Encoding Medieval Abbreviations for Computer Analysis (from Latin–Portuguese and Portuguese Non-literary Sources)¹

Stephen R. Parkinson  
University of Oxford, UK

António H. A. Emiliano  
Universidade Nova de Lisboa, Portugal

Abstract

This paper proposes a solution to the problem of handling scribal abbreviations in TEI-conformant transcriptions of medieval texts, following a conservative editorial strategy. A key distinction is drawn between alphabetic abbreviations, which represent sequences of letters, and logographic abbreviations which represent whole words. The TEI elements <expan> and <abbrev> can be used systematically to separate these two types: alphabetic abbreviations will be expanded in the main text, recording the abbreviated form (including TEI entities representing the main abbreviation marks) as an attribute of <expan>, while logographic abbreviations will be represented in their abbreviated form, with the expanded form recorded as an attribute of <abbrev>. The proposals are illustrated from common abbreviations and short text samples from tenth-century Latin–Portuguese and thirteenth-century Old Portuguese.

1 Introduction

Although there is currently much activity in the area of compilation of archives of medieval Portuguese texts, and, to a lesser extent, of Latin–Portuguese texts, discussion of the bases of encoding have been more limited. In many cases the discussion has simply been a rerun of previous discussions on principles of transcription (normas de transcrição), or has focused on questions such as part-of-speech tagging. This paper presupposes the choice of the TEI subset of the SGML coding system as a basic framework for text encoding, and proposes a set of principles and specific devices for solving one particular problem in the encoding of medieval texts, namely the treatment of abbreviations.²

Section 2 outlines a conservative style of transcription in which it is imperative to distinguish between two types of abbreviations, alphabetic
abbreviations and logographic abbreviations. Section 3 provides specific encoding methods for these two types of abbreviation, with extensive exemplification. In the particular case of alphabetic abbreviations, we make explicit three solutions for encoding them: the choice of how much information to encode and the particular way to encode it lies at all times with the editor. The proposed encoding system is illustrated by parallel transcriptions—palaeographical and TEI-conformant—of short extracts from a Latin–Portuguese document from the tenth century, and an Old Portuguese document from the late thirteenth century.

2 Problems in the Transcription and Editing of Medieval Primary Sources (Portuguese and Latin–Portuguese)

To transcribe a medieval manuscript text for publication always amounts to transliterating it, i.e. converting the original writing system into another system based on different conventions and graphemic principles. The linguistic study of medieval documents requires a special care in transliterating the original spelling conventions with the utmost faithfulness. This does not mean that an edition should aim at replacing facsimile reproductions by presenting with photographic precision all the minutiae of the manuscript, nor is it proposed that every letter-shape distinction should be rendered accurately in print, although this would not present an unsurmountable problem, given the possibility of creating special computer fonts to handle both on screen and in print the fundamental distinctions in letter-shape that any handwritten script presents. Such material aspects of the texts and of their supports will always demand a direct inspection of the manuscript or at least of a good reproduction of it. What is involved in transcription and encoding is the preservation and rendition of the basic elements of the linear and segmental structure of the script and of its layout in the original support: aspects that derive from the linguistic and textual intentions of the notaries and scribes, taking into account that not every letter-shape distinction presents a significant graphemic value or meaning. For the encoder who is attempting to make textual and graphemic structures explicit, the availability of several solutions for the same problem is one of the great advantages of SGML.

This philological perspective of preserving and observing the textual and orthographical data as they were produced by the material authors of the texts entails that no attempt should be made to modernize, regularize or correct the original spellings and forms of the manuscripts. The consequence of such an approach to medieval writing is a very conservative type of edition, one that should be both minimally interventionist and minimally interpretative, leading to a better understanding of the question of the linguistic reality of medieval documents and of its relationship to the language of the period.
The treatment of abbreviations is a clear case where special attention is required: brachygraphy was an important component of medieval writing, and must have figured prominently in scribal training and scribal competence. Many written elements (words, morphemes, syllables, groups of letters, and single letters) were frequently and consistently spelled in abbreviated form, and in many cases were used almost only in abbreviated form. Therefore, to transliterate mechanically an abbreviation as a sequence of discrete characters, an operation that the scribes chose not to do (or learned not to do), cannot avoid distorting the original graphic physiognomy of the texts.

In this respect, we make a clear distinction between alphabetic and logographic abbreviations (see Parkinson, 1983). An alphabetic abbreviation consistently represents a letter or sequence of letters, regardless of the word in which it is located; a logographic abbreviation represents a whole word (or possibly the root of an inflected word, the different forms of the same word constituting a family of related abbreviations).

