

LA RÉSURRECTION DU DICTIONNAIRE ANCIEN PAR LA DÉCONSTRUCTION POSITIVE DE L'INFORMATIQUE

Christophe REY & Corinne ZAOUÏ
DELIC, Université de Provence

SOMMAIRE

Introduction

1. Le point sur l'informatisation des dictionnaires anciens

1.1. Les méthodologies de rétroconversion

1.2. Quelques spécificités à prendre en compte

2. Déconstruire positivement le texte ancien

2.1. Pourquoi parler de déconstruction ?

2.2. L'apport "positif" de l'Informatique

Conclusion

Résumé : Les travaux d'étude et d'informatisation des dictionnaires anciens que poursuit l'un d'entre nous (C. Rey) mettent clairement en évidence les problèmes d'informatisation liés à la nature même de ces données, à savoir un certain manque de rigidité structurelle. Ce manque de rigidité occasionne indubitablement une moins bonne lisibilité des données, que les dictionnaires modernes – par le biais de leur informatisation systématique – semblent avoir réglée. Il est ainsi parfois difficile d'identifier les différents champs d'information constitutifs des tout premiers dictionnaires, et de repérer précisément, par exemple, ce qui relève du marquage grammatical, de la définition, etc., afin de pouvoir reconstruire, récupérer, retrouver ou mettre en évidence les informations les plus saillantes qui s'y trouvent.

Les soucis de structuration des données pour une manipulation et une interrogation grâce aux normes de codage informatique nouvelles prônées par C. Zaoui, permettent d'apporter des solutions nouvelles à notre projet de rétroconversion d'un dictionnaire encyclopédique, le dictionnaire Grammaire & Littérature (1782-1786) de l'Encyclopédie Méthodique (1782-1832).

L'un de nos objectifs communs étant de proposer une solution qui, à travers ses choix méthodologiques et informatiques, soit généralisable, ou tout au moins facilement réexploitable sur d'autres corpus, nous avons donc fait émerger le concept d'un balisage XML "souple" ou "flottant".

La solution des balisages précédents, soit minimal (Wooldridge 1994 et 1996, Leroy-Turcan 1996) soit analytique (Wionet et Tutin) n'étant pas pleinement satisfaisante, cette notion de balisage "souple" nous permettait d'allier la souplesse du premier et la richesse du second, tout en respectant la nature non rigide du document et en autorisant ainsi un balisage identifiant les grands champs informationnels et en mettant en évidence un certain nombre d'informations flottant à l'intérieur de ces grands constituants. L'efficacité de notre type de balisage a pu être mise à l'épreuve à travers la réalisation d'un outil d'interrogation de fichiers XML : CorpXML (<http://www.up.univmrs.fr/delic/perso/rey/methodique/index.htm>).

Introduction

Ainsi que l'illustre le foisonnement des ouvrages "grand public" paraissant à la fois sous leur forme papier et sous une forme informatique, le dictionnaire électronique semble s'être installé assez largement dans le panorama lexicographique actuel. Pour s'imposer, ce dernier a bénéficié non seulement de la place grandissante de l'outil informatique dans les techniques du monde de l'édition, mais aussi de l'aspect très structuré de nos dictionnaires modernes. L'évocation de cette dynamique très active d'informatisation des dictionnaires ne peut légitimement être faite sans que soit mise en avant la question très particulière de l'informatisation des dictionnaires anciens.

Initiés depuis les années 80 par le lexicographe Russon Wooldridge, les travaux de mise à disposition électronique des monuments lexicographiques de notre passé se distinguent comme une entreprise répondant à des contraintes bien particulières. L'article que nous proposons s'inscrit dans la lignée des études déjà existantes sur ce sujet et décrit la solution du balisage *souple* ou *flottant* pour l'informatisation des dictionnaires anciens de nature encyclopédique.

Après avoir dressé un bref rappel des solutions de rétroconversion déjà existantes, nous présentons les spécificités du balisage que nous avons mis au point et revenons sur l'application

que nous en avons faite sur l'un des trente-neuf dictionnaires de matière de l'*Encyclopédie Méthodique* (1782-1832).

