

**THE ORIGINS OF PORTUGUESE WRITING:  
REMARKS ON THE IMPORTANCE OF NOTARIAL DOCUMENTS  
FOR THE HISTORY OF THE PORTUGUESE LANGUAGE**

Instituto Camões Centre (Oxford), 7th March 2003

António Emiliano  
Universidade Nova de Lisboa  
ah.emiliano@fcsh.unl.pt

## **0. Introductory remarks**

The textualisation of Portuguese is commonly considered to have begun in the early 13th century: we know only two documents in Portuguese orthography from that period, the *Will of King Afonso II* from 1214, and the *Notícia de Torto*, dated indirectly from ca 1214. These two texts form what Ivo de Castro has called the “produção primitiva portuguesa”, i.e. “early Portuguese documental production”. Further documents written in Portuguese only appear considerably later, from 1255 on, in the chancery of King Afonso III.

The text known as *Notícia de Haver* can be dated from the late 12th century, and is clearly written in Portuguese (note specially the written forms of the definite article and the absence of Latin morphology): if it were dated or datable it should be regarded as the earliest known Portuguese text. The text known as *Notícia de Fiadores* dated from 1175, is a list of names with a single sentence in Romance spelling, is now regarded by some (but not by me) as the earliest known Portuguese text.

In order to understand the emergence of Portuguese writing in the early 13th century, we must bear in mind that it was preceded by a Latin-Portuguese tradition, which was based on the Late Latin Visigothic tradition, and which evolved with no cultural or linguistic interruption from Antiquity to the Middle Ages: it is crucial that we ask ourselves what this textual tradition was, how it came about, and how it evolved throughout several centuries of document production.

What it was, in terms of textual typology, and how it came to be is the object of Diplomatics, and I will not go into that.

How it evolved is of tremendous importance for the History of the Portuguese Language, because this Latin-Portuguese tradition is ultimately the origin of Portuguese writing, and its texts should indeed be regarded as the earliest known instances of the textualisation of Portuguese.

(many of the things discussed here apply also to the other Ibero-Romance domains with the exception of Catalonia)

The documents written before the 13th century have been referred to traditionally as Latin, because in a certain sense they look Latin, they have a certain Latin-like quality. However, close scrutiny of these texts shows that their latinity is peculiar, it is very different from Classical Latin, and that it clearly deviates from the grammatical standards prescribed by the *artes grammaticae* that were studied throughout the Middle Ages.

The concept of «latim bárbaro» (Barbarous Latin) still current among historians and latinists derives from the notion that legal documents were written by ignorant scribes, who strived without success to write good Latin, but could not help the interference of the vernacular on the written language.

## 1. Latin, Romance and Latin-Romance

One must regard the Notarial Latin found in documents from Castille, León, Galicia, and Portugal before the 13th century as being a special written code designed, as it were, to meet specific communicative needs. These documents formed a specific textual and communicative domain or genre, which had specific traits and characteristics. Their importance to the knowledge of the early periods of the Ibero-Romance languages is enormous and cannot be overstated: notarial documents were, so to speak, a “scripto-linguistic laboratory”, where for centuries scribes experimented with many of the graphemic solutions that would finally lead up to the development of autonomous Romance orthographies in the early 13th century.

(I use orthography in a loose and neutral sense; I should use instead scriptography)

Before the invention or creation of these orthographies there was no other way of representing the Ibero-Romance vernaculars: in order to write Romance before a Romance spelling system was available scribes had to make-do with the old Latin tradition. Romance was therefore written as if it were Latin, with some inflectional morphology and Latin lexis. But when the documents were read aloud they sounded like Romance, because they were read with Romance phonetics. They might have oldfashioned words and oldfashioned word-order, but that would have been perceived as features of medieval legalese, or simply as special features of the written language (it is still normal in many societies that the written and the spoken languages have distinct syntactic patterns and vocabulary).

