.09.2020). The lemmatizer *Lemlat* can be accessed at http://www.lemlat3.eu/ (last access 12.01.2021).

34   On modern languages, see, for instance, Coxhead/Nation/Sim (2015), 121–35. Modern Latin frequency lists tend to provide students with only a few thousand terms. For instance, Williams (2012) offers a 1425 word-list.

35   The database is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/ master/Mobile%20APP/Data_App_accentuation_Ipa.txt (last access 02.09.2020).

36   See Pavanetto (2009).

prosody, the Greek words introduced into Latin vocabulary after the disappearance of vocalic quantities around the third century CE maintained their original Greek accentuation (e.g., ἔρημος, *éremus*, 'hermitage'). Similarly, the words in which the penultimate syllable was short and was followed by a *muta cum liquida*, which were stressed on the third last syllable in Classical Latin (e.g., *íntĕgrum*; intact), tended to be accentuated on the penultimate syllable in late-antique and medieval Latin (e.g., *intégrum*).[37]

Section Three, which users can access through the button 'Info', explains the genesis of the app and acknowledges the work of the Classicists (T. Spinelli, C. Pavanetto) and Computer Scientists (Giacomo Fenzi, Kamil Kolosowski, Jan Rybojad) who developed it. Moreover, it contains links to the online repositories in which the codes and programs underpinning the app are stored. Overall, in its unique and unprecedented features, the *Handbook of Latin Phonetics* app contributes importantly to language teaching and to stylistic and prosodic studies by allowing even beginner students and non-expert users to learn the most common Latin words and their correct pronunciations, as recommended by the most recent studies on Latin linguistics.

The online toolkit: outline and features

Available in opensource, the *Latin Phonetics* online toolkit contains the source codes through which the data displayed in the app is generated. While the app offers only premade information that can be easily accessed (even offline) by every user who is able to use a smartphone, the online program allows users with good informatic skills to generate customized results. As I have anticipated in the introduction, the online program is accessible through the GitHub page of the *Latine Loquamur Project*.[38] The project's home page currently features two repositories containing, respectively, the *Online Dictionary of Latin Near Synonyms* (which I plan to discuss in another article), and the program on Latin phonetics, which can be accessed by clicking on the folder 'Latineloquamur-toolkit-IPA-transcriber-and-App'.
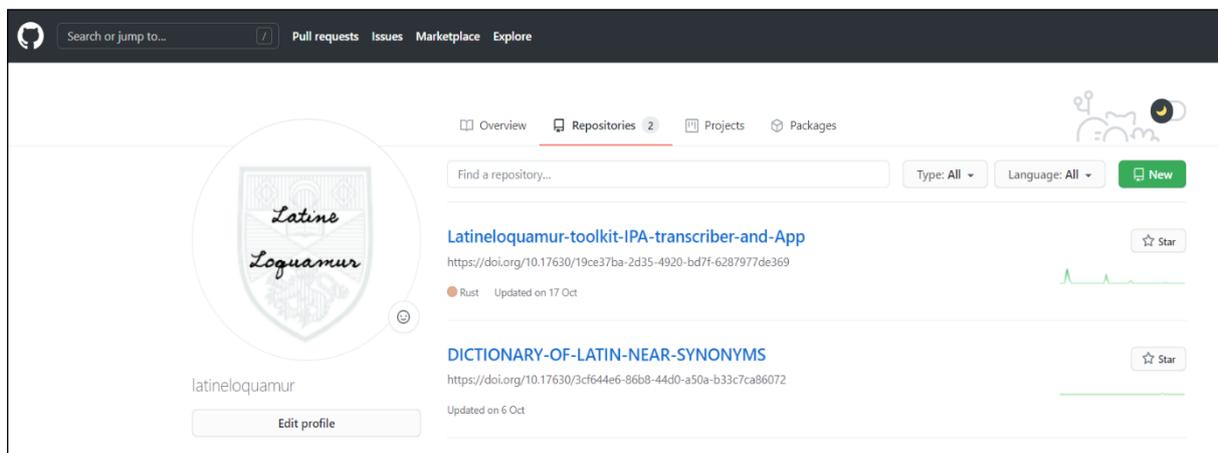

Fig. 4 The homepage of the Latine Loquamur repository in GitHub.

The repository that hosts the program on Latin phonetics is organized in different folders corresponding to the different functions performed by the program. This means that, while one can see the accentuation, syllabication, phonetic transcription and potential homographic forms of selected lemmas simultaneously in the app, the online program generates these results separately through different packages.

