



The banner features a row of six icons: a globe, a book, a handshake, a money bag with a Euro symbol, a scale of justice, and a bicycle. Below the icons, the text reads: "AIUCD 2021", "DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale", and "10° congresso annuale PISA 19-22 gennaio". On the right side, a list of topics is displayed in colored text: "DIGITAL PUBLIC HUMANITIES" (red), "OPEN CULTURE" (orange), "RETI SOCIALI" (yellow), "TECH ECONOMY" (green), "E-PARTICIPATION" (blue), and "TECNOLOGIE ASSISTIVE" (purple). The background includes binary code and a classical building facade.

AIUCD 2021

DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale

10° congresso annuale **PISA** 19-22 gennaio

DIGITAL PUBLIC HUMANITIES
OPEN CULTURE
RETI SOCIALI
TECH ECONOMY
E-PARTICIPATION
TECNOLOGIE ASSISTIVE

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Editing (and publishing) medieval vernacular inscriptions in a digital environment: potential and limitations

Nadia Cannata

Università di Roma “La Sapienza” – Italy – [nadia.cannata\(«»\)uniroma1.it](mailto:nadia.cannata@uniroma1.it)

ABSTRACT

I should like to discuss briefly the advantages of an integrated system of publication – digital and traditional – in relation to the critical editing and linguistic analysis of medieval texts, with particular reference to texts of a non-literary nature, composed in the numerous Romance varieties in use in Italy in the late Middle Ages. Digital publication constitutes a ground-breaking innovation in the methods of textual and linguistic analysis, and ultimately for the historical interpretation of our linguistic past. This is linked both to the potential for interrogation of marked up texts, and the invaluable information about the ‘grammar of writing’ and the spoken languages underlying written records it yields, and to the fact that important datasets, acquired independently by different teams of scholars, across time and space, are expanding very significantly the corpora of texts available to scholars and open to digital interrogation. Therefore we can engage in analysis and cross referencing of textual and linguistic (sometimes even material) features of medieval Romance texts in the numerous languages in which they were composed and gauge our findings in a much wider context than traditional philological methods would allow – a fact that does not only make more data available and brought to scholarly attention, but actually has the potential to influence very significantly our perception of history. A change in the means of documenting the past cannot but result in an important reassessing of its significance and of our abilities to interpret it.

KEYWORDS

Cultural history, orality and writing, multilingualism, digital editions, medieval vernaculars, digital epigraphy

1. INTRODUCTION

I have devoted the best part of ten years to a major digital project: *EDV: Epigraphic Database of Inscriptions in Italian Vernaculars (IXth-XVth c.)* now nearing conclusion. It records the corpus of all extant vernacular inscriptions produced in Italy from the middle of the 9th to the end of the 15th century, provided they were meant to be displayed publicly, are still extant, and were composed in any language other than Latin. The database currently records around 570 inscriptions nicely spread across the whole period.

The aim of the study is fundamentally threefold: to preserve a record of a significant and hitherto largely unknown corpus of medieval texts of a non-literary nature; to provide an archive of documentary evidence of the uses of language(s) other than Latin in public script in late Medieval and Early Modern Italy; and to devise a powerful research tool to study such languages.

It is my contention that the first two aims are best served by traditional forms of publication, in print and in book form, whilst the research potential unleashed by the possibility of interrogating texts marked up in XML is such that it should become an integral part of the research methods applied to the study of the linguistic heritage of the past. We are planning to publish the complete catalogue of inscriptions and its study in book form, as well as to keep the website open to update and add materials as appropriate. The two forms of publication are devised independently, but designed to support one another. Whilst the updating of data is an open ended task, which benefits greatly from the flexibility of digital means, scholarship needs books and libraries for which have provided for thousands of years the most reliable and durable form of cultural transmission.

Moreover, digital projects are very costly and do not guarantee sustainability over time. The issue of the long term preservation of the outcomes of such projects is therefore a pressing one. An integration between digital and traditional means of publication might help address such issues, and allow wide dissemination and accuracy of data, as well as durability in time, thus ensuring that the very significant investment of time and resources that normally goes into big digital projects reaps its expected rewards.

I shall devote the next two paragraphs of this contribution to illustrate the theoretical issues and research questions that I believe digital resources enable us to address in a way not possible before, then I shall say something more about how this is working for the EDV project.

2. RESEARCH QUESTIONS AND EXPECTED RESULTS

At the time of writing we have recorded for the entire corpus the general features (date, place of origin, location, type of object, material, language(s) used, function, script) and the complete bibliography available to date for each inscription, all openly available and easily browsable on our website.