So, the study of notarial documents must take account of the complex relationship that existed between the several types and modes of communication in medieval Romance-speaking societies.

Both the Latin written tradition, and the Romance vernaculars developed without any break, or discontinuity from Antiquity throughout the Early Middle Ages. Therefore, and this has been systematically overlooked by Romanists and Latinists, the relationship between the written and spoken languages in Romania was unbroken for centuries, until the Carolingian reforms, headed by Alcuin of York, introduced a completely new way of pronouncing Latin and also new standards of orthographic correctness. Written Latin became unintelligible to monolingual native Old French speakers. In the 11th century a similar reform, known as the Gregorian Reform (or Cluniac Reform), which, among other things, introduced Medieval Latin in Iberia, paved the way for the distinction between Latin and Romance as separate languages.

We can use the expression “Latin-Romance”, which was given a precise meaning by Roger Wright, to refer to the special written code of non-literary texts (like documents, legal codes, proceedings of councils and synods, chronicles), before the 13th century: it looked like Latin, but Romance-speakers of the time regarded it as the written representation of the vernacular; being used in special occasions it would seem natural that this written language possessed many archaic or even obsolete linguistic features, like certain word endings, or certain lexical items, or even a peculiar word-order in certain contexts. Formulae could be difficult to understand, but so is modern legal language today.

\*\*\*

Notarial documents, no matter how “barbarous” their language may seem to contemporary latinists, were valid communicative acts. Their language and spelling were accepted by their users as normal and adequate, and reflected a stable tradition of writing, which preceded for many centuries the development of Romance orthographies.

Any perspective of these texts based on a dysfunctional or handicapped conception of scribal competence is not acceptable, for the simple reason, that it is not grounded on solid evidence, and has no heuristic value.

To the historians and philologists that have written about the appalling “barbarousness” of Notarial Latin the documents were nothing more than result of widespread and ingrained

ignorance, and were written by poor semi-literate scribes: who were supposedly isolated and backward; who didn't know the *auctoritates* and the *litterae*, who struggled to no avail to master the intricacies of Latin orthography and grammar; whose vernacular kept creeping up as they tried to write good Latin, whose feeble attempts at writing Latin were systematically foiled and amounted to nothing but a corrupt and garbled language.

This pessimistic and negativist perspective of medieval Iberian literacy and of the early notaries of Spain, should yield to a more enlightened view; a view that must take into account the developments in several areas of research.

In discussing Latin-Portuguese documents we must nowadays use concepts such as “discourse community”, “textual interpretability”, “pragmatic competence”, “scripto-linguistic competence”, “graphemic structure”, “lexical access”, “grapho-phonemic transcodification vs grapho-semantic transcodification”, “graphemic polymorphism and variation”, and other concepts taken from research areas such as sociolinguistics, pragmatics, text linguistics, discourse analysis, psycholinguistics, scriptology, graphemics, cultural anthropology, social anthropology, etc.

It is also time that we get rid of what is basically a 19th century perspective of these texts. Throughout the 19th and 20th centuries Latin-Portuguese documents had the status of historical sources or records. Only seldom, and in a biased way, were they seen as linguistic sources or records. Even the romanists, who saw in the vulgarised spellings the emergence or interference of the vernaculars could not help looking at the texts as corrupt and decadent.

## **2. Representational models vs. operational models in the Latin-Romance written tradition**

Throughout the Middle Ages scribes had access to the old Latin grammatical tradition, as all clergy and monks had: they were taught to read and write according to that tradition, and they had books written in Latin. Models of correct latinity were available to the people who wrote the documents. It is known that the people who wrote documents also copied codices. So we may ask ourselves why was there such a difference between the linguistic models of the time and the scribes' practices.

