

Towards a common poetry model: challenges and solutions

Bermúdez-Sabel, Helena (1); Díez Platas, María Luisa (1); Ros, Salvador (1); González-Blanco, Elena (2)

1: Universidad Nacional de Educación a Distancia, Spain; 2: Coverwallet, Spain

Abstract

This paper stems from the analysis of multiple poetic resources that were available on-line, as well as the results of methodological discussions with scholars of European Literature. The goal was to retrieve the informational needs of all these different sources in order to build a common data model for European Poetry. Thus, by implementing a reverse engineering method, we have created the Domain Model for European Poetry, which is an important milestone for making existent poetry resources interoperable. In this paper, we will present some of the challenges we encountered while conceptualizing the information relevant to poetic analysis and how we have worked around them.

1. Rationale

During the history of European Literature, there have been different cultural centers that have irradiated their influence. Some traditions, due to historic and socio-political reasons, have leaned at some point in their history on other literary models (Even-Zohar 1978, 48). Thus, the relations between the different literary traditions are many and heterogeneous. This poses some difficulties for literary research, since these relations are not always easy to trace. An additional handicap is that it demands for researchers to closely know traditions and languages other than the ones of their specialization and the accumulation of all that knowledge is not always humanly possible.

We can find many poetic resources on-line. However, the access to these resources is fragmentary: there is no way to retrieve all relevant information at once. Researchers need to look for multiple resources and then, for each one of them, carry out different queries in order to retrieve the required information.

To work around this problem, the project POSTDATA[1] has a proposal that depends on two key concepts: standardization and interoperability, according to the linked open data paradigm (LOD).

After presenting some brief notes about the objectives, this paper will focus on modeling issues.

2. Contextualization

Linked data must endorse a semantic model before being published. This semantic model can take different formats one of them being an ontology. Considering that one of the main aims of POSTDATA is to provide a means to publish European poetry (EP) data as LOD, this project is building an ontology for this domain. With ontologies, shared and distributed knowledge can be managed in such a way as to allow the integration of information from different data sets (Davies, Fensel, & Harmelen, 2003).

The starting point of the ontology construction was the analysis of different databases with contents related to one or more EP traditions in order to represent the informational needs of the community of practice, that is, the EP one. [2]

Our goal is to enhance interoperability between existent repertoires and to facilitate the creation of new poetry resources (González-Blanco et al. 2018). With such an ambitious objective in mind, we must be very exhaustive when eliciting the data needs of our target.

Our sources to draw out the informational needs of the EP community were, on the one hand, a representative sample of existing resources and, on the other hand, a survey that allowed us to consult the EP community. [3] In addition, there were different validation processes through which we received the direct input of experts in EP in order to refine the model. [4]

We are dealing with miscellaneous sources of information that incorporate data of multiple languages and cultures. This matter complicates the process of modeling. In the following section we will present some of the issues we encountered.

3. Modeling challenges

The creation of a data model that covers all required concepts to analyze any European poem causes some difficulties.

a) Multilingualism: The most obvious problem we ran into arises from working with a multilingual reality. The modelers had to analyze on-line resources in languages they are not familiar with.[5] This knowledge gap is covered with either the documentation translated to English by the project being analyzed, or with additional bibliography. Nevertheless, the direct contact with the people in charge of that resource is at times inevitable but the response and willingness to collaborate is, for the most part, very positive.

b) Polysemic terms: Occasionally, the difficulties are due to ambiguities in the same language. For instance, we find that many European languages have a term derived from the Latin *versus* to describe the poetic line. In English, however, the term “verse” can describe either the line of poetry, a bigger division like the stanza or the whole poetic composition (‘Verse’ 2011).

c) Synonymic (or quasi-synonymic) terms: Literature scholarship is a field with thousands of years, which means that some of the concepts we are analyzing have been defined for many centuries and from different perspectives. The domain experts of the team cannot prioritize any school of thought or theory. On the one hand, we may find different terms for similar concepts, but the use of one term over the other is related to philological schools. In these cases, the less aligned term is selected. On the other hand, we may find the same term in different technical vocabularies but with distinct meanings. For example, the term “dieresis” in syllabic verse traditions describes the separate pronunciation into two syllables of two sounds which usually

form one syllable ('Diérèse' 2017). However, in the quantitative verse, a "dieresis" expresses the pause that occurs when the end of a foot coincides with the end of a word ('Dieresis' 2011). Therefore, it is unavoidable to rank the suitability of certain terms since absolute neutrality is unattainable. [6]

d) Semantic interoperability: Like in any other process of semantic modeling, there is some tension between interoperability and semantics. For instance, poetry of the Western world is most commonly divided between qualitative and quantitative meter (Aroui/Arleo 2009, 11–12). Thus, meter may depend on the length of syllables and their distribution, or on the pattern created by stressed syllables coming at regular intervals. [7] Some of the types of qualitative verse have many attributes that are interoperable with the quantitative ones. Therefore, we decided to make a conceptual division between metrical schemes that depend on patterns and those that are defined by "counting" elements (such as counting how many syllables are there before the last stressed one). In this manner, little semantics are lost, because other properties make the distinction between qualitative and quantitative. However, with this conceptualization we enable the comparison between types of meter that, even if they focus on different linguistic properties, have many things in common.

