Digitizing the Second Edition of Furetière's *Dictionnaire Universel*: Challenges of Representing Complex Historical Dictionary Data Using the TEI

Geoffrey Williams, Ioana Galleron

LiCoRN-LIDILEM, Université Grenoble Alpes
e-mail: Geoffrey.williams@orange.fr, ioana.galleron@gmail.com

Abstract

This poster will present preliminary conclusions about the pitfalls and the best practices in marking up entries from the *Dictionnaire universel* d'Antoine Furetière, in the revised and expanded version published in 1701 by Henri Basnage de Beauval (Furetière 1701). As the universality of the *Dictionnaire* is its main characteristic, the project decided to concentrate on terms for a start. Four domains have been spotted (architecture, maritime affairs, law and theatre), which are of interest for all 17th and 18th century French studies, and give a good idea about Furetière's lexicographical principles and techniques.

After an initial analysis conducted on the PDFs by using a CAQDAS system (Atlas.ti), encoding experimentations have been realised, aiming at:

1 Modelling the structure, while allowing to observe the experimentations conducted by Furetière in this regard.

2 Proposing a consensual conceptualisation of scientific domains covered by the Dictionnaire.

3 Linking the quotations to their sources.

Keywords: Furetière; Basnage de Beauval; CAQDAS; TEI; maritime terms; architecture terms; law terms; theatrical terms; Digital Humanities

1   Introduction

The Furetière project seeks to create an open access digital version of Antoine Furetière's *Dictionnaire Universel*. This dictionary, first published in three volumes in 1690 (Furetière 1690), remains an invaluable source for understanding late 17th and early 18th century language usage, most particularly in enlarged and revised edition published in 1701 by Henri Basnage de Beauval (Furetière 1701). Both editions of the dictionary are freely downloadable from the Gallica web archive of the French national library, the BNF, as PDF. In order to exploit this large amount of data, the project uses a dual technology approach by first analysing the content with Computer Assisted Qualitative Data Analysis Software (CAQDAS)[1] and then creating an XML version following the TEI guidelines.

A digital version of Furetière's 1690 dictionary has been made and can be found in the online collection of Classiques Garnier. The problem with this edition is that it has very simple mark-up and is hidden behind a commercial interface that limits access and renders in-depth analysis impossible. The long-term aim of our project is to mark-up up the much bigger revised version of Basnage de Beauval, as this is the one that became the reference edition. An earlier attempt to digitise the

---

[1] www.atlasti.com

Basnage de Beauval text was carried out in the late nineties (Wionet & Tutin 2001), but only tackled the letter D, in a customised SGML TEI. As this data is not available, we are starting from scratch using TEI P5 in XML.

Given the sheer volume of data, and the current total lack of financing, the project starts by concentrating on words designated in the text as terms (Williams Forthcoming). Indeed, what distinguished the work of Furetière, and subsequently of Basnage de Beauval, was that this *Dictionnaire Universel* was to cover the language of the sciences and crafts, to the difference of the *Dictionnaire de l'Académie* which ruled such words out, as being incompatible with the "bel usage". This is what made the *Furetière* a success at the time, and also which renders it a precious source of information about the Ancien Régime nowadays.

2   Digital Humanities: CAQDAS

Too often, Digital Humanists seem to expect humanists to adopt computer science research paradigms rather than developing tools that use the wealth and experience in the humanities and valorise the qualitative research that epitomises the field. In Sociology, qualitative analyses form the basis of much research and appropriate tools have been developed – Computer Assisted Qualitative Data Analysis Software - CAQDAS. We are using one such tool, Atlas Ti, to literary and lexicographical research.

In a first stage we are concentrating on four areas – the language of architecture, of maritime activities, of law and of the theatre. These domains have been chosen because they are of the highest interest for all 17th and 18th French studies, and the Furetière project is not about creating a lexical resource, but a digital resource for all humanities scholars. This means that rather than shoe-horning the dictionary into NLP accepted formats, we are trying to use CAQDAS and the guidelines of the Text Encoding Initiative so as show the sheer wealth of this highly innovative dictionary and to allow exploration of its complexity.