This discrepancy between models and practices corresponds to a distinction made in cultural anthropology between representational models and operational models: this means that what a culture does or thinks it does, and what it actually does may be – from the point of view of an

external observer – different things. In the 60's American sociolinguist William Labov detected this type of discrepancy between what informants thought they said and what they actually said, especially regarding sociolinguistic variables associated with social prestige.

What a culture does or thinks it does results from a representational model of their world, something which is more or less articulate and can be even verbalized, and is framed by a vision of the culture's past. This could be phrased like «We act the way we were taught to act» or «We do this the same way our forefathers did before us».

What a culture really does is based in implicit models which guide behaviour in certain situations and settings, and which are not necessarily conscious, i.e. cannot be verbalized. So when a culture states «We act the way we were taught», what they are actually doing may be noticeably different from the behaviour they think they are replicating. An outside observer can perceive this, but someone belonging to the culture, who is completely in synch with the culture's historical development, will not be able to notice any changes or differences.

Medieval scribal models of correctness belonged to a representational level of the culture, and had an ideological dimension, whereas the actual practices tended to meet the contemporary demands for communicative realism, and did not correspond isomorphically to the inherited models and norms.

Even if the scribes and their contemporaries perceived the enormous differences that existed between the Notarial language and say the language of the Scriptures, nevertheless they would have no qualms about stating that what was being written was “*latinus*”, Latin being simply a convenient label for the written language in general.

After the Gregorian Reform in the late 11th century the Latin of the charters seems for a while, or in certain centres, to become more correct in spelling and grammar: what happened was not the sudden “Restoration of Latinity” as Menéndez Pidal thought, but merely a profound readjustment of the practices to the models, especially at the level of orthography and morpho-syntax, a readjustment which would lead in time to a split between the written and the spoken languages (what Michel Banniard calls «*la rupture de la communication verticale*»).

When Romance orthographies emerged in the early 13th century in Iberia a new conception of the relationship between written and spoken language must have been in place: the conceptual distinction between Latin and Romance was first a distinction between written and

oral modes of language, and only later it became a full distinction between two separate languages.

We can say that the distinction between Latin and Romance as separate entities occurred at the representational level, changing the representational models, which explains why in the late 12th and 13th centuries there was a need for new spelling systems for the Ibero-Romance languages; these languages were now perceived as different things from Latin.

As for the ancient trends of orthographical vulgarisation, present in the earliest known charters, they were simply internal aspects of the development of the Latin tradition, and so they affected the operational models but not the representational models of literacy.

### **3. Graphemic polymorphism in notarial Latin**

The development of the Latin-Portuguese notarial language was based on a balance between the need to write traditionally, and the need for communicative realism (what Francesco Sabatini called the «esigenze di realismo comunicativo»).

One of the most striking aspects of Latin-Portuguese texts, and one could say this of medieval texts in general, is the lack of graphemic uniformity. This lack of uniformity, which is reflected in the existence of observable and quantifiable variation patterns, has erroneously been ascribed to the ignorance of the scribes.

The innovative spellings that appear in the earliest documents seem to signify a pressure towards greatest scribal awareness of specific features of the vernacular; on the other hand, the Latinate context in which these innovative spellings occur show that this awareness did not imply a conceptual distinction between Latin and Romance as separate entities.

Latinate spellings and Romance spellings coexisted and alternated in a stable situation of graphemic polymorphism, which showed no major changes the 9th to 11th centuries.

Polymorphism was an intrinsic aspect of the notarial tradition: one must recognise this in order to ascertain the importance of the texts for the History of the Portuguese language, and also to avoid the creation of mythological entities, such as Leonese Vulgar Latin, which Menéndez Pidal thought he had found in 10th century texts from the Kingdom of León, or the myth of a Latin/Romance diglossia in the Middle Ages.

Polymorphism resulted from the history of a written tradition which spanned many centuries in linguistic communities subject, as all linguistic communities are, to continuous language change, and without a strong and centralised tradition of linguistic standardisation.

