

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. дфн В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Linguistic Variation and Manuscript Transmission: A Case Study Using XML/TEI

Helena Bermúdez Sabel

Digital edition, linguistic variation, diachronic variation, Galician-Portuguese, medieval lyric

The union of textual criticism with information technology has been a very popular topic in academic discussions, and there are a considerable number of resources available for the medievalist. Some of these will be used to prepare a very specific digital edition which will shed light on the manuscript transmission of medieval Galician-Portuguese poetry. For this purpose, the edition will focus on linguistic textual variation since it can be crucial in determining the origin of manuscript transmission.

1. The object of study

The importance of Galician-Portuguese secular poetry lies in its status as the first cultural production written in this Romance variety. This is why scholars must make an effort to understand how this cultural patrimony came to exist, and all the social factors that led to part of this production being preserved for 800 years. Despite the extensive literature devoted to the study of secular Galician-Portuguese poetry, many questions about the genesis of the manuscript witnesses which have survived to the present day still remain unanswered.

A comprehensive study of all the manuscript witnesses seems mandatory for the study of their cultural transmission, thus answering crucial questions such as the origin of the copyists and the *scriptorium* or *scriptoria* involved, as well as the different sources which they used. Such a detailed research can be rather laborious. A possible alternative to limit the object of study is, to my mind, to examine the most significant linguistic phenomenon, namely linguistic variation. By focusing on all the linguistic variants found after collating every witness, we will report on the linguistic features of the manuscripts, thus making the study of manuscript tradition much more practicable.

In the case of Galician-Portuguese secular poetry, two factors facilitate this analysis: the small number of witnesses preserved and the low level of variation between them. Three *cancioneiros* (lyrical anthologies) are preserved: the *Cancioneiro da Ajuda* (A), the *Cancioneiro da Biblioteca Nacional de Lisboa* (B) and the *Cancioneiro da Biblioteca Vaticana* (V). There are also some minor fragments: the *Pergamiño Vindel* (R), the *Pergaminho Sharrer* (T), and two later copies of a *tensó* by the troubadour Afonso Sanchez, M and P. Therefore, this limited number of manuscripts, along with the limited signifi-

cant variation between the witnesses, makes the Galician-Portuguese secular poetry an ideal topic for an analysis such as the one presented here¹.

2. Methodology

This project deals with textual variation in the manuscript transmission of Galician-Portuguese secular lyric through the development of a digital edition of a corpus compiled according to diachronic criteria. This edition is based on the collation of encoded versions of every reading of every witness.

For this study of linguistic variation through a digital edition all texts are encoded according to TEI's parallel segmentation method to encode multiple versions of a text in a single XML document. To compare the texts we resort to the *Versioning Machine*, a software tool designed for displaying and comparing different versions of literary texts and conceived in 2000 by Susan Schreibman. The *Versioning Machine* is very convenient since it provides a frame to compare diplomatic versions of witnesses side by side, which makes the collatio much easier. Its documentation also provides numerous samples and explanations, which helps the encoding process.

Although linguistic variation is the main focus of this study, every type of textual divergence is identified—that is, any mistakes made during the copying of any of the witnesses, including those of the graphematic variety, and the equipollent readings. In other words, not only linguistic variation phenomena are highlighted, but we will also identify and categorize the typology of every textual variant, and therefore, some new tags not present in TEI must be developed so as to organize all this information.

At this point, the advantages of XML will become apparent again, since there are many tools we can use to extract and publish the data. The programming language XQuery is flexible enough for analysis since it allows an effective extraction of information for subsequent presentation.

3. Objectives

In the first place, this project involves a digital palaeographic transcription of the witnesses of the Galician-Portuguese secular poetry. This will fill a gap in the current Galician-Portuguese publishing field. The main goal is to develop a synoptic edition of the secular songs of the Galician-Portuguese tradition, using the TEI protocol which allows information retrieval for further study.

This encoding will allow for the description of the various graphic systems used in the copying, possibly revealing spelling rules present in the witnesses. A rigorous graphematic analysis will help to identify phenomena of language variation and, at the same time, clarify certain elements of the genesis of these manuscripts, such as graphic substrates, the origin of the copyists and the textual sources used.

¹ Digitized manuscripts are available at the online database *Cantigas Medievais Galego Portuguesas* (<http://cantigas.fcsh.unl.pt>), whose images have in many cases a better quality than the print facsimiles.

These data will allow for analysis of the phonetic, phonological, morphological, syntactic, and lexical variation in the corpus. This important step is also needed to conduct a descriptive study of Galician-Portuguese, explaining the main features of the language's grammar at the time (the High and Late Middle Ages) and concluding in a general study of historical linguistics, taking examples from this corpus.

Finally, scholars specialising in this literary tradition will find a very useful resource to prepare scholarly editions thanks to the detailed linguistic information created and the informative collatio.

References

- Lopes, Graça Videira et al. 2011. *Cantigas Medievais Galego Portuguesas [online database]*. Lisbon: Instituto de Estudos Medievais. <<http://cantigas.fcsh.unl.pt>>.
- Schreibman, Susan. *The Versioning Machine* <<http://v-machine.org>>.
- TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 2.6.0. Last modified date: 2014-01-20 <<http://www.tei-c.org/Guidelines/P5>>.