**BERMUDEZ — Using**

<0 figures>

**Using Feature Structures for the Study of Medieval Manuscripts**

Sabel, H. B.

Scholars would agree on the importance of doing a comprehensive linguistic study of every witness that integrates a manuscript tradition. However, such a detailed research on large traditions can be rather laborious.

A possible alternative to limit the object of study is, to my mind, to examine the most important linguistic phenomenon, namely linguistic variation. By focusing on all the language variants found after collating every witness, we report on the distinctive linguistic features of every manuscript, thus making the study of a manuscript tradition much more practicable (Roudil, 1986).

For a period such as the Middle Ages, which featured rather flexible writing codes, textual variation can be very informative. Taking a corpus of formal features as a reference and comparing it with other texts, for which there is a high degree of certainty about the conditions in which they were written, can actually provide useful information regarding their writing standards, that is, the *scriptorium*.

Linguistic variants usually occur as a result of the self-dictation that the copyist does after reading and memorizing the extract he intends to copy, hence reflecting features of the his idiolect. The study of those features contributes to gain a better understanding of the witness production, since it provides evidence of the human resources who participated in this production.

Copy errors are another gold mine, and their systematic classification and study provides information related to the sources, distinctive features of the model (e.g., a common graphic substitution in the copy means that the two graphemes were very similar in the model), and linguistic data, among others.

Therefore, in order to compile all these data I propose a digital edition based on the collation of every reading of every witness an edition that is applied to the Galician-Portuguese secular lyric.1

After using regular expressions and XSLT to create a preliminary interlinear collation, I manually corrected certain elements, such as the *scripta continua* found in some parts of certain witnesses or the accurate markup of lacunae and transpositions. I then transformed again the transcription, applying for this purpose a collation style sheet that I created and that tokenizes and compares each element generating an apparatus tagset that records every variant. This encoding is done according to the Text Encoding Initiative module ‘Parallel Segmentation’ method, a feature of the *TEI Guidelines* that allows multiple versions of a text to be encoded in one document.

Once all variants were marked up, I followed the TEI module ‘Feature Structures’ to categorize them. I created a very detailed feature-value library, in which every variant has its correspondent <fs> element defined by five to ten individual features. This kind of combination make it possible to query the corpus with high precision. The retrieval of information is thus very efficient, mainly because I can use the same language, XQuery, to query and transform the data (Siegel and Retter, 2015).

As has been mentioned, this encoding and its manipulation provides, on one hand, the description of the various graphic systems used, by decoding the spelling rules present in the witnesses. A graphemic analysis will not only help to identify phenomena of language variation, but it will also clarify certain elements of the genesis of these manuscripts, such as graphic substrates, the origin of the copyists, and the textual sources that were used.

On the other hand, these data will serve to analyse the linguistic variation in the corpus, at the phonetic, phonological, morphological, syntactic, and lexical levels. This information can be subsequently used to conduct a descriptive study, explaining the main features of the Galician-Portuguese grammar of the High and Late Middle Ages, concluding in a general study of historical linguistics, taking drawing examples from this corpus.

Finally, scholars specialised in this literary tradition will find a very useful resource to prepare scholarly editions thanks to the thorough linguistic information and the informative collation.

**Note**

1. For an initial approach, the *Cantigas Medievais Galego Portuguesas* [online database] offers all the *cantigas* (songs) contained in the medieval Galician-Portuguese *cancioneiros* (songbooks), the manuscripts’ images, and also the music—both the medieval and the contemporary versions.

**References**

**Lopes, G. Videira, et al.** (2011). *Cantigas Medievais Galego Portuguesas* [online database]. Lisbon: Instituto de Estudos Medievais, http://cantigas.fcsh.unl.pt.

**Roudil, J.** (1986). *Summa de los nueve tiempos de los pleitos. Édition et étude d’une variation sur un thème par Jean Roudil*. Klincksieck, Paris.

**Siegel, E. and Retter, A.** (2015). *eXist. A NoSQL Document Database and Application Platform.* O’Reilly, Sebastopol.

**TEI Consortium (eds)**. (n.d.). *TEI P5: Guidelines for Electronic Text Encoding and Interchange.* 2.7.0. http://www.tei-c.org/Guidelines/P5 (last modified date: 16 September 2014).