PARDONS, or the challenges of a multilingual crowdsourcing project on early modern 'true' crimes

Gert Gielis, State Archives Belgium

This short paper presents the PARDONS project (State Archives in Belgium, UCL, KULeuven & partners), which aims to disclose a large series of pardon letters kept in the Belgian State Archives for further research, with the help of crowdsourcing and DH tools, such as HTR.

In the late medieval and the early modern Low Countries a person convicted of serious crimes and risking severe punishment could request a pardon from the ruler. The pardon procedure required a thorough investigation of the criminal facts and thus left an abundant paper trail (the pardon letter or a copy thereof, witness reports, testimonials, correspondence...). As pardon requests stem from the entire Low Countries, we find documents in both Dutch and French. The account of the crime - homicide, fraud, religious offenses, abduction, misconduct in office, theft... - is often recorded in vivid detail. These 'true crime stories' of conflict, violence and disruption of daily life offer unique insights into local history, power structures, social relations, and the development of criminal law. These documents form a corpus that is among the most prominent early modern archive series in the Belgian State Archives, not only because of its significance for various disciplines and the richness of the narrative structures of these archives, but also because of the exceptionally long time span it covers.

The PARDONS project, therefore, provides for the digitization, metadating, valorization and research of this unique collection of pardon letters. To this end, the research team is calling on citizen scientists to help with the transcription of this extensive series of pardon letters. The transcriptions provided through crowdsourcing will be used to train HTR models in Transkribus. There is indeed an urgent need for 16th-century models for early modern Dutch. With such model(s), the rest of the corpus can then be transcribed automatically. The transcriptions, together with the images of the sources, will be disseminated on a website, providing access to these fascinating sources to a wide audience. The PARDONS project is a rare example of a bilingual (Dutch-French) crowdsourcing project in the humanities. The multilingualism of this project, which plays out on two levels, poses a major challenge.

Firstly, the source corpus is essentially bilingual. While most of the pardon letters in the archives of the Chamber of Accounts are predominantly in Dutch (DU: ca.90% - FR: ca.10%), that ratio is rather different in the archives of the Privy Council (DU: ca.25% - FR: ca.75%). Sporadically, in the latter archive, we also find sources in Latin and Spanish. These documents were drafted between the late 15th and late 18th centuries and were thus written in early modern Dutch and early modern French. These precursors of our modern languages differ so significantly in spelling, grammar and syntax that they can be considered foreign languages. Interpretation of specific words and of a text is very challenging for volunteers.

Secondly, the crowdsourcing part of the project is conducted in both Dutch and French. Recruitment, instruction, and communication... are done in both national languages.

Moreover, communicating about the status of the transcriptions requires working on a common working space (such as the Dutch platform Vele Handen), but this is not yet available in a bilingual format. How then can citizen scientists on both sides of the linguistic border collaborate?

In this short paper, I present the PARDONS project and the use of digital humanities within the context of this crowdsourcing project. In particular, I focus on the challenges of bilingualism, the solutions developed within the PARDONS project, as well as on opportunities to facilitate multilingual crowdsourcing projects.