

---

# Representing vague and uncertain historical data on places people and events in historical texts - a case-study of historical texts of Dimitrie Cantemir

Cristina Vertan\*<sup>1</sup>

<sup>1</sup>Universität Hamburg – Germany

## Abstract

The massive digitization activities in recent years made available many historical works, as scans with corresponding metadata, or even completely transcribed and annotated in TEI. For the analysis of such texts, the annotation of people and places of great importance. For modern texts (newspapers, blogs), text mining and language technology processes are used to identify, annotate and link so-called "named entities" and to link them with registers and / or geo-coordinates.

For historical texts, the problem is much more complicated for several reasons:

- Some geographic positions are not or just partially known. The places e.g. no longer exist, so linking with geo coordinates is practically impossible;
- The names have been changed throughout history (due to political changes);
- For many places there exist metaphorical expressions according to several traditions;
- The historical maps are themselves unreliable;
- People often do not know birth and death information, or are not sure of this information. Their names can also vary, sometimes with metaphorical descriptions.

The situation for Eastern Europe is particularly bad, since political changes and language changes were frequent on the one hand, and cartographic materials and personal registers on the other hand are rare.

In this contribution we will explain how to build a semantic network of places and people for the European regions of the Ottoman Empire.

The materials used are historical works by the universal scholar Dimitrie Cantemir. We create a resource that can be used for hermeneutic and computer-aided analysis.

Dimitrie Cantemir (1672-1723) was one of the last universal humanists. He dealt with philosophy, musicology, history and cartography of Wallachia and Moldavia, of the Ottoman and Russian empires. He has written books in Latin, Romanian, Greek and Turkish. In Istanbul as well as in St. Petersburg and Moscow he was part of a large network of important people in the immediate vicinity of the court circles and enjoyed the friendship with the rulers.

His work "Historia incrementorum atque decrementorum Aulæ Othomanicæ" represents his opus magnum and marks the beginning of modern scientific studies about the Ottoman

---

\*Speaker

Empire. The book was completed in 1716 and then translated into English in 1734-1735. Due probably to the unavailability of the Latin original, this translation then provided the template for the French (1743) and German (1745) translations. Until the middle of 19th century this work has a major impact on the reception of the Ottoman Empire in the Western countries. Several researchers claimed that Cantemir's assertions are wrong or invented and claimed that he is quoting works he did not have access to and persons he did not meet. Due to the fact that a manual comparison of Latin manuscripts and translations in several languages is very difficult (especially to the high number of languages and scripts involved) until now no study about the reliability of his works was performed.

In project HerCore1 we investigate among others, the reliability of his historical assertions by:

- Investigating and comparing the text of Cantemir with the quoted sources
- Creating and analysing the network of people who built his network at the sultan's court
- Highlighting the linguistic expression of uncertainty and vagueness which stress often the doubt about the reported facts of the author himself

The core of the investigation and annotation process is a Fuzzy OWL Ontology<sup>2</sup>, trying to model the Ottoman Empire world in its administrative, social, geographical and religious facets. Collection and networking of historical places and people in the historical texts - an exemplary study of works by Dimitrie Cantemir

Following aspects need a particular attention:

The modelling of geographical respectively political entities. Political entities (e.g. countries) tend often to share name with some geographical entities. Political entities keep name but change often borders. Thus we considered as fix, unambiguous individuals the geographical elements which are visible nowadays. Historically attested geographical zones which do not exist are modelled as fuzzy concepts.

Political entities are defined as a sum of several historical contexts. We introduce additionally the concept of historical zone<sup>1</sup> in order to model concepts as "Europe" which from the point of view of the Ottoman Empire e.g. began at the border with Hungary, or "Balkan" which for the Ottoman Empire was represented by the Wallachia and Moldavia principalities.

The ontology is still under development, and contains for the moment: 350 Classes, 130 Object properties, and 2000 Individuals.

For dealing with multilinguality we attach to each individual the name used in German translation in the official Ottoman documents, as well as in the Romanian sources.

Time Intervals and geographical positions include fuzzy concepts which allow us to model uncertain dates and coordinates. The ontology is encoded in OWL 2. For the fuzzy extension we use first the Portege plug-in developed in (Bobillo et al. 2013). In this way the ontology is kept in a general accepted encoded format. The fuzzy properties, concepts and data types are described by means of annotations within the crisp ontology. In a second step the convertor provided by (Bobillo and Straccia 2015) is used to transform the OWL 2 ontology into a fuzzy OWL representation which can be processed by a fuzzy reasoner.

<sup>1</sup> The work is performed within the HerCoRe project (<https://www.inf.uni-hamburg.de/inst/dmp/hercore/projects>), financed by the Volkswagen Foundation (Güney et al 2019)

<sup>2</sup> <https://www.w3.org/TR/owl2-syntax/>

## References

(Bobillo et al 2013) Bobillo, Fernando and Delgado, Miguel and Gomez-Romero, Juan, "Rea-

soning in Fuzzy OWL 2 with DeLorean, in *Uncertainty Reasoning for the Semantic Web II*, Bobillo, F., Costa, P.C.G., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K.J., Lukasiewicz, Th., Nickles, M., Pool, M. (Eds.), *Lecture Notes in Artificial Intelligence*, Springer Verlag, (2013)

(Bobillo and Straccia 2015) Bobillo, Fernando and Straccia, Umberto, "fuzzyDL: An Expressive Fuzzy Description Logic Reasoner," <http://www.umbertostraccia.it/cs/software/fuzzyDL/download/old/documents/fuzzyDL.pdf> (last retrieved 29.03.2020)

(Cantemir 1771) Cantemir, Dimitrie, 1771, *Beschreibung der Moldau*, Faksimiledruck der Originalausgabe von 1771, Frankfurt und Leipzig

(Cantemir 1745) Cantemir, Dimitrie, *Geschichte des osmanischen Reichs nach seinem Anwachsen und Abnehmen*, 1745, Herold, Hamburg (Güney et. Al 2019) Güney, A. and Vertan, C. and von Hahn, W. "Combining hermeneutic and computer based methods for investigating reliability of historical texts", *Proceedings of the "Twin Talks" workshop collocated with DHN 2019*, Steven Krauer and Darja Fiser (Eds), University of Copenhagen 2019, <https://cst.dk/DHN2019Pro/TwinTalksWorkshopProceedings.pdf>, pp. 25-38, (2019)

(Vertan and v. Hahn 2014) Vertan, Cristina and v. Hahn, Walther, 2014, *Discovering and Explaining Knowledge in Historical Documents*, In: Kristin Bjnadottir, Steven Krauer, Cristina Vertan and Martin Wyne (Eds.), *Proceedings of the Workshop on "Language Technology for Historical Languages and Newspaper Archives" associated with LREC 2014*, Reykjavik Mai 2014, p. 76-80.