Queste del Saint Graal: Textometry Platform on the Service of a Scholarly Edition

Lavrentiev, Alexei

Alexei.Lavrentev@ens-lyon.fr

UMR ICAR Université de Lyon / CNRS

Serge, Heiden

slh@ens-lsh.fr

UMR ICAR Université de Lyon / CNRS

Yepdieu, Adrien

Politecnico di Torino, student

In this poster/demo we will present an online scholarly edition of *La queste del saint Graal* (*The Quest for the Holy Grail*) based on a manuscript of Lyons public library (Lyon, BM, P.A. 77) and built in the Textometry platform (TXM). The particularity of this edition is that it combines rich paleographic and philological data (including digital photographs, various layers of transcription, translation in modern French and editorial notes) with advanced linguistic search and analysis tools provided by textometric software.

Despite some damage (torn miniature in the beginning, missing folios in the end), Lyons manuscript is considered to be one of the best witnesses of the *Queste del saint Graal*, a 13th century French novel, part of the famous "Arthurian" prose cycle. A well-known edition by Albert Pauphilet based on this manuscript was first published in 1923 and has been regularly reprinted ever since. However, this edition cannot be used as a trustworthy source of linguistic data, as it contains multiple corrections, and the readings of the primary source are not always accessible.

In the late 1990s Christiane Marchello-Nizia started working on a new edition, which was supposed to be closer to the witness, explicitly correcting only doubtless scribal errors and equipped with multiple tools to assist the reader in understanding the text and to explore various features of its language. This work was to a large extent inspired by the experience of the *Charrette* Project (Pignatelli & Robinson 2002). The first prototype of the edition was published on the web in 2002, and a completely new version was released in 2009. The edition is still under construction but its main components are already in place and can be accessed on the Web through an early TXM prototype.

These are the following:

- Old French text (108 000 words);
- digital facsimile of the manuscript (418 text columns);
- modern French translation;
- scholarly introduction;
- tools for textometric analysis.

The Old French text is established according to rigorous editorial principles stated in the Introduction. A particular feature of the edition is the respect of scribal punctuation. It is encoded according to the XML-TEI standard with special attention to linguistic data. All words are explicitly tagged and annotated for part of speech using the Cattex2009 tagset (Prévost&al, to be published). The text can be displayed and searched in several presentation forms: a "traditional" view (*vue courante*), close to the norms of French critical editions (Vielliard & Guyotjeannin 2001), a "diplomatic" view respecting linguistically significant graphical features of the manuscript (such as the absence of "phonetic" distinction between *u* and *v*), and a "facsimile" view where all noticeable graphical distinctions of the primary source are represented. For instance, medieval abbreviations are tacitly expanded in the traditional view, they are expanded but typographically highlighted (using italics) in the diplomatic view and they are represented by abbreviation marks in the facsimile view. The concept of three different views of the source text is based on the multi-level transcription model of the Medieval Nordic Text Archive (Haugen 2004). MUFI character codes (Haugen 2009) are used for "special" medieval characters in the facsimile view (such as abbreviation marks or letter variants). A MUFI compliant font (such as Andron Scriptor Web) needs to be installed on the system to display these characters correctly.

The three views of the Old French text, the translation and the photographs can be browsed

individually or side-by-side in two columns. The 'diplomatic' and 'facsimile' views, as well as the translation, are only available for a few folios at present but they will be progressively edited for the whole text.

The innovative aspect of this new edition consists first of all in the integration of textometric research tools. Those tools assist the reader by offering qualitative services like full text search engine to build KWIC concordances and browse the text through them, and quantitative services like comparing statistically the occurrences of different linguistic phenomena in various parts of the text or analyzing statistically their collocational attraction inside various contexts, such as sentences.

One can open the textometric panel by clicking on the *Outils* ('tools') button in the bottom of the window. The platform makes it possible to search for linguistic data (words, parts of words, parts of speech, phrases, etc.) in any presentation form using the powerful CQP query language (IMS Open Corpus Workbench) and to display KWIC concordances of the search matches with customizable context size and sort options. Concordances are displayed in a new tab next to the search form. The corresponding page (or column) of the edition is displayed upon a click on a line corresponding to an occurrence in the concordance. The words matching the query are then highlighted (see the figure below).
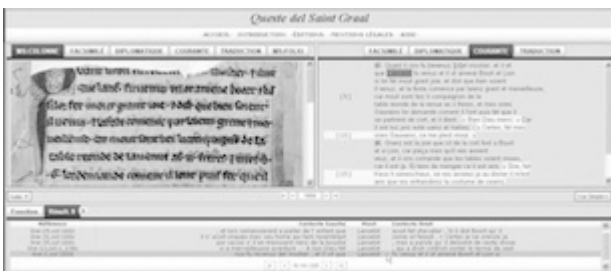

Figure 1

In the future, all the main textometric tools will be interfaced in the edition. These include specificity, collocates, lexicograms, etc. The source XML-TEI files of the Old French text and its modern French translation will be downloadable under a Creative Commons license (Attribution-Noncommercial-Share Alike) as soon as the edition is stable. The editorial and online

textometric research platform is distributed with an open source license (GPL v3) and can be used for publishing on the web various texts with XML-TEI encoding. The platform can handle any number of texts, up to several hundred million words in total.

The poster will display the editorial principles and several screenshots of the edition. Interactive work with the edition can be performed at a demo session.

References

Haugen, Odd E. (2004). 'Parallel Views: Multi-level Encoding of Medieval Nordic Primary Sources'. *Literary and Linguistic Computing.* **19.1**: 73-91.

Haugen, Odd E., ed. (2009). *MUFI character recommendation. Characters in the official Unicode Standard and in the Private Use Area for Medieval texts written in the Latin alphabet. Part 1: Alphabetical order. Version 3.0.* Bergen: Medieval Unicode Font Initiative.

Pauphilet, Albert. (1923). *La Queste del Saint Graal. Roman du XIII⁰ siècle.* Paris: Champion.

Pignatelli, Cinzia and Molly Robinson, eds. (2002). *Chrétien de Troyes : Le chevalier de la Charette (Lancelot) : le "Projet Charette" et le renouvellement de la critique philologique des textes, Œuvres et critique, 27.1.* Tübingen: Gunter Narr Verlag.

Prévost, Sophie, Céline Guillot, Alexei Lavrentiev and Serge Heiden (To be published). *Jeu d'étiquettes CATTEX2009.* Lyon: Équipe de la BFM. `http://ccfm.ens-lsh.fr`.

Vielliard, Françoise and Olivier Guyotjeannin (2001). *Conseils pour l'édition des textes médiévaux, Fascicule I, Conseils généraux.* Paris: CTHS, Ecole nationale des chartes.

Andron Scriptor font page. `http://www.mufi.info/fonts/#Andron`.

Charrette Project. `http://lancelot.baylor.edu`.

Creative Commons license. `http://creativecom` `mons.org/licenses/by-nc-sa/3.0.`

IMS Open Corpus Workbench. `http://cwb.sou` `rceforge.net.`

Medieval Unicode Font Initiative (MUFI). `http` `://www.mufi.info.`

Queste del Saint Graal edition prototype (temporary). `http://textometrie.risc.cnrs.f` `r/txm.`

Textometry Project. `http://textometrie.ens-l` `sh.fr.`

TXM source code . `https://sourceforge.net/p` `rojects/textometrie.`