1. Le point sur l'informatisation des dictionnaires anciens

Répondant à des caractéristiques macrostructurelles et microstructurelles qui ne caractérisent pas les dictionnaires modernes, les dictionnaires anciens constituent, en vue de leur informatisation, un objet d'étude particulièrement délicat. Parmi les nombreux projets de recherche consacrés à cette question de la rétroconversion des dictionnaires des siècles précédents, deux grandes solutions d'informatisation semblent s'être imposées : celle du balisage *minimal* et celle du balisage *analytique*.

1.1. Les méthodologies de rétroconversion

Les deux grandes solutions que constituent le balisage *minimal* et le balisage *analytique* possèdent chacune leurs spécificités.

Utilisée pour la rétroconversion du *Dictionnaire de l'Académie Française* (1694) (Wooldridge 1994), ou du *Dictionnaire Critique* (1787) de Jean-François Féraud (Caron, Dagenais, Gonfroy 1992), la solution du balisage *minimal* est une approche minimaliste guidée par le souci de préserver l'intégralité du texte informatisé. Elle se traduit par un balisage restreint de la structure dictionnaire grâce à l'identification de "points d'accès" au document caractérisant à la fois des indices typographiques ou de mise en page du texte balisé, pouvant ainsi en mentionner l'édition, la page, la colonne, ou préciser l'existence de fontes particulières (grandes et petites capitales, italique, gras, etc.). La figure 1 ci-dessous¹ fournit un exemple de balisage *minimal* :

TIMBRE. s. m. Sorte de cloche ronde qui n'a point de battant en dedans, & qui est frappée en dehors par un marteau. *Le timbre d'une horloge. timbre d'un reveille-matin. le timbre de cette horloge est tres-bon.*[...] Timbrer. v. a. Terme de blason, Accompagner d'un timbre. *Timbrer une armoirie.*
Timbrer. v. a. Terme de Pratique, Ecrire au haut d'un Acte, la nature de cet acte, sa date & le sommaire de ce qu'il contient. *Timbrer des pieces.*
On dit aussi, *Timbrer du papier, timbrer du parchemin*, pour dire, Imprimer la marque du Roy sur du papier, sur du parchemin, pour faire qu'il puisse servir aux actes de Justice.

```
<page n="563"><col n="1">[...]<p><lc>TIMBRE</lc>. s. m. Sorte de cloche ronde qui n'a point de battant en dedans, & qui est frappée en dehors par un marteau. <i>Le timbre d'une horloge. timbre d'un reveille-matin. le timbre de cette horloge est tres-bon</i>.<p> [...] <sc>Timbrer</sc>. v.a. Terme de blason, Accompagner d'un timbre. <i> Timbrer une armoirie</i>. <p><sc>Timbrer</sc>. v.a. Terme de Pratique, Ecrire en haut d'un Acte, la nature de cet acte, sa date & le sommaire de ce qu'il contient. <i> Timbrer des pieces</i>. <p>On dit aussi, <i> Timbrer du papier, timbrer du parchemin</i>, pour dire, Imprimer la marque du Roy sur du papier, sur du parchemin, pour faire qu'il puisse servir aux actes de Justice.<p>2
```

Figure 1. Exemple de balisage minimal (Dictionnaire de l'Académie 1694)

En plus de cette première forme d'indications, l'approche minimaliste permet également l'accès à certains champs informationnels de l'article balisé et ce grâce à l'utilisation de mots-clés métalinguistiques (Wooldridge, Leroy-Turcan 1996).

Cette solution du balisage *minimal* contraste assez radicalement avec la solution du balisage *analytique* retenue pour l'informatisation du *Dictionnaire Universel de Furetière revu par Basnage de Bauval* (1702) (Wionet, Tutin 1998 et 2001), dans la mesure où, en plus d'un balisage de nature

¹ Exemple tiré de l'article de T.R. Wooldridge, "L'informatisation du Dictionnaire de l'Académie française (DAF)", dans *Actes du colloque-atelier international DictA1998* organisé par le Groupe d'Études sur l'Histoire de la Langue Française (GEHLF) et la Société Internationale d'Études Historiques et Linguistiques des Dictionnaires Anciens (SIEHLDA), Université de Limoges, 19-20 novembre 1998.