A typical case of an alphabetic abbreviation is the superscript letter 'a', and its later developments as a superscript diacritic; these both represent the sequences ‘-ra-’ or ‘-ar-’ in words such as *carta* ‘letter, deed’ (written ‘cªta’), *Maria* (written ‘Mªia’), *Martim* (written ‘Mªim’), *Braga* (written ‘Bªga’), *laurador* ‘farmer’ (written ‘lauªdor’), etc. Some alphabetic abbreviations have a wide range of possible expansions, notably the ‘general abbreviation mark’ represented by a horizontal line, typically indicating nasal consonants, the vowel letter ‘e’, the sequence ‘-en-’, and others.

Logographic abbreviations, illustrated in Fig. 1, involve the removal of a number of letters from the whole expanded form (a process known as suspension) and the addition in most cases to this abbreviated form of a general-purpose abbreviation sign. The sequences of letters and abbreviation marks indicate the complete individual words in ways not derivable from normal abbreviatory conventions; their interpretation is often context dependent, as in the case of morphological endings of words in Latin–Portuguese texts, or multifunctional abbreviations such as [the Tironian sign 9]. Indeed, given the long lifespan of many logographic abbreviations, used first in Latin texts and subsequently in Portuguese

<table>
<thead>
<tr>
<th>dd</th>
<th>corresponds to Lat. <em>denarios</em> / Port. <em>dinheiros</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>pps</td>
<td>corresponds to Lat. &amp; Port. <em>presentes</em></td>
</tr>
<tr>
<td>Po</td>
<td>corresponds to Port. <em>Pedro</em></td>
</tr>
<tr>
<td>ss</td>
<td>corresponds to Lat. <em>solidos</em> / Port. <em>soldos</em></td>
</tr>
<tr>
<td>dco</td>
<td>corresponds to Lat. <em>dicto</em> / Port. <em>dito</em></td>
</tr>
<tr>
<td>sca</td>
<td>corresponds to Lat. <em>santa</em> / Port. <em>santa</em></td>
</tr>
<tr>
<td>xpo</td>
<td>corresponds to Lat. <em>christo</em> / Port. <em>cristo</em></td>
</tr>
</tbody>
</table>

Fig. 1 Logographic abbreviations.
texts, there need not even be a direct correspondence between letters in the abbreviation and letters in the expansion. These abbreviations reflect an important characteristic of scribal competence, where lexical items were related to minimum sequences of letters with no graphemic decoding interface: i.e. logographic abbreviations were used holistically without any internal letter-to-phoneme analysis.

Many transcription systems fail to make this distinction, and work on the assumption that expansion of abbreviations consists entirely of the addition of letters, so that any letters present in the abbreviated form must be retained in the expanded form, as exemplified by Fig. 2 (examples drawn from Portuguese texts).

Our choice is to allow the expansion of alphabetic abbreviations, as part of the process of transcription, but to transliterate logographic abbreviations without expansion. It goes without saying that all expansion of abbreviations should be explicit and recoverable: this contrasts with the traditional practice of expanding all abbreviations, with the presumed missing letters supplied by the editor, in italics or between parentheses, or without any indication that they are not in the manuscript (this latter option being especially common in editions made by and for historians).

Other aspects such as word separation and punctuation show how the perception and segmentation of the structure of a written utterance were distinct from what happens in modern Western scripts and texts. All too frequently modern editors remark on the chaotic character of medieval punctuation, and proceed to replace the original punctuation with modern punctuation, thereby destroying nexus, divisions, meanings, emphases, and textual structures that the medieval scribe deliberately inserted in his text or copy.

Current editions of medieval Portuguese documents pose several problems to language researchers, as a result of the absence of a common standard of transcription. The main problem is that some of the existing editions (besides the fact that they may contain erroneous readings) are based upon editorial criteria that are not very explicit and are not suited for the linguistic, philological, and diplomatic study of the documents. Editions may also be made for different purposes and by people who pursue different goals: the resulting editions often reflect the interests of the editors. Thus, an edition made by an historian for the purpose of historical research will typically be of little use for a linguist, who will then find him/herself forced to go back to the manuscript and make a new edition more adequate for the linguistic study of the texts. This is particularly true for all existing editions of Latin–Portuguese documents.