---

37   On the evolution of Latin through Late Antiquity, see Norberg (1999), 33–35.

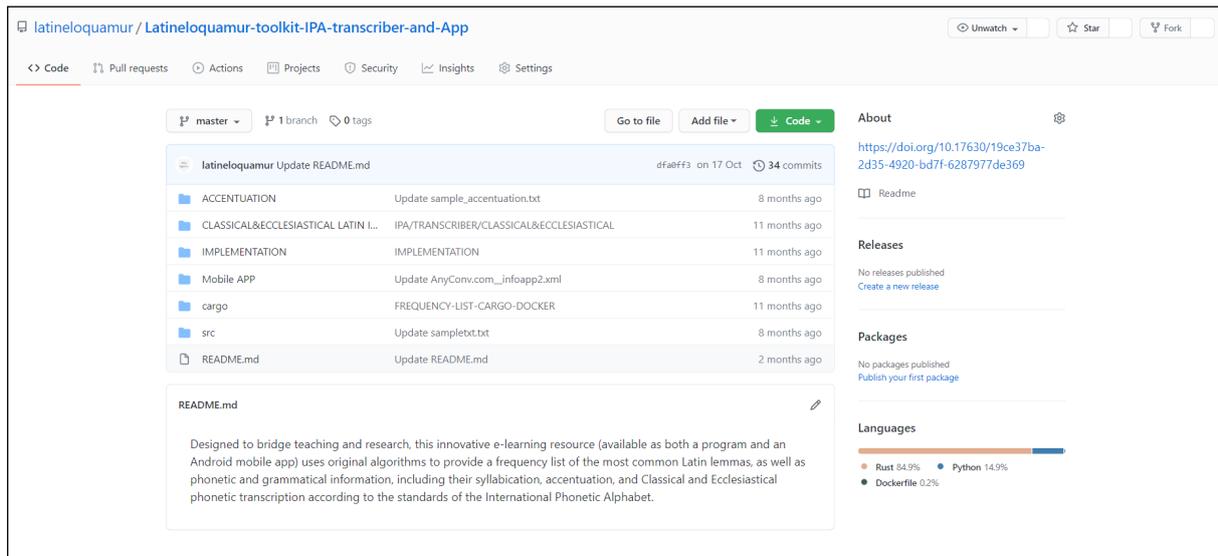38   https://github.com/latineloquamur?tab=repositories (last access 02.09.2020).

Fig. 5 The folders in which the Latin Phonetics online program is organized.

The folder 'Accentuation' contains the codes to accentuate automatically macronized Latin words organized in a CSV-file.[39] The output of this program is a new CSV-file displaying both the original word and its accentuated form.[40] The folder 'Classical and Ecclesiastical Latin IPA Transcriber' hosts the codes that perform the transcription of given Latin lemmas into phonetic characters according to the standards of the International Phonetic Alphabet.[41] Here the file 'Readme.md' provides users with a detailed guide on how to run this complex package. In extreme synthesis, by using in sequence the commands *cargo build* and *cargorun--{path to the file}* users can operate the phonetic transcription of Latin words (organized one per line in a txt file) and generate two files containing, respectively, the Classical and Ecclesiastical pronunciation of those words. A sample of the results that can be achieved using this package is provided by the section 'sample IPA'.[42] The folders 'Implementation' and 'Mobile App' can be disregarded by users as they contain, respectively, work-in-progress material that will be used in the implementation of the toolkit (as described in the final section of this article) and the codes that have been used to build the app. Using the folders 'Cargo' and 'Src' expert users can generate frequency lists of Latin lemmas attested in a customizable set of Latin texts.[43] Specifically, the folder 'cargo' governs the functioning of the packages hosted in 'Src' and contains a 'Dockerfile' with instructions. To use this program, users can upload the texts that they wish to process in the sub-folder 'data_dir' (within 'Src').[44] Files in this repository must be organized in folders, one for each author, and named with the authors' names. Inside each author-folder, there must be a list of folders corresponding to the author's works, which must be stored in 'txt' format. An example of how to organize personalized textual corpora effi-

---

39   Both functions are available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/ACCENTUATION/Latin_accentuation_code.py (last access 02.09.2020).

40   A sample is available at (https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/ACCENTUATION/sample_accentuation.txt (last access 02.09.2020).

41   https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/CLASSICAL%26EC-CLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER (last access 02.09.2020).

42   https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/CLASSICAL%26EC-CLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER/sample%20IPA (last access 02.09.2020). Note that slight differences may be caused by the manual checking operated on the data used in the app.

43   https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/cargo; https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/src (last access 02.09.2020).