² Les codages <page n="563">, <col n="1">, <p>, <lc></lc>, <i></i>, <sc></sc>, servent respectivement à marquer la page dans laquelle se situe l'article balisé, les colonnes occupées par l'article dans la page, la présence d'un paragraphe, les grandes capitales (Low capitales), le caractère italique de l'information codée, et les petites capitales (Small capitales).

typographique fin, Chantal Wionet et Agnès Tutin revendiquent un repérage exhaustif et systématique¹ des différents champs informationnels de l'article de dictionnaire, en vue de permettre des requêtes très fines de sa structure. La figure 2 proposée ci-dessous illustre le balisage *analytique* de l'article *Daguet* du *Dictionnaire Universel de Furetière revu par Basnage de Bauval* :

DAGUET. Terme de Venerie. Jeune cerf, qui est à sa première tête; qui pousse son premier bois.
Daguet. adv. Sourdement; en cachette. Il s'en est allé, il a tiré ses chausses daguet. Cela est bas et populaire.²

```
<Entry>
  <Form Type=LEMMA><Orth Rend=CAPS>DAGUET</Orth>. </Form>
  <GramGrp><Pos Type=S></Pos><Gen Type=M></Gen></Gramgrp>
  <Sense><CDomain><Lbl>Terme de</Lbl><Domain> Venerie</Domain>.</CDomain>
  <Def>Jeune cerf, qui est à sa première tête; qui pousse son premier bois.</Def></Sense>
  <Re><Form Type=HOMOGRAPHE><Orthre Rend=SCAPS>Daguet</Orthre>.</Form>
  <GramGrp><Pos Type=ADV>adv. </Pos></GramGrp>
  <Sense><Def>Sourdement; en cachette. </Def>
  <Eg><Q>Il s'en est allé, il a tiré ses chausses <Oref Rend=IT>daguet</Oref>. </Q></Eg>
  <CUsg><Lbl>Cela est </Lbl><Usg>bas et populaire.</Usg></Cusg></Sense></Re>
</Entry>
```

Figure 2. Exemple de balisage analytique

Ces deux stratégies de rétroconversion offrent bien évidemment un rendement différent en termes de capacité d'interrogation des données rétroconverties. À ce titre, mais également parce qu'elle s'appuie sur l'utilisation d'une norme de codage³ – le langage SGML (Standard Generalized Markup Language) – la solution du balisage *analytique* apparaît aujourd'hui comme la solution la plus adaptée pour la redécouverte des monuments de notre lexicographie grâce à l'outil informatique.

1.2. Quelques spécificités à prendre en compte

Dans la perspective de fournir une édition électronique du corpus de thèse⁴ de C. Rey (Rey 2004), nous avons essayé de choisir la solution de balisage la mieux adaptée à la nature encyclopédique de ce dernier. Notre volonté étant de fournir une édition électronique autorisant une interrogation la plus riche possible de la structure informationnelle, la solution du balisage *minimal* a d'emblée été écartée. Nous avons donc dans un premier temps opté pour la mise en application de la solution plus séduisante du balisage *analytique*.

Néanmoins, la nature encyclopédique du dictionnaire à rétroconvertir nous a rapidement confrontés à deux obstacles que ce type de balisage ne règle pas.

Nous avons précisé plus haut que la richesse du balisage *analytique* découlait du repérage de la totalité des champs informationnels de l'article, or, la structure "molle" de notre dictionnaire – loin d'être aussi rigide que celle des dictionnaires modernes – met en évidence l'existence de certains chevauchements d'informations. Certains articles sont en effet construits de telle manière qu'il nous est difficile de savoir où commence et où finit un champ particulier. C'est notamment ce que nous avons essayé d'illustrer à travers les découpages multiples envisageables pour l'énoncé suivant tiré de l'article *Sous-entendu* du *Dictionnaire français* (1680) de César-Pierre Richelet :

¹ Lorsqu'un champ informationnel "stable" n'existe pas, il est tout de même procédé à un balisage spécifiant à l'utilisateur que ce champ est effectivement absent.

