DH Poetry Modelling: a Quest for Philological and Technical Standardization.

María Gimena del Rio Riande, Elena González Blanco García y Clara Martínez Cantón.


Dirección estable: https://www.aacademica.org/gimena.delrio.riande/8

ARK: https://n2t.net/ark:/13683/pdea/swF

Esta obra está bajo una licencia de Creative Commons. Para ver una copia de esta licencia, visite https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es.

DH Poetry Modelling: a Quest for Philological and Technical Standardization

1. Introduction

Standardization has become an increasingly important process in relation to academic research, as it provides a better way for exchanging information. Humanities and cultural studies have followed, however heterogeneous path in which creativity and tradition play an essential role. Comparative studies in literature, and especially poetry are a clear example of this ecletic situation. There is not a uniform academic approach to analyze, classify or study the different poetic manifestations. Things get even worse when comparing poetry schools from different languages and periods. The result of this uncoordinated evolution is a bunch of varied terminologies that means to explain analogous metrical phenomena through the different poetic systems whose correspondences have been hardly studied (González-Bianco and Solé, 2014).

The existing digital poetic repertoires and databases are a good example of this situation, as they constitute a rich kaleidoscope of multilingual virtual postconstituted by lyrical collections in French (Nouveau Nanteau, Italian (BedT), Hungarian (RPHA), Medieval Latin Corpus Rhythmorum Musicum, Annalecta Hymnica Digitalis, Pedecerdo Gallego-portuguese Oxford Cantigas de Santa María, MedDEP Castillian (ReMeCa), Dutch (Dutch Song Database), Occitan (BedT: Poésie Neotroubadouresque, The last song of the Troubadours), Catalan (Repertori d’obres en ver). Skaldic (Skaldic Project), or German repertoires Lyrick des Minnesänger, among others.

Interoperability among these different poetic corpora would be desirable, as having a common search engine to extract information from all of these databases at the same time would have a deep impact for comparative studies in literature, linguistics and other humanities disciplines. Were, however, far from this reality as interoperability is not possible due to a lack of standardization both in technological and philological fields.

2. Philological standardization

During the Middle Ages and the Renaissance, the powerful influence of Latin made scholars inherit the terminology of Classical poetry treatises and apply it to Romance languages, regardless of their different linguistic traits and verse structures. When vernacular theories started to arise, each literary school set up its own terminology and classification system. This multiplicity led to complex situations, such as the creation of conceptual genres that only exist in some traditions.

Spanish conceptualization models are a good example to illustrate this situation: the classical system of Bello (1955), first published in 1835, divided all the structures into binary and ternary feet (imitating the classical Latin terminology):

<table>
<thead>
<tr>
<th>Binary</th>
<th>Ternary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trochaic: óo</td>
<td>Dactylic: óoo</td>
</tr>
<tr>
<td>Iambic: oó</td>
<td>Amphibrachic: ooo</td>
</tr>
<tr>
<td>Anapestic: oo</td>
<td></td>
</tr>
</tbody>
</table>

Although unified criteria is translated into many different uncoordinated technologies when research data are transformed to build digital projects and do not even follow a standard, in most cases. The multiplicity of technologies includes SQL databases, TEI and XML markup languages, semantic web technology standards (RDF, SKOS), natural language processing systems (NLP) and visualization tools.

Relational databases have been deeply used by the first digital poetic repertoires combining an ER (Entity-Relationship) model, together with the data model based on records for the logical implementation (Elmasri and Navathe, 2011, 27-ss). The problem of representing ER composition model is that the result shown is data centered, but it is not enough to mark textual items that need to be analyzed from a metrical point of view.

There are other projects based on XML solutions, as TEI has a specific module for poetry analysis, “è”, with a rich set of tags to describe metrical schemas, rhymes, accentual structure and syllabic varieties. However this model is not widely used by the different projects, and the lack of philological unified criteria makes it difficult to translate literary schemas into XML tags, making researchers create new tags or express nuances with customized attributes for each project.

The key for interoperability both in philological and technological fields is a common reference system, for which semantic web technologies are a powerful solution. Building a linked data model by adding a semantic layer of metadata to the existing databases does not alter their internal structure. This solution requires, however, to assume unified criteria on the philological model that serves as a reference.
What we propose in this paper is not a new method for analyzing poetry but an abstract model based on a working methodology supported by a double standardization system, both at philological and technological levels. In relation to this aspect, it must be highlighted that it needs to be carried out by an interdisciplinary and coordinated team, which requires a careful design of data architecture in different levels. Our proposal aims to set up a procedure to combine philological criteria to map vocabularies and concepts which might have common means and properties in the different traditions and to insert them into an abstract framework in which each of these elements can fit as individuals of an ontology which gathers the main poetics concepts shared by most traditions. We have worked on some first approaches in this sense, building our first ontology prototype, based on our ReMetCa Spanish project www.purl.org/net/remetca (González-Blanco, and Del Río, et al., 2014), populated with a controlled vocabulary in SKOS, that can be found in http://vocabularios.caicyt.gob.ar/pmc/ (http://vocabularios.caicyt.gob.ar/pmc/). These preliminary results, which served as prototype and a basis to build the current model, have been applied to different poetry projects, such as the edition of the Cancionero de Baena http://sade.textgrid.de/exist/apps/SADE/Dialogo_Medieval/index.html (http://sade.textgrid.de/exist/apps/SADE/Dialogo_Medieval/index.html) and the Starting Grant ERC-2015-STG-679528 POSTDATA.

This paper has been developed thanks to the research projects funded by MINECO and led by Elena González-Blanco: Acción Europa Investigaría EUIN2013-50630: Repertorio medieval www.purl.org/net/remetca (http://www.purl.org/net/remetca) and the TEI Conference 2014 http://tei.northwestern.edu/files/2014/10/gonzalez­blanco­20wnj3q.pdf (http://tei.northwestern.edu/files/2014/10/gonzalez­blanco­20wnj3q.pdf) on the description of poetic collections in http://www.poetriae.linhd.es/ (http://www.poetriae.linhd.es/) but they need to be expanded and improved to analyze other poetic systems.

This paper has been developed thanks to the research projects funded by MINECO and led by Elena González-Blanco: Acción Europa Investigación EUI2013-50630: Repertorio Digital de Poesía Europea (DIREPO) and FFI2014-57961-R. Laboratorio de Innovación en Humanidades Digitales: Edición Digital, Datos Enlazados y Entorno de Investigación para el trabajo en humanidades, and the Starting Grant ERC-2015-STG-679528 POSTDATA.