44   The repository is accessible at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/src/Data_dir (last access 02.09.2020).

ciently is offered by the file 'sampletxt.txt'.[45] While the Android app simply provides a list of the 6500 most common Latin lemmas, expert users can perform much more complex searches with the online program to generate data on the use of a term or only on some of its forms by specific authors within a selected time frame. The tremendous potential of this tool can be appreciated by looking at the sample that, stored in the GitHub repository, shows a part of the data generated by running the program on our large textual corpus.[46]

Technologies

The original technologies used to develop the program underpinning the app are highly advanced and closely tailored to its aims and function. This complex program is organized in different packages that govern, respectively, the frequential statistics of the words most commonly attested in Latin literature, the syllabication and accentuation of Latin lemmas, their transcription into the characters of the International Phonetic Alphabet, and their visualization through a user-friendly mobile app.[47] The following section will discuss the most important technologies and methodological issues concerning each component of the backend.

Lemmatizing and counting Latin

The first stage of the development of the *Latin Phonetics* toolkit was the creation of a unique frequency list of the 6500 most common Latin lemmas, as attested across a large corpus of both Classical and Christian texts. Making this list involved parsing, lemmatizing, and categorizing the data directly from the sources, which means scanning a pre-assembled and standardized textual corpus in order to identify the different inflected forms of each word, and to calculate which lemmas are the ones most frequently used by Latin authors.

While most previous Latin dictionaries have relied on a manual processing of texts, this toolkit uses original algorithms and the capabilities of the opensource lemmatization service offered by *Lemlat*.[48] Lemmatization is the process through which the variants of a term, and its inflected or graphically different forms (e.g., *amat*, *amant*, *amas*, *amavi*, *amatum*; to love), are attributed to their lemma: the standard form of the word (e.g., *amo*) as it appears in a dictionary. Many programs can perform this task on Latin texts quite successfully (e.g., the Schinke algorithm, the Perseus lemmatizer, *PROIEL*, *Parsley*, *Morpheus*, Whitaker's *Words*, *LatMor*), but none of these technologies provides entirely correct data.[49] Among them, we have chosen to use a freely adapted version of *Lemlat*, which was developed between 2002 and 2004 by the National Research Centre (CNR) of Pisa in collaboration with the University of

---

45 https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/commit/b475af49a9dcbabb3a9cb-70582da84b5df18ecd9 (last access 13.12.2020).

46 Samples are available on Github (https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/src/sample_frequency_list_data.txt; last access 02.09.2020) and in a dedicated, work-in-progress webpage (https://latin.netlify.com/; last access 02.09.2020).

47 These packages are freely accessible through the program's repository: https://doi.org/10.17630/19ce37ba-2d35-4920-bd7f-6287977de369 (last access 02.09.2020); https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcri-ber-and-App (last access 02.09.2020).

48 Cf. http://www.ilc.cnr.it/lemlat/lemlat/index.html (last access 12.01.2021).

49 See, for instance, *LatMor* (http://cistern.cis.lmu.de; last access 02.09.2020), *Words* (http://archives.nd.edu/words.html; last access 02.09.2020), *Parsley* (https://github.com/goldibex; last access 02.09.2020), *PROIEL* (https://github.com/proiel/proiel-treebank; last access 02.09.2020), and *Morpheus* (https://github.com/tmallon/morpheus; last access 02.09.2020). On these technologies see also Springmann/Schmid/Dietmar (2016).

Turin, because it has proved to be the most consistent and reliable technology of this kind.[50] Based on a database of 40,014 lexical entries and 43,432 lemmas including many late antique and medieval terms, *Lemlat* adopts the standards of the *Oxford Latin Dictionary* (Glare [1982]). Being able to recognize over 97% of Latin terms including many anthroponyms and toponyms, it successfully lemmatizes 319,725 lexemes into 30,413 lexical entries (around 3,500 more than the modern Liège dictionary by Denooz). Moreover, its automatic analysis is very accurate and takes into account many spelling variations and even rare or archaic forms of a lemma which the former frequency dictionaries neglect.[51] However, this technology, which is still undergoing further development, cannot disambiguate homographic forms, which are therefore counted under all the lemmas which they can belong to.

To create the frequency list used in the app, we have fed into the program (written in RUST) a large textual corpus yielding some 9,484,029 words, and covering the works of 307 authors, which has been created using different opensource textual databases available online such as Perseus, the PHI database, and the *Bibliotheca Augustana*.[52] This textual corpus, which has not been made publicly available in accordance with its distribution licence, was stored in the repository *data_dir*.[53] As I have anticipated, this folder has been left empty in the program's repository, so that users can input a personalized corpus on which they can run our program by using, for example, the big textual databank provided by *Perseus*, or the *Packard Humanities Institute* both online and on CD. The most important element of this package is the 'lemmatizer.' This file is a 'CSV' directly exported (with adaptations) from Lemlat to specify the lemmatization bases that are used to operate on the literature.[54] This section also contains an original 'runner' program that is in charge of the full *GraphQL* endpoint, being used to query the text through the generic command: '*cargo run--bin {program_name}---aAUTHORS_FILE-dDATA_DIR-lLEMM_FILE*.' There, the options *authors_file*, *data_dir*, *lemm_file* are used to specify the data files on which to operate. Specifically, *authors_file* is used for an advanced function which is still being perfected; this folder contains a database with the chronology of Latin authors which can be used to perform