The development of a data model that expects to serve a whole community of practice in the LOD ecosystem entails a great complexity. The type of final user that will consume that data is very diverse. Also, the applications that might be built with these data are many and very heterogeneous. This factor complicates the elicitation of the functional and no functional requirements, thus arising very interesting issues during the modeling process.

Acknowledgments

The authors would like to credit Mariana Curado Malta (Polytechnic of Oporto, Portugal) as the semantic modeler who engineered the method and its implementation.

We would also like to credit Clara Martínez Cantón (UNED, Spain) for her examination of the metrical concepts which are part of the data model referenced in this paper.

Finally, would like to thank the delegates of the analyzed repertoires for their availability and willingness to share information and to discuss issues related to their projects with the POSTDATA team.

Research for this paper has been achieved thanks to the Starting Grant research project Poetry Standardization and Linked Open Data: POSTDATA (ERC-2015-STG-679528) obtained by Elena González-Blanco. This project is funded by the European Research Council (<https://erc.europa.eu>) (ERC) under the research and innovation program Horizon2020 of the European Union.

Bibliographic references

- Aroui, Jean-Louis, and Andy Arleo. 2009. *Towards a Typology of Poetic Forms: From Language to Metrics and Beyond*. John Benjamins Publishing.
- Bermúdez-Sabel, Helena, Mariana Curado Malta, and Elena Gonzalez-Blanco. 2017. 'Towards Interoperability in the European Poetry Community: The Standardization of Philological Concepts'. In *Language, Data, and Knowledge: First International Conference, LDK 2017, Galway, Ireland, June 19-20, 2017, Proceedings*, edited by Jorge Gracia, Francis Bond, John P. McCrae, Paul Buitelaar, Christian Chiarcos, and Sebastian Hellmann, 156–65. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-59888-8_14.
- Curado Malta, Mariana, Helena Bermúdez-Sabel, Ana Alice Baptista, and Elena Gonzalez-Blanco. 2018. 'Validation of a Metadata Application Profile Domain Model'. *International Conference on Dublin Core and Metadata Applications*, 65–75.
- Davies, J., Fensel, D., & Harmelen, F. van. (2003). *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley & Sons.
- 'Diérèse'. 2017. In *Larousse. Dictionnaire de Français*. <http://www.larousse.fr/dictionnaires/francais>.
- 'Dieresis'. 2011. In *American Heritage® Dictionary of the English Language, Fifth Edition*. <http://www.thefreedictionary.com/verse>.
- Even-Zohar, Itamar. 1978. *Papers in Historical Poetics*. Tel Aviv: Porter Institute for Poetics and Semiotics.
- Gonzalez-Blanco, E., Ros Muñoz, S., Ruiz Fabo, P., Diez Platas, M. L., Bermúdez, H., Caminero, A., Martínez Cantón, C.I, Ayciriex, L. (2018). Poetry and Digital Humanities making interoperability possible in a divided world of digital poetry: POSTDATA project. Zenodo. <https://doi.org/10.5281/ZENODO.2203807>
- 'Verse'. 2011. In *American Heritage® Dictionary of the English Language, Fifth Edition*. <http://www.thefreedictionary.com/verse>.

Notes

[1] Poetry Standardization and Linked Open Data (POSTDATA) is an ERC-funded project. Please visit the project's website for more details: <http://postdata.linhd.es/>

[2] For a detailed exposition about how these informational needs were elicited and other methodological aspects, please see Bermúdez-Sabel, Curado Malta & González Blanco (2017).

[3] See the map available at <http://postdata.linhd.es/partners/> to see the projects that have collaborated with us. In Curado Malta et al. (2018) there is more information about all the resources that were analyzed and what type of study was done to each one of them.

[4] To learn about the validation processes, please see Curado Malta et al. (2018).

[5] The perfect team would have an expert on every poetic tradition, that is, a scholar for every European language and literary period. Regretfully, it is hard to find a project in Humanities with that type of resources.

[6] In the aforementioned issue with the word “diaeresis,” we selected that term to describe the type of pause and used “hiatus” to express the separation into two syllables, taking the term from its counterpart concept in Linguistics.

[7] In the case of qualitative meter, instead of demanding a fixed pattern of all the stresses, some traditions only care about the position of a certain stressed syllable, like the last one.