Atlas Ti allows us to work directly on the PDF. Working systematically through the 3200 pages of dense data, we code the terms marked as such, and code the various formulae used to do so:

Formule : Ce terme est particulier au [utilisateur]
Formule : Terme de [domaine ou métier]
Formule : Terme qui affirme
Formule: Ancien terme de [domain ou métier]
Formule: c'est un terme de [...]
Formule: ce terme signifie la même chose que
Formule: en termes de [domaine ou métier]
Formule: est aussi un terme de [domaine ou métier]
Formule: est encore un terme de [domaine ou métier]
Formule: On dit en termes de [Domaine]
Formule: se dit en termes de [Domaine]
Formule: Terme [modifiant]
Formule: vieux terme de [...]

Using Atlas Ti, we thus build a list of the terms that can then be marked up in TEI. However, Atlas Ti allows us to carry out a full analysis using codes to annotate all aspects of usage as well as creating interactive networks to show interaction between terms.
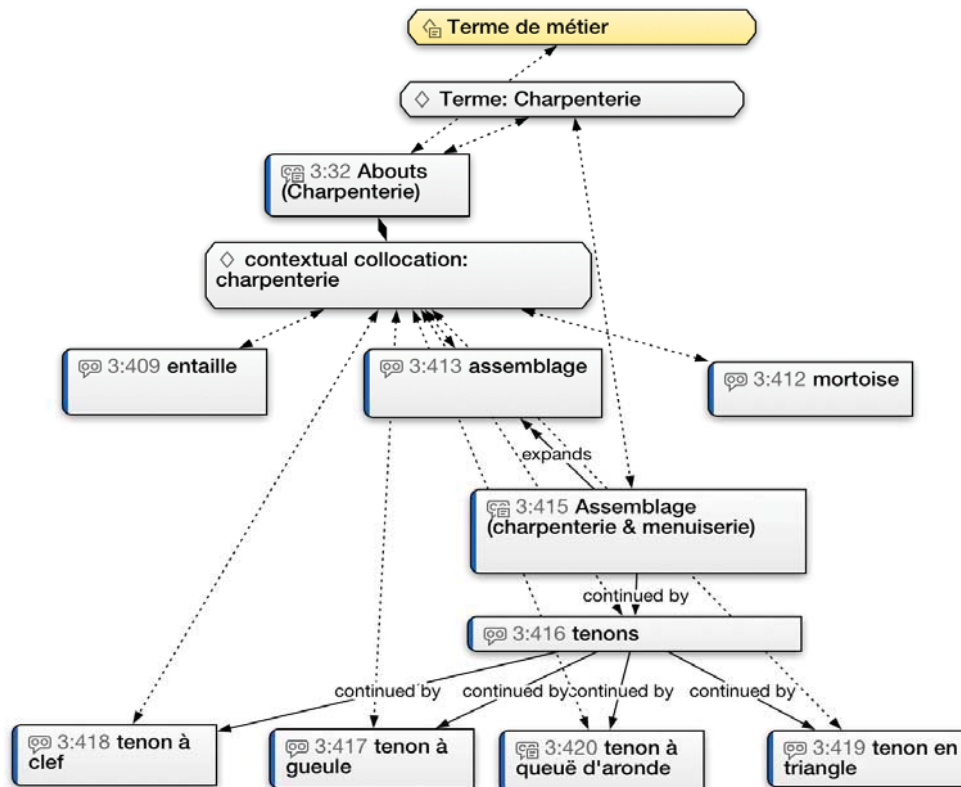
Figure 1: A network of related terms in the field of carpentry.

Atlas Ti can be used simply as a digital highlighter, but can do much more than that in allowing the organisation and analysis of complex codes with annotations and comments. It also allows the export of analytical data in XML format. However, although it allows the use of humanistic methodologies on image data, the data remains an image and therefore of a limited use. Thus, the ultimate need is to create a TEI-conformant version of the dictionary, using the input coded through Atlas.