---

50   The program is available at http://www.lemlat3.eu/ (last access 02.09.2020). On its features and assessment see Passarotti/ Baudassi/Litta/Ruffolo (2017), 24–31 and Springmann/Schmid/Dietmar (2016). The adapted version of the lemmatizer is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/src/latin_ lemmatizer (last access 02.09.2020).

51   See Passarotti/Baudassi/Litta/Ruffolo (2017), 26. A way to check the ways in which forms are lemmatized in our dictionary through *LEMLAT* is through http://www.ilc.cnr.it/lemlat/lemlat/index.html (last access 12.01.2021). The lemmatizer Lemlat has successfully lemmatized more than 97% of the word-forms attested in our corpus, leaving unrecognized only the 2.88% of the forms. Among them many are names, Greek forms used in Latin texts, indication of books given in Roman letters (e.g., LXV), or Latin endings (e.g., -*ar*; -*or*) that are mentioned by ancient grammarians in their discussions of Latin morphology but do not correspond to any lemma.

52   The RUST source code is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcri- ber-and-App/tree/master/src (last access 02.09.2020). On *Perseus* see http://www.perseus.tufts.edu/hopper/ (last access 02.09.2020); the Bibliotheca Augustana can be accessed at http://www.hs-augsburg.de/~harsch/a_impressum.html (last access 02.09.2020). The list of authors included in our textual corpus is stored in this repository: https://github.com/la- tineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/src/authors_chrono/AUTHORS-LIST (last access 02.09.2020). Our corpus uses the *PHI* standards to name Latin authors in order to facilitate the use of the toolkit by other users who will likely use the *PHI* textual database. The *Packard Humanities Institute* (*PHI*) corpus is one of the widest opensource Latin corpora currently available online https://latin.packhum.org/ (last access 02.09.2020) and, al- though it does not match our corpus perfectly, it can be effectively used to look up the large majority of the Latin passages in which our lemmas or their inflected forms appear. However, while our program lemmatizes every inflected form, the *PHI* searching tool performs only simple pattern-matching queries. Thus, if one searches 'ultor' the program shows also results like '*multorum*', unless the search is made for a specific form like #*ultor*#. In this case the program displays only the occurrences of this specific graphic form and not of the lemma *ultor* and of its inflected forms.

53   The repository is accessible at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/ master/src/Data_dir (last access 02.09.2020).

54   To run the lemmatizer use https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/mas- ter/src/latin_lemmatizer/src/parsers (last access 02.09.2020).

diachronic linguistic searches on specific centuries.[55] The data so created can be exported using the *CSV* function. Conceptually, this system works as a 'forest' of sorts, with one tree for each lemma, one leaf for each form, and each form with a collection of occurrences. The data obtained from this system is semi-structured (data is fully tagged, but its structure is not rigidly defined, allowing for great flexibility in terms of exporting it), and the entire system operates without interacting with the disk. Leveraging this core system, two applications can query the relational databases (texts, lemmatizer, and chronological authors map), and generate the results requested by users in different formats (e.g., Json, txt, Excel).[56]

Prosodic processing

After building the list of the most common Latin lemmas, other algorithms were used to divide words into their syllables, and to mark them as long or short respectively, which is indispensable for a correct phonetic transcription. In Latin, the quantity of syllables does not always coincide with the quantity of the vowels that they contain. However, syllables are always long, except where a short vowel is in an open syllable (a syllable that does not end with a consonant). For instance, the *u* in the second syllable of the word *sepultus* (buried) is short by nature (*se-pŭl-tus*). However, because it is closed (it ends in consonant), this syllable is long, and takes the accent (*sepúltus*). For this reason, the program displays the quantities of syllables. Several opensource tools can divide Latin words into syllables, and mark the long ones as such.[57] Among them, we used the syllabifier shared by CLTK and Collatinus because it has been recently implemented to correctly process 'exceptional' forms that do not follow the standard rule of syllabication, using a list made by Rev. Frère Romain Marie de l' Abbaye Saint-Joseph de Flavigny-sur-Ozerain in 2016.[58] Thus, this algorithm can correctly syllabify compound words in which consonants are counted in the same syllable (e.g., *de-scri-bo*; to describe). This tool also offers the most efficient macronizer currently available, which is based on eight Latin dictionaries.[59]