² Sans trop détailler l'exemple fourni, nous pouvons mentionner le repérage des champs informationnels principaux tels que <Form> indiquant si l'entrée ou la sous-entrée est un lemme ou un homographe, la marque de domaine <Domain>, l'information grammaticale <GramGrp>, la définition <Def>, et des champs secondaires comme les remarques sur la typographie des entrées, des sous-entrées, ou des références, <Orth Rend=CAPS>, <Orthre Rend=SCAPS>, <Oref Rend=IT>, les marques d'usage <Usg>, les particules introduisant un champ, nommées libellés <Lbl>, etc.

³ La solution du balisage *minimal* s'appuie elle sur des balisages propriétaires, à savoir des jeux de balises ne faisant pas l'objet de recommandations par un organisme de standardisation.

⁴ Ce corpus va être présenté dans la seconde partie de cet article.

- 1- (C'est une figure de Grammaire) [qui consiste à n'exprimer point, par élégance, un ou plusieurs mots.]
- 2- [C'est une figure (de Grammaire) qui consiste à n'exprimer point, par élégance, un ou plusieurs mots.]
- 3- [C'est une figure] (de Grammaire) [qui consiste à n'exprimer point, par élégance, un ou plusieurs mots.]

Dans cet exemple précis, l'identification de la marque de domaine grammatical – délimitée par les parenthèses – et de la définition – mise entre crochets – ne peut se faire sans que l'arbitraire du lexicographe procédant au balisage ne se manifeste. Une telle difficulté, déjà présente pour des dictionnaires de langue et exacerbée pour les dictionnaires encyclopédiques, réside dans notre volonté de vouloir imbriquer chaque portion textuelle au sein de "boîtes" pré-étiquetées et distinctes les unes des autres.

En lien étroit avec ce premier obstacle, il semblerait par ailleurs que la nature encyclopédique de notre dictionnaire génère également une multiplication des informations susceptibles d'être mises en valeur dans le but d'une investigation lexicographique informatique plus fine.

À titre d'exemples, il peut en effet paraître intéressant de repérer les informations issues du développement encyclopédique que ne fournit pas le dictionnaire de langue. À un autre niveau d'analyse, il peut également s'avérer important de baliser les titres d'ouvrages ou les noms propres mentionnés dans le corps de l'article. Comment peut alors se faire le repérage de ces données sans que de nouveaux chevauchements informationnels ne soient générés ?

Les réflexions que nous avons été amenés à conduire pour mettre au point la rétroconversion de notre corpus se sont concrétisées par l'émergence d'une nouvelle solution de balisage que nous allons détailler ci-dessous : le balisage *souple* ou *flottant*.

2. Déconstruire positivement le texte ancien

La solution du balisage *souple* ou *flottant*, en plus de constituer une alternative intéressante aux solutions déjà existantes, participe selon nous à la "déconstruction positive du texte ancien". Pour illustrer les particularités de ce balisage, nous nous proposons d'évoquer l'exemple de la "déconstruction informatique positive" que nous avons fait subir à notre corpus de travail : les articles de Grammaire de l'*Encyclopédie Méthodique*.

Conçue comme une édition corrigée, remaniée et augmentée de la célèbre *Encyclopédie* ou *Dictionnaire raisonné des arts et des sciences* (1751-1777) de Diderot et d'Alembert, l'*Encyclopédie Méthodique* (1782-1832) publiée par Charles-Joseph Panckoucke constitue l'un de nos monuments lexicographiques du siècle des Lumières. À plusieurs égards décisive pour la traduction de la mutation épistémologique s'étant imposée entre le milieu et la fin du XVIII^e siècle, la "Méthodique" mérite ainsi de sortir de l'ombre imposante de la première encyclopédie. C'est à ce titre qu'après avoir fourni une réflexion d'historien de la langue sur la nature des connaissances véhiculées sur les sons du français au sein du dictionnaire *Grammaire & Littérature* (1782-1786) – l'un des trente-neuf dictionnaires de matière de cette encyclopédie – nous nous sommes penchés sur l'élaboration d'une méthode d'informatisation qui puisse restituer toute la richesse et l'originalité de cet ouvrage.