---

S. R. Parkinson and A. H. A. Emiliano

---

Fig. 2 Logographic abbreviations expanded by addition of letters.
3 Encoding Medieval Abbreviations using TEI Principles

The treatment of abbreviations presented here is based on the distinction between alphabetic and logographic abbreviations, discussed in Section 2. Alphabetic abbreviations fall into the following types:

- abbreviations with suspension of a single letter with either a special sign or a general-purpose sign of abbreviation;
- abbreviations with suspension of a continuous group of two or three (compare the case of ‘con-/com-’) letters with a special sign of abbreviation; these abbreviations correspond to a number of very frequent sequences of letters, among others, the endings ‘-us’, ‘-os’, and the groups ‘ue’, ‘en’, ‘er’, ‘re’, ‘ar’, ‘ra’, etc.

The following types of logographic abbreviations are found:

- abbreviations with suspension of a discontinuous group of two or more letters, with a single general-purpose sign of abbreviation in most cases;
- abbreviations with suspension of a continuous group of more than two letters, with a single general-purpose sign of abbreviation in most cases;
- abbreviations with conservation of the first and last letters of the word and suspension of the rest of the letters, with a single general-purpose sign of abbreviation in most cases;
- Roman numerals (figures and dates).

According to the principles established in Section 2, alphabetic abbreviations should generally be expanded editorially, whereas logographic abbreviations should be conserved. SGML allows various ways of implementing this general decision, by virtue of allowing significant quantities of annotation, which is accessible through SGML-conformant viewers but which is not part of the text as automatically displayed. We refer to the text displayed (i.e. the content of SGML elements) as visible text, and the annotation (the attributes of those elements) as hidden text.

The TEI conventions provide two basic elements for transcribing abbreviated text: `<expan>` (expansion), which locates the expansion of an abbreviation in visible text, and stores the representation of the abbreviation in hidden text, and `<abbr>` (abbreviation), which locates the abbreviated form in visible text and stores its expansion or interpretation in hidden text (see Robinson (1994) for extensive discussion).

These two devices may be exemplified using the example of the Portuguese word `parte` ‘part’ as usually abbreviated using the superscript ‘a’ diacritic over the ‘p’, usually transcribed ‘p(ar)te’. Assuming the availability of an SGML entity ‘&a-sup;’ giving a platform-independent representation of the ‘a’ symbol, the word could be represented in one of two ways (here and in the following examples, visible text is presented in bold):

```
p<abbr expan='ar'> &a-sup; </abbr> te or p<expan abbr='&a-sup;'>te &lt;expan&gt;ar</expan>te
```

Similarly, the much-used ‘q’ abbreviation for the conjunction `que`, made up of the letter ‘q’ with a vertical stroke above (provisionally represented by the entity ‘&dash;’), could be coded:
We assume a catalogue of entities representing the central components of Latin and Latin–Portuguese abbreviations, similar to that recently proposed by Driscoll (2000) for the Latin-based abbreviations of Old Icelandic. This would itself be linked to an online database of abbreviations and graphic devices, based in the Portuguese case on the authoritative catalogue of abbreviations published by Nunes (1981). The establishment of a definitive inventory of abbreviations is a prerequisite of any transcription system, and the creation of a standard set of entities to encode them is a basic requirement of SGML-based systems, as well as providing a minimally interpretative type of transcription. The extent to which such entity lists can be generalized across different scribal traditions is still an open question.

Most commentators present the choice between `<abbr>` and `<expan>` elements as one of editorial preference. Our proposal is to use the two options to distinguish systematically between the two types of abbreviations identified above.

### 3.1 Encoding of alphabetic abbreviations

Alphabetic abbreviations, if editorially expanded, will be coded by using the `<expan>` element (so that editorial expansions form part of visible text), with the attribute ‘`abbr`’ containing the SGML entity representing the abbreviation, as illustrated in Figs 3a and 4a. To make the editorial