We have also completed the diplomatic transcription and critical edition of one hundred and twenty inscriptions from Central and Southern Italy and marked up the text in Epidoc, which is an established implementation of TEI, used widely by epigraphers working on inscriptions from the most diverse cultures, times, and languages. For these we were therefore able to generate a complex indexing of spelling systems and linguistic features, which is allowing us to catalogue important information for the interpretation of the linguistic data gathered in the corpus. It is a known fact to sociolinguistics that any sociolinguistic investigation depends upon the ability to engage with a comprehensive dataset, and to contextualize whatever the linguistic document under scrutiny is into the wider picture of general language use. When studying documents written in languages which were not standardized, the relationship between spelling and pronunciation, or the influence of the standard on common usage, and of the usage on the perceived standard, as well as the possibility of comparing such data with contemporary production of a different nature yields information that can prove very interesting to understand linguistic heritage. It is particularly useful to identify language uses which – because of their marginality or lack of prestige – may well have escaped our attention. As a philologist and textual critic mostly engaged in the study of languages and literary traditions in Romance Europe, with particular reference to Italy and its vernaculars, I have often wondered through what methods – other than those which historical linguistics and Romance Philology in particular have passed down to us – might we be able to add to our knowledge of the languages actually spoken in Medieval and Early Modern Europe, given that the available documentation testifying to the linguistic culture of that time and place is constituted by written records. Attempting to engage in historical sociolinguistics is so much harder since it is very questionable if the languages whose use was solely, or mostly, oral and confined to everyday exchanges are somehow accounted for in writing, or rather if – as one would be inclined to conclude – they have mostly fallen through the net of historical memory, due to the lack of positive means apt to secure their preservation.

Interestingly, such veil of haze surrounding the language varieties in most frequent use in the period and places under consideration is mirrored by the lack of appropriate terms we have at our disposal to describe them. When considering the languages in use in Italy in the period 5th – 13th c. one has to contend, as a preliminary consideration, with the fact that our description of the languages of Italy in the Middle Ages is based on categories which were not shared by the communities whose languages are the focus of our attention.

There are around 5 centuries of linguistic usage from the 5th to at least the 9th c. which appear unaccounted for, since we would be quite unable to answer the simple question “What language(s) were spoken in Italy in the 5th, 6th, 7th, and 8th centuries?” We define languages in use in Italy after the 9th c. by using as our pole of attraction no longer Latin, but the dominant literary vernaculars at the time: Sicilian, Tuscan, Provençal, Old French, Catalan and so on, which does not help with the issue of how to account for languages not graced with a written tradition, nonetheless largely in use by local communities. But even when we move forward in time, to consider the centuries between the 10th and the 15th we are at pains at finding names for the languages recorded by epigraphy or documents (which are widely understood to record ‘natural’ everyday language more faithfully than other kinds of text) and those supposedly spoken other than as Tuscan, Central Italian, Sicilian etc. vernaculars.

Roger Wright and Michel Banniard, among others, have argued extensively about the need to devise a new paradigm to understand language change, its periodization and to deal with the issues that the rigid protocols of 19th c. philology seem to have passed to us and which now seem to hinder, to an extent, our understanding of linguistic and cultural change in Neo-Latin Europe.

In this context I believe that digital interrogation of marked up texts might contribute to our finding an answer to two important research questions, closely related to one another, which have rather escaped us so far: 1) Is there a way for us to retrieve the voices of languages used in Romance Europe before the birth of ‘Italian’, ‘French’, ‘Spanish’, or indeed of ‘Catalan’ ‘Occitan’ ‘Galego portuguese’ etc. and after the demise of ‘Latin’, thence restoring them as far as possible to memory?

2) Might it be possible to develop a methodology which will allow us to better understand the grey zone connecting written and spoken language, and the hitherto scarcely charted territory from which Italian, as other modern Romance languages, issued and was subsequently circulated and preserved?

These issues may seem of a ‘pure’ philological and linguistic nature, but they can only be addressed if we look at language questions in the wider perspective of cultural history. For nearly a millennium in most of Europe Latin obliterated to posterity the wide variety of languages spoken in the 1000 years or so of the Roman domination in Europe. In the early modern period more languages graduated to writing, and we date their birth based on their written records. They do not document, however, the birth of new languages (which were alive long before any of them came to be written), but rather, more crucially, the development of a culture of writing which enabled them to acquire a written

form and therefore to transfer a memory of their existence in a form which is quintessentially different from the features we would like to believe are preserved. The creation of a new written culture also entails the birth of a new shared written culture and a common language to serve the purpose. Modern Europe was born multilingual, indeed as a consequence of a process through which new voices, many issuing from Latin, established themselves as independent from one another.

