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

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

DH Home (http://www.dh2016.adho.org) / Abstracts (/abstracts/) / 73 (/abstracts/73)

| Show info | How to cite | XML Version (/static/data/426.xml) |

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, a heterogeneous path in which creativity and tradition play an essential role. Comparative studies in literature, and especially poetry, are a clear example of this eclectic 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-Blanco and Sélaf, 2014).

The existing digital poetic repertoires and databases are a good example of this situation, as they constitute a rich kaleidoscope of multilingual virtual poetry constituted by lyrical collections in French (*Nouveau Naetebus*), Italian (*BedT*), Hungarian (*RPHA*), Medieval Latin (*Corpus Rhythmorum Musicum, Annalecta Hymnica Digitalia, Pedecerto*), Gallego-portuguese (*Oxford Cantigas de Santa María, MedDB2*), Castilian (*ReMetCa*), Dutch (*Dutch Song Database*), Occitan (*BedT, Poèsie Neotroubadouresque, The last song of the Troubadours*), Catalan (*Repertori d'obres en vers*), Skaldic (*Skaldic Project*), or German repertoires (*Lyrik 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. We are, 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):

Binary	Ternary
Trochaic: óo	Dactylic: óoo
Iambic: oó	Amphibrachic: oóo
	Anapestic: ooó

Later, the musical analysis system applied by Navarro Tomás (1956) was followed as a valid system through many years, using concepts like *anacrusis*. In the last years, there have been many different approaches to explain the Spanish panorama, as it is shown by the semantic comparative model designed by the Czech Oldrich Bělič.

The international context is richer, especially in English, with two prominent schools: 1) A traditional approach based on stress and classical feet; and 2) A generative approach based on the terminology and concepts shown through text grids that take into account word boundaries, with a strong impact on poetic theories (Gerber, 2013: 147).

The models described are just an example of the idiosyncrasy that can be observed in each literary tradition. Although the current ICT infrastructures are prepared to harvest different types of collections and models, it is necessary at a first stage to standardize metadata and map vocabularies and terminology at the philological level in order to build a consistent able to be shared between the different traditions.

3. Technological standardization

The lack of 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 used includes SQL databases, TEI and XML markup languages, semantic web technology standards (RDF, OWL, 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, "verse", 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.

Although semantic web technologies have had success in archives libraries and museums (group known as LODLAM, http://lodlam.net/ (http://lodlam.net/)), its application to poetic corpora is very different, as there are only a couple of studies dealing with some of the above mentioned aspects (Bootz and Szoniecky, 2008 and Zöllner-Weber, 2009), but there is not a standard conceptual model of ontology referred to metrics and poetry. The closest works related to this topic are probably the conceptual model of CIDOC, the controlled vocabularies of English Broadside Ballad Project (http://ebba.english.ucsb.edu/ (http://ebba.english.ucsb.edu/)) and the linked data relations offered by the Library of

Congress (http://id.loc.gov/ (http://id.loc.gov/)), which do not offer enough information on metrics vocabulary There are also interesting computational approaches which use automated linguistic analysis or text mining, based on the morphological and phonetic structure of each language. Results have been impressive, as one of the greatest advantages is the speed of the analysis of big amounts of text (Gervás, 2000). Nevertheless, the integration of these technologies with the previous models described is not easy, and solutions are often customized for the variation of natural language used, most times standard English.

4. Our approach

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 (http://www.purl.org/net/remetca) (González-Blanco, and Del Rio, et al., 2014), populated with a controlled vocabulary in SKOS, that can be found in http://vocabularios.caicyt.gov.ar/pmc/ (http://vocabularios.caicyt.gov.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 *Cancionero de Baena* http://sade.textgrid.de/exist/apps/SADE/Dialogo_Medieval/index.htm (http://sade.textgrid.de/exist/apps/SADE/Dialogo_Medieval/index.html) or 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.

Bibliography

1. **Bělič, O.** (2000). *Verso español y verso europeo: introducción a la teoría del verso español en el contexto europeo*, Santafé de Bogotá, Instituto Caro y Cuervo.

2. **Bello, A.** (1955). *Estudios filológicos. Principios de la ortología y métrica de la lengua castellana y otros escritos en Obras completas* de Andrés Bello, Caracas, Ministerio de Educación. (1st ed. 1835)

3. **Bootz, P. and Szoniecky, S.** (2008). Towards an ontology of the field of digital poetry, *Paper presented at Electronic Literature in Europe.* http://elmcip.net/node/415 (http://elmcip.net/node/415) Accessed: 30/10/2015.

4. **Burnard, L. and Bauman, S.** (Eds.) TEI P5: Guidelines for Electronic Text Encoding and Interchange. Ver. 2.5.0. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/) Accessed 30-10-2015.

5. **Gerber, N.** (2013). Stress-Based Metrics Revisited: A Comparative Exercise in Scansion Systems and their Implications for Iambic Pentameter. *Thinking Verse III*, pp. 131-68.

6. **González-Blanco, E., G. Del Rio, C. Martínez, M. Martos** (2014). When TEI Verse becomes linked data: using TEI tags as a model to build a linked poetry system. *Abstract from the paper presented at the TEI Conference 2014*, http://tei.northwestern.edu/files/2014/10/gonzalez-blanco-20wnj3q.pdf (http://tei.northwestern.edu/files/2014/10/gonzalez-blanco-20wnj3q.pdf)

7. **González-Blanco, E. and L. Seláf** (2014). Megarep: A comprehensive research tool in medieval and renaissance poetic and metrical repertoires. In Soriano, L. et al. (Eds.) *Humanities on the web: the medieval world.* Oxford, Peter Lang, pp. 321-32.

8. **González-Blanco, E. and J. L. Rodríguez** (2013). ReMetCa: a TEI based digital repertory on Medieval Spanish poetry at The Linked TEI: Text Encoding in the Web, *Book of Abstracts - electronic edition*. Edited by F. Ciotti and A. Ciula, DIGILAB Sapienza University and TEI Consortium, Rome 2013, pp. 178-85. http://digilab2.let.uniroma1.it/teiconf2013/abstracts/ (http://digilab2.let.uniroma1.it/teiconf2013/abstracts/) Accessed 30-10-2015.

9. **Navarro Tomás, T.** (1991). *Métrica española*, Barcelona, Labor. (1st ed., 1956).

10. **Zöllner-Weber, A.** (2009). Ontologies and Logic Reasoning as Tools in Humanities? *Digital Humanities Quarterly*, **3**(4). http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html (http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html) Accessed: 30/10/2015.