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>ENCODED TRANSCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>cor(um)</td>
<td>co&lt;expan abbr='&amp;r-vbar;'&gt;r(um)&lt;/expan&gt;</td>
</tr>
<tr>
<td>p(er)</td>
<td>&lt;expan abbr='&amp;p-ubar;'&gt;p(er)&lt;/expan&gt;</td>
</tr>
<tr>
<td>monagor(um)</td>
<td>monago&lt;expan abbr='&amp;r-vbar;'&gt;r(um)&lt;/expan&gt;</td>
</tr>
<tr>
<td>p(er)manent</td>
<td>&lt;expan abbr='&amp;p-ubar;'&gt;p(er)&lt;/expan&gt;manent</td>
</tr>
<tr>
<td>insup(er)</td>
<td>insu&lt;expan abbr='&amp;p-ubar;'&gt;p(er)&lt;/expan&gt;</td>
</tr>
<tr>
<td>p(ro)</td>
<td>&lt;expan abbr='&amp;p-stroke;'&gt;p(ro)&lt;/expan&gt;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>ENCODED TRANSCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>cor(um)</td>
<td>co&amp;r-vbar;</td>
</tr>
<tr>
<td>p(er)</td>
<td>&amp;p-ubar</td>
</tr>
<tr>
<td>monagor(um)</td>
<td>monago&amp;r-vbar;</td>
</tr>
<tr>
<td>p(er)manent</td>
<td>&amp;p-ubar;manent</td>
</tr>
<tr>
<td>insup(er)</td>
<td>insu&amp;p-ubar;</td>
</tr>
<tr>
<td>p(ro)</td>
<td>&amp;p-stroke;</td>
</tr>
</tbody>
</table>

Fig. 3 Latin–Portuguese alphabetic abbreviations encoded with (a) the element and (b) abbreviation entities.

Encoding Medieval Abbreviations for Computer Analysis

intervention clear, the expansion of the abbreviation will be transliterated between parentheses '( )', in accordance with traditional practice. If unexpanded, the abbreviation is transcribed simply as the relevant entity, with no interpretative coding, as illustrated in Figs 3b and 4b. This latter solution provides a minimally interpretative and interventionist

4 The <expan> element makes the editorial intervention recoverable, but it is only explicit when the text is viewed in 'tags on' mode. The alternative traditional means of indicating expansion of abbreviations, italicization of the expanded text, can be incorporated by use of the SGML formatting elements <hi> or <rend>, as developed by Driscoll (2000, p. 89); we reject these solutions, as we reserve the use of formatting elements for indicating the appearance of the original text.

Fig. 4 Old Portuguese alphabetic abbreviations encoded with (a) the element 'expan' and the attribute 'abbr', and (b) abbreviation entities.
type of annotation; it provides the user with a less heavily annotated text, but leaves him at the same time with the burden of interpreting each occurrence of an abbreviation in its particular context.

There is a need to distinguish between abbreviations whose basic form is that of a superscript letter, and the more general (and optional) superscripting of letters or sequences of letters, as is common in numerals and some logographic abbreviations: in the former case, the name of the entity concerned is the explicit marker of superscription; in the latter it will be directly represented in the encoded text by the SGML-conformant formatting element \texttt{<sup>}.

The representation of abbreviations can be enriched by internal reference to the associated database of abbreviations and graphic devices (p. 350), using one or more \texttt{id} attributes.

3.2 Encoding of logographic abbreviations

Logographic abbreviations will not be expanded in visible text. Only those letters that are actually present in the source should be transcribed, as the textual content of the \texttt{<abbr>} element, with a suggested expansion as part of the ‘\texttt{expan}’ attribute’. Figures 5 and 6 present examples of logographic abbreviations drawn from the complete texts of which short extracts are presented in the Appendix. The transcription system will need to distinguish between abbreviations incorporating a general-purpose abbreviation sign and their otherwise identical counterparts where the abbreviation is implicit. In these and subsequent illustrations the presence of the general-purpose abbreviation sign is indicated by placing the abbreviation between curly brackets ‘{ }’; to maximize the use of TEI-conformant devices, the presence or absence of the general abbreviation mark sign could equally well be handled inside the \texttt{<abbr>} element by adding a ‘\texttt{type}’ attribute, as illustrated in Fig. 7 (Portuguese examples).

Unexpanded logographic abbreviations are encoded without the ‘\texttt{expan}’ attribute. This solution, apart from simplifying significantly the annotation system and process, would be particularly well suited to diplomatic editions made for the graphemic study of the texts, where no expansion of the abbreviation is in fact required or of use. This minimal encoding makes explicit just the presence of an abbreviation and its character-content, as illustrated by Fig. 8, using examples explained more fully in Fig. 9.

Logographic abbreviations with a general-purpose sign are the rule. However, in some cases the abbreviation sign is omitted, as the examples in Fig. 7 show.

The \texttt{<expan>} element can be nested inside an \texttt{<abbr>} element, as can be seen by the examples in Fig. 9, which combine in the same form a logographic and an alphabetic abbreviation. If one chooses not to expand alphabetic abbreviations, simple abbreviation entities can be combined with the \texttt{<abbr>} element.