Writing has had the formidable task of constituting the fundamental means for the composition, reading, transmission and preservation of texts. All written documents are also linguistic documents, and survive thanks to a material support. It follows that any critical approach to the study of text and language cannot ignore either textual transmission, or, indeed, palaeography and the study of the history of writing in its material aspects.

It is my contention that it would be impossible to study and understand fully any written message if we were unable to decode not just the components that form it but the grammar of writing presiding over the way in which that language is written. Any critical approach to the study of texts cannot ignore either textual transmission, or, indeed, the study of the material aspects (scripts, writing surfaces, actors) of linguistic documents.

There appears therefore to be a third research question which begs to be answered:

3) Is here a way to separate, in as far as possible, the issues of 'how they spoke' from 'how they wrote' and to define, accordingly, language use in previously uncharted areas?

3. IN SEARCH OF A NEW METHODOLOGY

Digital publication allows us to assemble online and interactive libraries of texts, marked up through TEI P5 XML to interrogate phonology, spelling vs pronunciation, morphology, syntax, lexis, writing materials, scripts. We have applied this to the edition of epigraphic texts in the medieval vernaculars in use in Italy up to the 15th c., and are proceeding to cross reference this body of texts with other corpora of texts located and dated with a reasonable degree of accuracy, in order to establish the social context in which the documents were produced and their areas of circulation, and trace chains of influence between Latin and modern vernaculars, and across the vernaculars. This is helping to promote the appreciation of the sociolinguistics and the sociology of writing in late Medieval and Early Modern Italy, and the multilingual reality in communities in which Latin and vernaculars coexisted and nurtured each other; and it guides us in a better understanding of the components of the *scripta* by relating the linguistic evidence of the documents 'horizontally' to each other, rather than to ideal linguistic standards which had not yet been established. To begin with, we were thus able to identify 15 different vernaculars and produce indexes of the morphological and syntactic features through which we have categorized them.

As regards devising a methodology which would allow us at least to attempt to answer our research questions, here again digital interrogation is essential, since digital tools can support an innovative interaction between historical linguistics and the history of writing by allowing the cross referencing of huge amounts of textual data which, through their sheer volume, may give us better grounds to infer the relation between sign and sound in recording languages with a very limited written tradition.

An area of particular interest in this respect is constituted by spelling, for example in the number of graphic solutions elaborated to account for the development of palatals (which varied greatly in time from area to area and according to the script adopted, albeit not as haphazardly as one may think). The possibility of interrogating thousands of texts to look for the ways in which palatal and affricate consonants are represented in script would allow us to acquire new information about the varying distance over time of sound and sign. As an example: many Christian inscriptions display COIUX for 'spouse'. More often than not the form is considered as simply lacking the required *titulus*. However, the frequency with which the *titulus* slipped from the chisel, when seen alongside the use of 'I' for the palatal /ɲ/ in later Latin and vernacular documents (capital 'I' used when small 'I' would be expected in Beneventan script to signal in an unequivocal way a pronunciation of the palatal /ɲ/) invites the inference that the palatal had developed much earlier than current scholarship allows and that written culture strove to accommodate it.

Similarly the differing graphic solutions for /ʃ/, /ʎ/, /ɲ/ provide some indication on the pronunciation of sounds which failed to be preserved as phoneme in Italian (but constitute phonemes in Italian Romance vernacular varieties) because they did not find an accepted graphic representation; whereas the wealth of solutions adopted to represent /tʃ/ /dʒ/, /ts/, /dz/ suggests that for certain areas from the Latin consonants coupled with a glide (S+J, L+J, N+J, D+J, T+J) the development of an affricate rather than the palatal and thus give indications as to the pronunciation which have previously escaped our radar. We have noticed, for example, developments of spellings specific of certain regions: in texts written in Southern Italy, the grapheme <cz> as opposed to <ç> or <z> is generally used for /tʃ/. This use has proven to be systematic and contradicts common wisdom about the development of affricates in southern dialects of the Neapolitan type.

Similarly, the graphic solutions adopted in Christian inscriptions, both in Latin and in Greek alphabets, once studied systematically, have yielded very precious information about language development in Rome 5th-9th c. by allowing relations which have resulted in the identification of phenomena belonging to Romance linguistic varieties at a very

early stage. They, however, were not recorded in script and therefore belong for linguists to the realm of late or vernacular Latin, or else Protoromance, without acquiring the identity of a modern language. Such detailed analysis of spelling features is fundamental to try and chart a sociolinguistic of lost language varieties and to understand matters relating to the history of linguistic culture which cannot be studied independently from the history of written culture. The complex indexing we have elaborated for our data which can be perused freely at www.edv.uniroma1.it/it testify to it.