After marking the quantities of each syllable, original Python scripts were used to accentuate Latin lemmas.[60] This program, which I co-developed in collaboration with Jan Rybojad, takes as an input a CSV-file containing Latin lemmas, and parses words so as to break them into an array of Unicode characters. For each lemma, this array is further converted into two new arrays, one for sounds, and one for vowels (including diphthongs). The actual accentuation is performed through the functions *findStress* and *isLongVowel* that replace long vowels and diphthongs with the appropriate stressed vowels and diphthongs, according to the rules of Latin accentuation.[61] In particular, if the second-last vowel of a lemma is marked as long or is a diphthong, the program accentuates it; if the second-last vowel is marked as short

---

55   The runner program is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/src/runner (last access 02.09.2020).

56   Samples of potential results are available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/src/sample_frequency_list_data.txt; last access 02.09.2020) and https://latin.netlify.com/ (last access 0.09.2020).

57   For instance, this service is provided by: http://marello.org/tools/syllabifier/ (last access 02.09.2020); https://github.com/cltk/cltk/blob/master/cltk/stem/latin/syllabifier.py (last access 02.09.2020); https://github.com/biblissima/collatinus/blob/master/doc-usr/scander.md (last access 02.09.2020).

58   Cf. https://github.com/biblissima/collatinus/blob/master/bin/data/hyphen.la (last access 02.09.2020).

59   The dictionaries are De Valbuena (1819); Noël (1824); Quicherat (1836); De Miguel (1867); Franklin (1875); Lewis/Short (1879); Du Cange (1883); Georges (1888); Calonghi (1898); Gaffiot (1934); Gaffiot (2016).

60   See https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/ACCENTUATION (last access 02.09.2020).

61   Both functions are available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/ACCENTUATION/Latin_accentuation_code.py (last access 02.09.2020).

and not a diphthong, then it replaces the third-last vowel or diphthong with the corresponding stressed vowel. The output is a new CSV-file with lines in the format <word>, <accentuated word>.[62]

IPA transcriber for Classical and Ecclesiastical Latin

A unique feature of the *Latin Phonetics* toolkit is the phonetic transcription of both Classical and Ecclesiastical Latin using the *International Phonetic Alphabet*, which is performed by original algorithms. Because the program takes as input Latin words that have the quantities of their syllables fully marked, this operation is conceptually simple. Given an input word, the algorithm applies iteratively a number of replacement rules that, co-developed in collaboration with Giacomo Fenzi, convert a combination of Latin characters into the corresponding IPA symbols (e.g., $x \rightarrow$ /ks/). However, several factors make this process more complex. Firstly, the pronunciation of Classical and Ecclesiastical Latin has to be treated separately because it follows different rules. For instance, while the long e (*ē*) is pronounced as long /e:/ in Classical Latin, it is pronounced as a normal closed /e/ in Ecclesiastical Latin, where the quantity of vowels is no longer perceived as a phonetically significant element. Similarly, the nexus *gn*, which sounds /ŋn/ in Classical Latin, is softened in Ecclesiastical Latin (/ɲ/). In this important respect, the phonetic transcription of Classical Latin operated by the *Latin Phonetics* toolkit differs from that of CLTK in so far as it is based on the new phonetic transcriptions recommended by recent studies of Latin linguistics. Secondly, combinations of letters are sometimes pronounced as just one sound. While a *nexus* can have different lengths, the same letters that appear in a two-character group can be pronounced differently when they occur in a three-character nexus. For instance, in the term *ămīcĭtĭa* (friendship), the nexus '*tia*' is pronounced /tsja/ in Ecclesiastical Latin. However, in the plural form *ămīcĭtĭae* the same group '*tia*' appears in the longer group '*tiae*' which is formed by the nexus '*ti*+vowel' and the diphthong '*ae*'. In this case, the replacement /tsja/+/e/ would be wrong, because the group *ae* is monophthonged in /ɛ/ or /e/ in Ecclesiastical Latin, and the word is consequently pronounced /a.miˈtʃi.tsje/. To fix these problems, the program, which is written in RUST, processes Classical and Ecclesiastical Latin separately.[63] In particular, using as an input a path to a file containing a list of words (one per line), the functions 'cargo build' (the executable being 'target/debug/ipa_latin(.exe)') and 'cargorun--{path to the file}' operate parallel replacement for Classical and Ecclesiastical Latin, using strings such as 'if Classical {subs.push((„aei", „aei"));} else {subs.push((„aei", „ɛi"));}', where *else* is the Ecclesiastical pronunciation. As a result, the program generates two different files: 'Classical.txt' and 'Eccl.txt'. In order to efficiently treat *nexus*, such replacements are based on conversion rules that, specifying each possible combination, are applied in descending length-order, so as to match longer structures first.[64] In this way, for instance, the group *oe* is successfully transcribed as /e/ in Ecclesiastical Latin, rather than as /o/+/e/. The results of this process can be seen in the files stored in the GitHub folder 'sample IPA'.[65]