Logographic abbreviations are thus never expanded in visible text; the expanded form, if included, is always given as hidden text. The form given in hidden text is not a true expansion (such as the ones provided
### Table 1: Encoding Medieval Abbreviations for Computer Analysis

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>ENCODED TRANSCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>xpe</td>
<td><code>&lt;abbr expan=&quot;christe&quot;&gt;{xpe}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>xpi</td>
<td><code>&lt;abbr expan=&quot;christi&quot;&gt;{xpi}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>dni</td>
<td><code>&lt;abbr expan=&quot;domini&quot;&gt;{dni}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>lhu</td>
<td><code>&lt;abbr expan=&quot;hesu&quot;&gt;{lhu}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>kls</td>
<td><code>&lt;abbr expan=&quot;kalendas&quot;&gt;{kls}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>ma</td>
<td><code>&lt;abbr expan=&quot;mea&quot;&gt;{ma}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>mla</td>
<td><code>&lt;abbr expan=&quot;millesima&quot;&gt;{mla}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>nne</td>
<td><code>&lt;abbr expan=&quot;nomine&quot;&gt;{nne}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>nsi</td>
<td><code>&lt;abbr expan=&quot;nostri&quot;&gt;{nsi}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>nsm</td>
<td><code>&lt;abbr expan=&quot;nostrum&quot;&gt;{nsm}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>pbtr</td>
<td><code>&lt;abbr expan=&quot;presbyter&quot;&gt;{pbtr}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>r+o</td>
<td><code>&lt;abbr expan=&quot;roboro&quot;&gt;r&lt;mark type='cross'&gt;{signum}&lt;/mark&gt;o&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>sca</td>
<td><code>&lt;abbr expan=&quot;sancta&quot;&gt;{sca}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>sce</td>
<td><code>&lt;abbr expan=&quot;sancte&quot;&gt;{sce}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>sci</td>
<td><code>&lt;abbr expan=&quot;sancti&quot;&gt;{sci}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>scis</td>
<td><code>&lt;abbr expan=&quot;sanctis&quot;&gt;{scis}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>spu</td>
<td><code>&lt;abbr expan=&quot;spiritu&quot;&gt;{spu}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>ts</td>
<td><code>&lt;abbr expan=&quot;testis&quot;&gt;{ts}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>ul</td>
<td><code>&lt;abbr expan=&quot;uel&quot;&gt;{ul}&lt;/abbr&gt;</code></td>
</tr>
</tbody>
</table>

**Fig. 5** Latin–Portuguese logographic abbreviations.

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>ENCODED TRANSCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>dca</td>
<td><code>&lt;abbr expan=&quot;dita&quot;&gt;{dca}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>dco</td>
<td><code>&lt;abbr expan=&quot;dito&quot;&gt;{dco}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>gra</td>
<td><code>&lt;abbr expan=&quot;graça&quot;&gt;{gra}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>ihns</td>
<td><code>&lt;abbr expan=&quot;iohnes&quot;&gt;{ihns}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>lbs</td>
<td><code>&lt;abbr expan=&quot;libras&quot;&gt;{lbs}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>port</td>
<td><code>&lt;abbr expan=&quot;portugal&quot;&gt;{port}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>rodrig</td>
<td><code>&lt;abbr expan=&quot;rodriguiz&quot;&gt;{rodrig}&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>Staren</td>
<td><code>&lt;abbr expan=&quot;Santaren&quot;&gt;{Staren}&lt;/abbr&gt;</code></td>
</tr>
</tbody>
</table>

**Fig. 6** Old Portuguese logographic abbreviations.
by traditional scholarly editions), but rather an interpretation of the abbreviation: in Medieval Portuguese dco and dto are logographic signs for the same word form, and so they can have the same expansion value of dito ‘aforesaid’. The same applies to sco and sto corresponding to santo ‘saint’. In the cases that the same abbreviation is used in Latin and Portuguese texts it will have different expansion values. It should, however, be noted that a clearcut distinction between Latin and Romance (Portuguese) is a matter of dispute among scholars, many of whom see a Latin–Romance continuum extending up to the thirteenth century, rather than a sharp division or opposition between the two languages.