3   Digital Humanities: Text Encoding Initiative

Europe has a tremendous wealth of historical data, and this should be readily accessible to researchers. This is what tries to achieve the workgroup on retrodigitalisation of the COST action IS1305, where issues involved in marking up complex historical data are discussed.

In creating a digital version of the dictionary, our first audience is that of the dictionary user who is familiar with paper editions and simply wants to read a transliterated and easier to handle version online; 18th century typography and spelling are therefore replaced by modern norms, while preserving the initial orthography through the use of <orig> and <reg> tags. The modern reader is therefore no longer held back by the need to disambiguate between ſ and f, while the scholar interested in the evolution of writing norms and of prononciation can still retrieve this information.

However, the mark-up must allow us to transcend the actual document and make the information available using computational search systems for different users as linguists, historians and literary scholars. But document annotation, allowing to handle data via XQueries, poses various challenges. The first problem is to build a model for the data in a way that lets the experimental structuring of the 17th century lexicographer remain visible. Take for instance the entry "abatre", which can be broken

down into 7 senses, two of which are deemed terminological in content. Some senses, as the first, can be broken into a series of sub-senses, often introduced by a series of synonyms for the headword. We can observe that even the first general sense can be broken into two groups based on two series of synonyms each with extensive examples and citations. A closer look shows that although Furetière presents a class of synonyms, each example exemplifies only one of the listed words, and not necessarily in the order in which these appear in the list:

\<def>Renverſer, demolir, faire tomber, coucher par terre.\</def>
\<cit type="example">\<quote>\<hi rend="italic">Abbatre\</hi> une maiſon pour la rebâtir. \</quote>\</cit>
\<cit type="example">\<quote> Ce Luiteur a \<hi rend="italic">abatu\</hi> ſon homme ſous lui. \</quote>\</cit>
\<cit type="example">\<quote> Ce Chaſſeur \<hi rend="italic">abat\</hi> bien du gibier. \</quote>\</cit>
\<cit type="example">\<quote> On \<hi rend="italic">abat\</hi> les noix avec la gaule. \</quote>\</cit>
In addition to this language information, each entry is linking to a broad usage field and introduces both cultural and language information as in a modern learner's dictionary. The aim is clearly pedagogical, showing how the word is used in context, rather than purely having the prestige of literary usage, which became the norm in the following centuries. In addition to examples, we also find quotations from well-respected authors from the 17th century, near to collocations observed in the language of craftsmen of all categories:

\<cit type="citation">\<quote>\<bibl ana="#Abl">ABL.\</bibl> Les Moiſſonneurs ont \<hi rend="italic">abatu\</hi> trois arpens de blé aujourd'huy.\</quote>\</cit>
\<cit type="example">\<quote> Les Bouchers diſent \<seg type="collocation">\<hi rend="italic">abatre\</hi> le cuir d'un boeuf\</seg>, pour l'écorcher.\</quote>\</cit>
Such structuring had to wait until the innovative COBUILD dictionaries to appear again on the market. The question is here how to represent this wealth and internal contrasts with the available tags.

A small extract from "abatre" as a maritime term shows the wealth of information that must be managed. The definition describes the naval manoeuvre whilst the examples list collocations each of which details actions, who carries out the action and in what circumstances.

\<sense n="2">\<usg type="dom" ana="#marine">\<oRef corresp="#abatre">\</oRef>ABATRE en termes de Marine,\</usg> \<def>ſignifie, Decheoir, deriver, s'écarter de la vraye route: ce qui ſe fait par la force des courants, ou des marées, ou par les erreurs du pointage, ou du mauvais gouvernement du timonier\</def>.
\<cit type="example">\<quote>On dit auſſi, qu'\<seg type="collocation">un Pilote \<hi rend="italic">abat\</hi> ſon vaiſſeau d'un quart de Rumb ou d'un autre aire de vent\</seg>,\</quote>\</cit>
\<def>quand il vire ou change ſa courſe, & gouverne ſur une autre Rumb de celui de ſa route.\</def>

Thus, we have not only language information to be managed, but also detailed descriptions of naval practice that are of interest to historians. This is naval practice seen by an intellectual, working with specialists in maritime history we are linking this data to sources written by naval officers.