The gap between the Latin tradition and the everchanging vernacular brought about by centuries of language change gave rise to a complex relationship between the written and the spoken codes, in which variation and polymorphism stand out.

Notarial documents, as legal acts, had a direct impact in the lives of people. It was crucial that the texts recorded faithfully the intentions and obligations of the parties involved in a legal act. The documents thus encoded communicative contents that should be delivered and received unambiguously by all people attending the reading, both literate and illiterate: this would have been especially true of the dispositive sections of the documents.

Now if Notarial Latin was read aloud with Classical Latin pronunciation, the texts would have become unintelligible for monolingual speakers of Old Portuguese. If the reader just spelled the texts out, mapping each letter or graph to a phonic unit, as we do today when we read Classical Latin, the result would sound like gibberish to an Old Portuguese native speaker.

This strain between traditional standards of writing and the demands for communicative realism resulted in time in the gradual creation and development of alternative spellings. These new spelling conventions, which did not displace for several centuries the old Latin conventions, were generally more transparent and isomorphic in grapho-phonemic terms.

In time, if no spelling reform happens, all alphabetic orthographies become logographic, i.e. graphemically opaque. This poses no problem to the mature users of the system but puts a burden on learners, who have to memorise and learn to recognise thousands of whole-word patterns rather than a few dozen letter-to-sound correspondences.

#### **4. Logography vs. phonography in notarial Latin**

Experimental research in the mental processing of spelling in several modern orthographies based on the Roman alphabet leads to the conclusion that grapheme-phoneme correspondence rules belong to the internal structure of the writing system but not to the mental processes involved in reading and understanding written language.

Grapheme-phoneme correspondence rules seem to have no psychological existence or relevance in lexical access to a mature fluent reader. This is not to say that these rules do not

exist, but simply to state that they are used as subsidiary strategies by children, or by learners in general. Fluent readers read holistically, i.e. they recognise whole words as single shapes, with no internal analysis by means of grapho-phonemic interface. Fluent readers read words as logograms, and lexical access is done directly, from written form to mental lexicon.

This explains how Latin-looking written forms could be used to represent Old Portuguese phonology, and how the written inflectional morphology, such as noun cases, could be used in writing to represent the morphology of languages with no morphological case, like the Ibero-Romance languages.

Many Latin morphological endings would simply have no phonetic realisation, and in an extreme view, that I favour, certain morphemes could be interpreted holistically and be given a completely different phonetic representation than that that the spelling might suggest. I call this the “logographic hypothesis”.

One would write UOLUERIT but would pronounce [ki'zɛr] in Old Portuguese, or [ki'zɛr] in Old Castilian; one would write UXOR but would pronounce [mo'λɛr] in Old Portuguese, or [mu'zɛr] in Old Castilian, one would write ALIUS, -UM, -O but would pronounce ['ou̯tro] in Old Portuguese, or ['otro] in Old Castilian, and so on. A Latin-Romance pronunciation was also possible, and there are in fact spellings that attest to this: UOLUERIT could be pronounced [vo'leɾ] in Old Portuguese, UXOR could be pronounced ['u̯ɔɾ] in Old Portuguese, ALIUS, -UM, -O could be pronounced ['aλo] in Old Portuguese.

These more learned pronunciations could be required in situations such as grammar classes, reading among literate people, the reading of formulaic sections of documents, or reading in solemn circumstances, as the liturgy. The *lectores* had to be able to assign the right pronunciation to the right context.

## 5. Contemporary functionality of notarial texts - acceptability and interpretability

Notarial documents, being the written embodiment of legal acts, were validated, when they were read aloud in a special setting: validation required the presence of the parties involved, of the *lector* (which could also be the *notator* but not necessarily) and of witnesses. A legal act was not in effect just by being written down, it had to be read aloud and understood by all parties concerned. The charter could then be stored as proof that the act had taken place.

Thus, notarial documents were written in order to convey a communicative content in a specific communicative context, which included the presence of illiterate people.