The expansion of Portuguese logographic abbreviations implies a high degree of regularization regarding the supplied letters, and therefore presents in some cases a very different form from traditional editorial expansions; e.g. mar (expan = ‘martim’) is expanded by different editors as ‘mar(ti-)’, ‘mar(tim)’, and ‘mar(tin)’; st (expan = ‘st(euam)’) is expanded by different editors as ‘st(eua-)’, ‘st(euam)’, and ‘st(evam)’;

---

**ABBREVIATION** | **ENCODED TRANSCRIPTION**
--- | ---
P | <abbr expan='Pero' sign='y'>{P}</abbr>
port | <abbr expan='portugaeses' sign='y'>{port}</abbr>
port | <abbr expan='portugal' sign='y'>{port}</abbr>
purt | <abbr expan='pertugal' sign='y'>{purt}</abbr>
Po | <abbr expan='Pedro' sign='n'>Po</abbr>
P’ | <abbr expan='Pero' sign='n'>P&amp;sup;</abbr>
Port | <abbr expan='Portugal' sign='n'>Port</abbr>

---

**ABBREVIATION** | **ENCODED TRANSCRIPTION**
--- | ---
(con)f. | <abbr>&con;f.</abbr>
(con)fr | <abbr>&con;fr</abbr>
(con)frmo | <abbr>&con;frmo</abbr>
dgn(us) | <abbr>{dgn&amp;us;}</abbr>
dign(us) | <abbr>{dign&amp;us;}</abbr>
scissim(us) | <abbr>{scissim&amp;us;}</abbr>
kld(rum) | <abbr>{kld&amp;r-vbar;}</abbr>
nnib(us) | <abbr>{nni&amp;b-col;}</abbr>
tprib(us) | <abbr>{tpri&amp;b-col;}</abbr>

---

Fig. 7 Alternative encoding of Old Portuguese logographic abbreviations.

Fig. 8 Alternative encoding of unexpanded logographic abbreviations.
this type of editorial difference is not recorded in the encoding that we propose. As the `<expan>` element should be meant to allow word search and word listing in a semi-lemmatized manner, its content cannot and does not reflect actual scribal spelling practices when the abbreviated form was written in full in the text; these interpreted forms generally preserve the letters actually written in the logographic abbreviation (especially the first letter), whereas the supplied letters correspond to an ideal regularized form, which is close in many cases to modern spelling conventions or corresponds to a more or less standardized form from the early or mid-fourteenth century in the case of Portuguese.

Notable cases of divergence between the abbreviated form and the expanded form are abbreviations such as

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Transcription with embedded abbreviation expanded</th>
<th>Transcription without expansion of embedded abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(con)f.</td>
<td><code>&lt;abbr expan='confirmo/confirmat'&gt;confirmo&lt;/expan&gt;</code></td>
<td><code>&lt;abbr expan='confirmo/confirmat'&gt;confirmat&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>(con)fr</td>
<td><code>&lt;abbr expan='confirmo/confirmat'&gt;confirmo&lt;/expan&gt;</code></td>
<td><code>&lt;abbr expan='confirmo/confirmat'&gt;confirmat&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>(con)frmo</td>
<td><code>&lt;abbr expan='confirmo'&gt;confirmo&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='confirmo'&gt;confirmat&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>dgn(us)</td>
<td><code>&lt;abbr expan='diagonus'&gt;diagonus&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='diagonus'&gt;diagonus&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>dgn(us)</td>
<td><code>&lt;abbr expan='diagonus'&gt;diagonus&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='diagonus'&gt;diagonus&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>scissim(us)</td>
<td><code>&lt;abbr expan='sanctissimus'&gt;sanctissimus&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='sanctissimus'&gt;sanctissimus&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>kld(rum)</td>
<td><code>&lt;abbr expan='kalendarum'&gt;kalendarum&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='kalendarum'&gt;kalendarum&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>nnib(us)</td>
<td><code>&lt;abbr expan='nominibus'&gt;nominibus&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='nominibus'&gt;nominibus&lt;/abbr&gt;</code></td>
</tr>
<tr>
<td>tprib(us)</td>
<td><code>&lt;abbr expan='temporibus'&gt;temporibus&lt;/abbr&gt;</code></td>
<td><code>&lt;abbr expan='temporibus'&gt;temporibus&lt;/abbr&gt;</code></td>
</tr>
</tbody>
</table>

Fig. 9 Latin–Portuguese logographic abbreviations incorporating alphabetic abbreviations
which are functionally true logograms. In forms like these, one can in fact consider that any expansion is superfluous (and misleading regarding the true nature of scribal practices), and so the encoder may well choose not to give an expanded form:

```
  dd   <abbr>[dd]</abbr>
  ppss <abbr>[ppss]</abbr>
```

This latter solution, with the advantage previously mentioned of being minimally interpretative, is ideally suited for graphemic analysis; nevertheless, it produces a more difficult text to read for less palaeographically trained readers. The encoder must weigh these advantages and disadvantages according to the goals that s/he aims at in his/her edition. It must be stressed that whereas more interpretative elements imply a more heavily annotated text, non-interpretative (i.e. lighter) encoding will reduce the accessibility of textual and graphemic structures.