CONCLUSIONS

Digital publication allows for the interrogation of the materials in ways traditional publication does not countenance. Its limitations consist mainly in the volatility of the repository of the data. This, consequently, poses a tangible risk for the dissemination and impact in the long term of the results achieved. Scholars must therefore make sure they elaborate a system integrating digital and more traditional means of publication, so that both can support and enhance each other's potential. Only a solid institutional network constituting an international platform into which the data are fed and to which institutions are committed can counterbalance such volatility and help unleash the research potential of connecting a great wealth of data that may normally not 'communicate' among each other and help interrogating them with the systematic approach with which only texts in TEI and machine readable forms can afford as well as ensure their preservation to memory in the long run.

It might also be added that digital means allow for a wealth of new forms of dissemination, designed to encourage the general public to develop an interest in matters traditionally confined to the close cloisters of scholarship, by helping 'exhibit' and share whichever progress we make in learning through the visualization in websites (or even in actual physical spaces) of maps for the geolocalization of the documents, sound and image-files to hear the voices and bring back to life lost everyday language, or helping the viewer to elaborate some personal itinerary across the documents.[14]

REFERENCES

- [1] M. Banniard, *Viva voce. Communication écrite et communication orale du IVe au IXe siècle en Occident Latin*, Turnhout: Brepols, 1992.
- [2] Francesco Bruni, *Italia. Vita e avventure di un'idea*, Bologna: il Mulino, 2010.
- [3] Nadia Cannata, *Lontano da dove? Tradizioni culturali e coscienza linguistica in Europa*", *Critica del testo*, 19/1 (2016): 63-91.
- [4] Nadia Cannata, «Le parole sono pietre. *Lingua communis* e *lingua literata* in alcune epigrafi romane (secc. IV VI)», in *Scrivere, leggere, conservare. A colloquio con Armando Petrucci*, ed. Nadia Cannata and Maddalena Signorini, *Studi Romanzi* 10 (2014), Special Issue: 281-310.
- [5] *Écritures Exposées. Discours, matérialité et usages* a research project jointly coordinated by the École des hautes études hispaniques et ibériques (Casa de Velázquez, Madrid), the Grupo de Investigación «Lectura, escritura, alfabetización» (LEA), Seminario Interdisciplinar de Estudios sobre Cultura Escrita (SIECE) (Universidad de Alcalá) and the Centre d'Etude des Littératures et Langues Anciennes et Modernes (CELLAM), Groupe de recherche sur culture écrite et société (GRECES) (Université Rennes 2).
- [6] Giulio C. Lepschy, "*Ethnicity and language in Italy*" in *L'Italia allo specchio*, ed. Fabio Finotti, Venezia: Marsilio, 2015: 97-112
- [7] Isa Lori Sanfilippo and Giuliano Pinto (ed.), *Comunicare nel Medioevo. La conoscenza e l'uso delle lingue nei secoli XII-XV*, Roma: Istituto Storico Italiano per il Medioevo, 2015.
- [8] Maddalena Signorini, «Il ritmo cassinese: cultura grafico-libraria e qualche proposta di correzione», in *Scrivere il volgare fra Medioevo e Rinascimento* ed. Nadia Cannata Maria Antonietta Grignani, Pisa: Pacini, 2009: 1-26
- [9] Estelle Ingrand-Varenne, *Langues de bois, de pierre et de verre. Latin et français dans les inscriptions médiévales*, Paris: Classiques Garnier, 2017.
- [10] Nigel Vincent, «Languages in Contact in Renaissance Italy», in *Rethinking Languages in Contact. The Case of Italian*, ed. Anna Laura Lepschy and Arturo Tosi, Oxford: Legenda, 2006: 12-24.
- [11] *Visibile parlare: le scritture espone nei volgari italiani dal Medioevo al Rinascimento*, ed. Claudio Ciociola, Napoli: ESI, 1997.
- [12] Roger Wright, (ed.), *Latin and the Romance Languages*, London: Routledge, 1990
- [13] Roger Wright, *A SocioPhilological Study of Late Latin*, Turnhout: Brepols, 2002
- The Cambridge History of Romance Languages*, ed. M. Maiden, Adam Ledgeway and John C. Smith, Cambridge: CUP, 2013.
- [14] *Museums of language and the display of intangible cultural heritage*, ed. Nadia Cannata, Maia W. Gahtan, Margaret Sönmez, London-New York, 2020.