One of the most delicate issues is that of using attributes, as for example "ana" for linking terms to a specific scientific or arts and crafts domain. Conceptualisation of domains and of data is far from

being consensual amongst scholars from different disciplines involved in the project. Furetière and Basnage do not proceed with a "knowledge tree" in head, as the one appearing later in the preamble of Diderot's and d'Alembert Encyclopaedia, therefore pockets of specialised language appear quite unexpectedly. Sometimes, they link to well-known disciplines, such as the anatomy or the geometry, to give but two examples; it will be interesting to see, when the project is completed, what falls under certain domains as opposed to others who are intellectually close to them: what difference between philosophical and logical terms, for instance, and to what extent this difference is supported by treatises of that time? In some other cases, specialised terms do not relate to a formalised "science"; Furetière and Basnage resort then, most frequently, to the names of the agents ("terme de marchand de drap", "terme de cordier", "terme de jardinier") to designate the specialised use. For the moment, we encode data following Furetière's and Basnage's denominations, but at a later stage we intend to build a more complex taxonomy to provide links between related items.

One of the aims of the project is to link quotations with known source text. On the one hand, this will allow to see who Basnage quotes in relation with what scientific domain or craft, a question of particular interest as he was engaged, in parallel, in a journal, called somewhat misleadingly *Histoire des ouvrages des savants*. On the other hand, in so far as Basnage had access to much of what was being written in literary and scientific circles of the time, the *Histoire* and the source texts might well be seen as a corpus. It will be interesting to confront the use of the term in the dictionary with its occurrences in the initial text. True enough, to conduct such studies asks for electronic editions not only of the *Dictionnaire universel*, but also of the full library on which it draws – a titanic encoding endeavour, and still a most valuable one if we want to understand how modernity emerged gradually in French language and thinking. For the moment, the problem we have is in simply identifying the quotations, as titles are seldom indicated.

4   Conclusion

The project is an ambitious one and presents numerous technical difficulties that have still to be overcome. We are also aware about the fact that solutions and practices forged by dealing with terms may have to evolve when tackling other words. Our aims are threefold: developing the use of CAQDAS and other tools that can be readily adopted in the humanities; contributing to the Text Encoding Initiative; and above all providing an open access resource destined to add to the growing corpus of data freely available to scholars.

5   References

Furetière, A. (1690). *Dictionnaire Universel, contenant généralement tous les mots françois tant vieux que modernes et les termes des sciences et des arts*. La Haye et Rotterdam. http://gallica.bnf.fr/ark:/12148/bpt6k50614b. [26/02/2014]

Furetière, A. (1701). *Dictionnaire universel, contenant généralement tous les mots françois tant vieux que modernes, & les termes des sciences et des arts*. Tome 1 / ,... par feu messire Antoine Furetière, 2e édition revue, corrigée et augmentée par M. Basnage de Bauval. La Haye et Rotterdam: Arnoud et Reinier Leers. http://gallica.bnf.fr/ark:/12148/bpt6k5841680f. [26/02/2014]

Williams, G. Forthcoming. «Le temps des termes: les termes et la phraséologie dans les dictionnaires du 17 siècle». In *Fraseologia e paremilogia: Passato, presente e futuro*, édité par Cosimo De

Giovanni. Milan: FrancoAngeli.

Wionet, C. & Tutin, A. (2001). *Pour informatiser le Dictionnaire universel de Basnage (1702) et de Trévoux (1704) : Approche théorique et pratique*. Paris: Honoré Champion.

Acknowledgements