4 Conclusions

Medieval writing and medieval scribal practices must be grasped within the cultural and intellectual framework where they developed and functioned: they must not be evaluated in terms of criteria that will decontextualize them culturally and chronologically. The principles of transcription and encoding presented here will allow such study of medieval document to proceed within SGML-encoded corpora. The coding system systematically incorporates the key distinction between alphabetic and logographic abbreviations, while providing alternative types of encoding corresponding to different types of transcription.

The establishment of a central list of abbreviations can also shed light on the vexed question of the continuity of Latin and Portuguese. It is interesting to note that most of the abbreviations that occur in the Portuguese text also appear in a tenth- or eleventh-century Portuguese–Latin texts. Our system of encoding recognizes this, in that the same elements are used in both corpora: it is the expansion of the abbreviation in invisible text that indicates whether an abbreviation corresponds to a Latin or a Portuguese form. It is also noteworthy that, although the Portuguese text is significantly larger than the Latin text, it shows a smaller number of distinct logographic abbreviations. Bearing in mind that many abbreviations in Portuguese texts are identical to Latin abbreviations, this enterprise reveals an interesting aspect of the evolution of the Latin–Portuguese tradition of writing into the Portuguese tradition, which has so far passed unnoticed in the literature.

References

Encoding Medieval Abbreviations for Computer Analysis


Appendix: Sample Texts

Notes on transcriptions:

(1) Palaeographical transcriptions:

expanded abbreviations are transcribed between ( ),
logographic abbreviations are transcribed between { },
editorial word separations are indicated by ‘-’,
superposed characters such as the tilde ~ and the plica /H11032/ are transcribed following the character that they accompany.

(2) Encoded transcriptions:

words divided at line ends are joined on the earlier line, with the point of division marked by ‘-’,
editorial word separations are indicated by ‘=’ for full words and by ‘= =’ for clitics.
1 Latin–Portuguese document (Concelho de Feira, 977)

Source: Arquivo Nacional da Torre do Tombo, Sé de Coimbra, maço 1, doc. 5
Type: original private document in semi-cursive visigothic hand
Notary: Inuenando
Subject: Penedruia donates several land-holdings to the Monastery of S. João-de-Ver
Sample: lines 9–13

1.1 Palaeographical transcription
L09 et do_pro remedium anime mee et pro uictum et uestimentum .
monago(rum) fradrum {ul}_sororum qui Ibidem In_seruicio
p(er)ma
L10 nent et non damus ei_licitum ad_nullo omine~ proInde aligo
deuindigare nisi ad_ipsius domnis {scis} . nodum die quod est X
tei
L11 maii era {mla} X
c. V. si_quis tamen qod fieri non_creditis aliquis
ohmo ueneri ad_Inrunpenu~ contra unc factum nsm Inrumpere
temtaueri
L12 Inprimis siat exconmunigadus ad_corpus . et_sanguinis {dni} {nsi}
{lbu} [xpi] et_con Iuda traditore parte susciptat In eterna
damnatjo
L13 ne et numqua~ finemda et_insup(er) pariat ipsum quod sursun
resonat . dubladum factum est anc series testamentum

1.2 Encoded transcription
<text>
<body>
... <line n="09"> et do= pro remedium anime mee et pro uictum et
uestimentum <punct>.
<punct> monago<expan abbr="/r-
vbar">r(um)</punct> <expan fradrum <abbr expan="/u">ul</abbr>
</expan> sororum qui Ibidem In= seruicio <expan
abbr="/u">ul</abbr> <punct>ma
L10 nent et non damus ei_licitum ad_nullo omine~ proInde aligo
deuindigare nisi ad_ipsius domnis {scis} . nodum die quod est X
tei
L11 maii era {mla} X
c. V. si_quis tamen qod fieri non_creditis aliquis
ohmo ueneri ad_Inrunpenu~ contra unc factum nsm Inrumpere
temtaueri
L12 Inprimis siat exconmunigadus ad_corpus . et_sanguinis {dni} {nsi}
{lbu} [xpi] et_con Iuda traditore parte susciptat In eterna
damnatjo
L13 ne et numqua~ finemda et_insup(er) pariat ipsum quod sursun
resonat . dubladum factum est anc series testamentum