	A	B	C	D	E	F
1	SEARCH	GRAMMAR AND PROSODY: [ACCENTUATION], Classical /CLASSICAL PRONUNCIATION/, Ecclesiastical /ECCLES	GRAMMAR AND PROSODY	ACCENTUATION	CLASSICAL PRONUNCIATION	ECCLESIASTICAL PRONUNCIATION
2	a	ā, prep. + abl. : [a], Classical /a:/, Ecclesiastical /a/.	ā, prep. + abl.	a	a:	a
3	ab	āb + abl.: [ab], Classical /ab/, Ecclesiastical /ab/.	āb + abl.	ab	ab	ab
4	abdico	ābdīco, abdīcas, abdīcāre, abdicavi, abdicatum: [ab-di-co], Classical /a:bdɪko/, Ecclesiastical /abdiko/.	ābdīco, abdīcas, abdīcāre, abdica	ab-di-co	a:bdɪko	abdiko
5	abdico	ābdīco, abdīcis, abdīcēre, abdixi, abdictum: [ab-di-co], Classical /a:bdi:ko/, Ecclesiastical /abdiko/.	ābdīco, abdīcis, abdīcēre, abdixi, ab	ab-di-co	a:bdi:ko	abdiko
6	abditus	ābdītus, a, um : [ab-di-tus], Classical /a:bdɪtus/, Ecclesiastical /abditus/.	ābdītus, a, um	āb-di-tus	a:bdɪtus	abditus
7	abdo	ābdo, is, ere, didi, ditum : [ab-do], Classical /a:bdo/, Ecclesiastical /abdo/.	ābdo, is, ere, didi, ditum	ab-do	a:bdo	abdo
8	abduco	ābdūco, is, ere, duxi, ductum : [ab-dú-co], Classical /a:bdu:ko/, Ecclesiastical /abduko/.	ābdūco, is, ere, duxi, ductum	ab-dú-co	a:bdu:ko	abduko

Fig. 6 Sample of the database deployed by the app.

62   A sample is available at (https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/ACCENTUATION/sample_accentuation.txt (last access 02.09.2020).

63   The RUST transcriber is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/CLASSICAL%26ECCLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER (last access 02.09.2020).

64   https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/CLASSICAL%26ECCLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER/src/ipa.rs (last access 02.09.2020).

65   https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/tree/master/CLASSICAL%26ECCLESIASTICAL%20LATIN%20IPA%20TRANSCRIBER/sample%20IPA (last access 02.09.2020). Note that slight differences may be caused by the manual checking operated on the data used in the app.

App design

The project has also pioneered the creation of an intuitive mobile app that, titled *Handbook of Latin Phonetics*, makes the phonetic program easily accessible for students and non-expert users. Because the amount of data is relatively small, rather than connecting the app to the program through a rest API, the application has been designed to work offline, using as input a SQLite database containing the information created by the program. In this file, a list of Latin lemmas without diacritics is matched with the lemmas complete with grammatical and prosodic information and their phonetic transcriptions. Therefore, when one searches a word without diacritics, all the possible corresponding Latin lemmas, which may include different syllabic quantities as in the aforementioned case of *praedico*, are returned. The app was built using a Software Development Kit called *Flutter*, which allows one to build high-performance apps for iOS & Android from a single codebase, using Dart programming language.[66] In the future, it will become the native framework of Google's Fuchsia *OS*, so that a project developed in Flutter will work on three platforms: iOS, Android, and Fuchsia. The architecture of the app is simple, and uses a Business Logic Components pattern, meaning that everything in the app is represented as a stream of events, in which widgets submit events and other widgets respond.

Future developments

Two new implementations of the *Latin Phonetics* toolkit are being developed to further support both academic research and language teaching.[67] The first is a diachronic function which, based on an original diachronic mapping of Latin authors, will allow users to see by which Latin authors and in which century each lemma was used.[68] This feature will support not only stylistic choices in exercises of Latin composition, but also commentary writers (by providing a concise 'story' of each word and of its occurrences) and philological conjectures (by showing which words were more likely to be used by an author). The second new feature that is being designed will leverage my app of Latin synonyms to describe the meaning of each lemma (for which phonetic information is provided) directly in Latin through the list of its most important Latin synonyms.[69] This function will also support new digital technologies that are being developed to detect similarities of meanings and ideas between Latin texts, independently of precise lexical repetitions. Finally, a version of the app for Apple devices will be released soon.