Contemporary acceptability of a text depends on a combination of intra-textual, inter-textual and extra-textual factors. Interpretability of a text depends, not so much on intra-textual aspects (i.e. aspects of the structure, layout and contents of the text) but rather on the set of beliefs and expectations that the reader or hearer has, as a member of a textual community. Interpretability depends crucially on the assumptions that a textual community has about itself and about the texts that circulate in its midst.

We can say, in this sense, that *every and each text in every and each culture is based on shared expectations that exist in the textual community concerning what a text is and how it should look like (or sound like if read aloud); each and every text must fit into an accepted model or genre of textual production*, and the notarial texts were no exception. Their structure and their language fitted into what the textual community of the time considered adequate and expectable.

This means that we must try to grasp the actual patterns of textual production that were learnt and used by medieval scribes, instead of applying our own conceptions of latinity, correctness, grammar, style, composition, logic, etc.

Literacy, as anthropologists know, and many linguists have come to know also, is not a universal category: there is no such thing as a single autonomous concept of literacy. Literacy practices cannot be isolated from other practices of a given culture. Medieval literacy was not modern literacy. And because it was not we must make the effort to integrate medieval literacy in the set of beliefs, knowledges, prejudices, and social and cultural practices that existed and functioned in medieval times: it was by their set of beliefs and standards, not by our own, that medieval people received and accepted the texts of the time.

Both philologists and linguists must strive to grasp and understand the discourse conventions that were in place in medieval society: only then can philologists and linguists approach the textual data and interpret the texts as linguistic source-materials.

\*\*\*

These are, I believe, some of the issues whose adequate study allows us to understand notarial documents without devaluating, decontextualizing ou distorting them.

## Conclusion

The Latin-Romance language in Iberia attests to an unbroken chain of slow orthographic vulgarisation from the earliest texts until the 13th century.

If we just look at the late 12th and early 13th centuries, we can note that the development of different Ibero-Romance orthographies happened in a time of profound cultural and political change (in some of the Christian Kingdoms we can even speak of political upheaval). But if we look at the larger picture, the development of the Romance orthographies began centuries earlier, and resulted in the continued de-latinisation of the Late Latin tradition.

Traditional Romance Philology has failed to answer several simple questions:

- how was the Latin-Romance tradition able to stay in a relatively stable state throughout so many centuries, mysteriously or systematically (depends on your point of view) presenting the same errors and deviations over and over again?
- why was there not some sort of linguistic censorship against the «barbarous» Latin of the notaries? (many legal acts were performed in the presence of high Church officials and in centres of learning where there were libraries and schoolmasters — the absence of censorship continued even after the Reform when there were many French clergymen in Portuguese monasteries and bishoprics; there were even Bishops of French origin acting as witnesses or confirming legal acts);
- why did the notaries insist on learning Latin so poorly, when the Latin language was all around them in its purity, in the *auctoritates* (both pagan and Christian), in the Scriptures, in the liturgy, in the sermons, in the lives of saints, in grammatical treatises, etc.?
- why are the deviations of the notaries not random but systematic? which leads us to:
- why do the scriptographic deviations of the notaries fit into well-defined patterns of variation?
- why can we find in the same *scriptorium*, the same document, and sometimes the same paragraph or line alternative graphemic forms for the same word, morpheme or phoneme? i.e. why did they get their Latin so wrong when they themselves proved that they knew the correct spellings?

The attempt to provide adequate answers to these questions, and many others that Notarial Latin raises, forces us to look at the texts as a whole, as texts, as speech acts, as products of a culture with different literacy standards and practices from our own modern cultures.

To look at the Latin-Portuguese texts as valid communicative acts and as valid speech acts (in themselves and to their medieval users) is to give them the status and the dignity of linguistic documents, and to enable their analysis in a well-grounded and well-defined linguistic, cultural and anthropological framework.