S. R. Parkinson and A. H. A. Emiliano

<line n='12'> In primis siat excommunigadus ad corpus <line n='13'> et = sanguinis <line n='14'> et = abbr expan = domini > [dni] <line n='15'> abbr expan = nostri > [nsi] <line n='16'> abbr expan = ihesu > [Ihu] <line n='17'> abbr expan = christi > [xpi] <line n='18'> abbr = con Iuda traditore parte susciap In eterna damnatjo-ne <line n='19'> et numqua~ finemda et = insu<line n='20'> abbr = ’&q-obar;’ <line n='21'> pariat ipsum quod sursun resonat <line n='22'> et = abbr expan = anc series testamentum <line n='23'> <line n='24'> <line n='25'> <line n='26'> <line n='27'> <line n='28'> ... </line>
</text>

2 Medieval Portuguese document (1289)
Source: Arquivo Nacional da Torre do Tombo, Gaveta 1, maço 1, no. 11
Notary: Domingos Eanes of Santarem
Subject: The widow and daughter of D. João de Portel transfer the Castle of Portel and the patronage of the church to King Denis
Sample: lines 39–46

2.1 Palaeographical transcription
L39 e o bispo e os ome´e’s bo´o’s de suso ditos dissero~ e dero~lhis por conselho. q(ue) desse~
L40 a´a’s (dit)as donas a´ a`uença e o canbo q(ue) eles teyna~ por be~ e se o as donas q(ui)sessem outorgar
L41 ou caber o canbo ou a´ a`uença q(ue) eles por be~ teuessem e q(ue) gardassem e comp(ri)ssem a nosso senhor elRey
L42 a condiço~ assj como fora posto. feyto o estrumento na era e no dya de suso dito e eu
L43 Domi~gos i(o)h(a)n(e)s publico Tabellio~ de Santare~ a rogo dos de suso ditos do~ Martin
L44 gil & de Lourenço scola destas cousas de q(ue) p(re)sent~ tej este publico estrumento p(re)sen
L45 te fuj & aqueste estrumento con ma ma~o p(ro)sp(r)i a escreuj & antralinhey tres diço~es co~ue~
L46 a saber e q(ue) gardassem & este meu sinal y pugi q(ue) tal e

2.2 Encoded transcription
<text>
<body>
...<line n='39'> e o bispo e os ome´e’s bo´o’s de suso ditos dissero~ e dero~lhis por conselho. <expan abbrev='&q-obar;' id='Q1a'>q(ue)</expan> desse~<line>

S. R. Parkinson and A. H. A. Emiliano

*a* = *a’s*<abbr expan='ditas'>{dias}</abbr> donas

*a* = *a’uença* e o canbo<abbr expan abbrev='&q-obar;' id='Q1a'>

q(ue)<</expan> eles teyna~ por be~ e se o as donas<expan abbrev='&q-i;' id='Q5'>q(ui)</expan> sessem outorgar</line>

<line n='41'> ou caber o canbo ou a’ = a’uença<expan abbrev='&q-obar;' id='Q1a'>

q(ue)</expan> eles por be~ truessem<add place='supralinear'>e</add> <expan abbrev='&q-obar;' id='Q1a'>

q(ue)</expan> gardassem</add>:> e con<expan abbrev='&p-i;' id='P5'>p(ri)</expan> semen a nosso senhor el_= = Rey</line>

<line n='42'> a condiço~ assj como fora posto. feyto o estrumento na era e no dya de suso dito e eu</line>

<line n='43'> Domi~gos<abbr expan='iohnes'>{ihns}</abbr> publico Tabellio~ de Santare~ a rogo dos de suso ditos do~ Martin</line>

<line n='44'> gil & de Lourenço scola destas cousas de<expan abbrev='&q-obar;' id='Q1a'>

q(ue)<</expan><expan abbrev='&p-dash;' id='P3'>p(re)</expan> sente fuj este publico estrumento<expan abbrev='&p-stroke;' id='P2'>p(ro)</expan> a escreuj & antralinhey tres diço~es co~ue~</line>

<line n='45'> a saber e<expan abbrev='&q-obar;' id='Q1a'>

q(ue)<</expan> gardasem & este meu sin—al (signum) y pugi<expan abbrev='&q-obar;' id='Q1a'>

q(ue)</expan> tal e</line>

<line n='46'>

... </body> </text>