Overall, in its innovative cross-fertilization of recent developments in the fields of Latin linguistics, pedagogy, and digital humanities, the *Latin Phonetics* toolkit bridges teaching and research, using original algorithms to provide scholars and students with the first IPA phonetic transcription of the Classical and Ecclesiastical pronunciations of the most common Latin lemmas, as attested across the entire corpus of Latin literature. Besides facilitating the teaching of Latin in Latin and contributing to the creation of a shared methodology for the study of Latin phonology, this tool also supports a more interactive inde-

---

66 An overview of this innovative technology is provided by Kuzmin/Ignatiev/Grafov (2020). Cf. https://flutter.dev/ (last access 02.09.2020).

67 Originally, we had planned to assess the impact of our toolkit (published at the beginning of 2020) by using students' and professors' feedback. However, due to the Covid-19 pandemic this has been impossible so far. We now aim to collect and examine feedback after the development of the two new implementations.

68 A draft is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/src/authors_chrono/AUTHORS-LIST (last access 02.09.2020).

69 A sample is available at https://github.com/latineloquamur/Latineloquamur-toolkit-IPA-transcriber-and-App/blob/master/IMPLEMENTATION/IMPLEMENTATION_SYNONYMS.txt (last access 02.09.2020).

pendent learning of Latin, displaying the benefits of truly interdisciplinary approaches to the study of classical languages.[70]

References

Allen (1966): W. S. Allen, Vox Latina: a guide to the pronunciation of Classical Latin, Cambridge 1966.

Avitus (2018): A. G. Avitus, Spoken Latin: Learning, Teaching, Lecturing and Research, Journal of Classics Teaching 19.37 (2018), 46–52.

Barnett (2006): F. J. Barnett, The second appendix to Probus, The Classical Quarterly 56.1 (2006), 257–278.

Calonghi (1898): F. Calonghi, Dizionario Latino-Italiano, Milan 1898.

Chiesa (2012): P. Chiesa, L'impiego del ,cursus' in sede di critica testuale: una prospettiva diagnostica, in: F. Bognini, Meminisse iuuat. Studi in memoria di Violetta De Angelis, Florence 2012, 279–304.

CIL = Corpus Inscriptionum Latinarum, ed. Th. Mommsen et alii, Berlin 1862ff.

Collins (2012): A. Collins, The English pronunciation of Latin: its rise and fall, Cambridge Classical Journal, 58 (2012), 23–57.

Collins (1988): J. F. Collins, A primer of Ecclesiastical Latin, Washington 1988.

Coxhead/Nation/Sim (2015): A. Coxhead/ P. Nation/ D. Sim, Measuring the Vocabulary Size of Native Speakers of English in New Zealand Secondary Schools, NZ. J. Educ. Stud. 50 (2015), 121–135.

Cupaiuolo (1991): F. Cupaiuolo, Problemi di lingua latina. Appunti di grammatica storica, Naples 1991.

Davies (2005): M. Davies, A frequency Dictionary of Spanish: Core vocabulary for learners, Abingdon 2005.

De Miguel (1867): R. De Miguel, Nuevo Diccionario Latino-Español Etimológico, Leipzig 1867.

De Valbuena (1819): M. De Valbuena, Diccionario Universal Latino-Español, Hermanos 1819.

Delatte/Evrard/Govaerts/Denooz (1981): L. Delatte, É. Evrard, S. Govaerts, J. Denooz, Dictionnaire fréquentiel et index inverse de la langue latine, Liege 1981.

Denooz (2010): J. Denooz, Nouveau lexique fréquentiel de latin. Alpha-Omega, Liege 2010.

Diederich (1939): P. B. Diederich, The Frequency of Latin Words and their Endings, Chicago 1939.

Du Cange (1883): C. Du Cange, Glossarium Mediae et Infimae Latinitatis. Ed. L. Favre, Niort 1883–1887.

Franklin (1875): A. Franklin, Dictionnaire des noms, surnoms et pseudonymes latins de l'histoire littéraire du Moyen Age, Paris, 1875.

Gaffiot (1934): D. L. F. Gaffiot, Dictionnaire Latin-Français, Paris 1934.

Gaffiot (2016): D. L. F. Gaffiot, Dictionnaire Latin-Français, London 2016.

Georges (1888): K. E. Georges, Kleines deutsch-lateinisches Handwörterbuch, Hannover / Leipzig 1888.

Harrington/Pucci (1997): K. P. Harrington, J. Pucci, (eds.), Medieval Latin, Chicago 1997.

Kuzmin/Ignatiev/Grafov (2020): N. Kuzmin, K. Ignatiev, D. Grafov, Experience of Developing a Mobile Application Using Flutter. Lecture Notes in Electrical Engineering 621 (2020), 571–75.

Lewis/Short (1879): C. T. Lewis, C. Short, Latin Dictionary, founded on Andrews' Edition of Freund's Latin Dictionary: Revised, Enlarged, and in Great Part Rewritten, Oxford 1879.

Martinazzoli (1953): F. Martinazzoli, Hapax legomenon: Parte prima, Roma 1953.

Muccigrosso (2004): J. D. Muccigrosso, Frequent vocabulary in Latin instruction, The Classical World 97.4 (2004), 409–433.

Noel (1822): F. Noel, Dictionarium Latino-Gallicum: dictionnaire latin-français, compose sur le plan de l'ouvrage intitulé: Magnum totius latinitatis lexicon, de Facciolati, septième edittion, Paris 1822.

Norberg/Oldoni (1999): D. Norberg, M. Oldoni, (eds), Manuale di Latino Medievale, Napoli 1999.

Oniga (2003): R. Oniga, La sopravvivenza di lingue diverse dal latino nell'Italia di età Imperiale: alcune testimonianze letterarie, Lexis: poética, retórica e comunicaciones nella tradizione classica, 21 (2003), 39–62.

Oniga (2014): R. Oniga, Latin: a linguistic introduction, Oxford 2014.

Passarotti/Baudassi/Litta/Ruffolo (2017): M. Passarotti, M. Budassi, E. Litta, P. Ruffolo, The 'Lemlat3.0' Package for Morphological Analysis of Latin, Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language, 133 (2017), 24–31.

Paulus VI (1976): Paulus P. P. VI, Romani Sermonis, Roma 1976.

Pavanetto (2009): C. Pavanetto, Elementa Linguae et Grammaticae Latinae, sexta editio aucta et emendata, Roma 2009.

Préaux (1957): J. Préaux, Premier congrès international pour le latin vivant: Avignon 3–6 septembre 1956, Latomus, 16.3 (1957), 509–511.

Quicherat (1836): L. Quicherat, Dictionnaire Français-Latin, Paris 1836.

Ramage (1963): E. S. Ramage, Urbanitas: Cicero and Quintilian, a contrast in attitudes, The American Journal of Philology, 84.4 (1963), 390–414.

Serianni (1998): L. Serianni, Lezioni di grammatica storica italiana, Roma 1998.

Short/George (2013): E. Short, A. George, A Primer of botanical Latin with vocabulary, Cambridge 2013.

Sinclair (1991): J. Sinclair, Corpus, concordance, collocation. Oxford (1991).

Spinazzè (2014): L. Spinazzè, Cursus in clausula. An Online Analysis Tool of Latin Prose, Association for Computing Machinery, 10 (2014), 1–6.

Springmann/Schmid/Dietmar (2016): U. Springmann, H. Schmid, N. Dietmar, LatMor: A Latin finite-state morphology encoding vowel quantity, Open Linguistics – Topical Issue on Treebanking and Ancient Languages: Current and Prospective Research, 2.1 (2016), 386–392.

Traina/Bernardi-Perini (1998): A. Traina, G. Bernardi-Perini, Propedeutica al latino universitario, Rome 1998.

Williams (2012): M. A. Williams, Essential Latin Vocabulary: The 1,425 Most Commom Words Occurring in the Actual Writings of Over 200 Latin Authors, Milan 2012.

Winge (2015): J. Winge, Automatic Annotation of Latin Vowel Length. Bachelor's Thesis in Language Technology, Uppsala University, https://cl.lingfil.uu.se/exarb/arch/winge2015.pdf (last access 02.09.2020).

Figure references

Fig. 1: Latin Phonetics App, by Tommaso Spinelli.
Fig. 2: App's introductory section, by Tommaso Spinelli.
Fig. 3: Sarcophagus of Scipio Barbatus (Vatican Museums), image by the Center for Epigraphical Studies of the Ohio State University.
(http://db.edcs.eu/epigr/bilder.php?s_language=de&bild=$OH_CIL_06_01284_1.jpg;$OH_CIL_06_01284_2.jpg;$OH_CIL_06_01284_3.jpg;$CIL_01_00006.jpg;PH0010886;PH0010887;pp; last access 12.01.2021).
Fig. 4: The homepage of the *Latine Loquamur* repository in GitHub, by Tommaso Spinelli.
Fig. 5: The folders in which the *Latin Phonetics* online program is organized, by Tommaso Spinelli.
Fig. 6: Database of Latin phonetics, by Tommaso Spinelli.
Tab. 1: Phonetic Transcription Table, by Tommaso Spinelli.

Author contact information[71]

Dr Tommaso Spinelli

University of Manchester
School of Arts, Languages & Cultures

E-Mail: tommaso.spinelli@manchester.ac.uk