2.1. Pourquoi parler de déconstruction ?

Ainsi que nous l'avons sommairement évoqué plus haut, notre conception de déconstruction du document est issue de l'initiative présente au sein du balisage *analytique* de Wionet et Tutin.

En effet, à la différence du balisage *minimal* qui ne propose qu'un accès réduit au contenu de l'article, le balisage *analytique* propose de poser au sein de la microstructure de ce dernier des jalons issus d'une norme de codage, le langage SGML. Notre solution propose le même "éclatement" informationnel mais privilégie la norme de codage XML (eXtensible Markup Language), standard incontesté de codage et de diffusion des documents électroniques.

L'éclatement occasionné par un balisage *souple* ou *flottant* se distingue néanmoins de celui obtenu grâce à un balisage *analytique* par le fait qu'il s'affranchit comme nous l'avons déjà dit plus haut, de l'idée de faire coïncider chaque portion de texte avec un champ informationnel préétabli. Toute la flexibilité et la souplesse de notre balisage vient d'un dégroupement des données de l'article en deux grands blocs : un bloc <entree> qui renferme le lemme (ou ses formes fléchies et dérivées) ainsi que l'information grammaticale, et un bloc <corps> qui renferme la totalité des autres informations de l'article. C'est précisément au niveau du bloc <corps> que les problèmes

d'identification et de découpage des éléments sont les plus présents et que le balisage *souple* ou *flottant* prend tout son sens.

Ainsi que l'illustre la figure 3 ci-dessous représentant la forme balisée de l'article *Prosonomasie*¹, les données appartenant à ce bloc peuvent flotter à l'intérieur de cet espace en étant ou non identifiées comme appartenant à un champ informationnel clairement défini.

```
<ARTICLE>
  <STATUT TYPE="DIFFERENT"/>
  <ENTREE TYPE="EP">
    <FORME>PROSONOMASIE</FORME>,
    <INFORMATION_GRAMMATICALE TYPE="SUBSTANTIF
      FEMININ">
      <PARTIE_DU_DISCOURS TYPE="SUBSTANTIF">s.
    </PARTIE_DU_DISCOURS>
    <GENRE TYPE="FEMININ">f.</GENRE>
    </INFORMATION_GRAMMATICALE>
  </ENTREE>
  <CORPS>
    <EXTRA TYPE="AJOUT">(<ETYMOLOGIE><LANGUE
      TYPE="GREC">Προσωνομασία</LANGUE>, du verbe
      <LANGUE TYPE="GREC">προσωνομάζω </LANGUE>,
      <LANGUE TYPE="LATIN">insuper nomino
      </LANGUE>. </ETYMOLOGIE> <DEFINITION>C'est un autre nom de la
      figure appelée ordinairement Paronomase).</DEFINITION>
    <SIGNATURE TYPE="BEAUZEE">(M.BEAUZÉE.)</SIGNATURE>
    </EXTRA><DISCOURS_ENCYCLOPEDIQUE>Figure de Rhétorique, par
    laquelle on fait allusion à la ressemblance du son qui se trouve entre différents
    noms ou différents mots, comme dans ces phrases : <LANGUE
    TYPE="LATIN">Is verè consul est qui Reipublicæ saluti consulit ; Quum
    lectum petis de lethocogita </LANGUE>. Voyez <REFERENCE
    TYPE="VEDETTE">PARONOMASE</REFERENCE>.
    </DISCOURS_ENCYCLOPEDIQUE>
    <SIGNATURE TYPE="ANONYME">(ANONYME.)</SIGNATURE>
  </CORPS>
</ARTICLE>
```

Figure 3. Exemple de balisage Souple ou Flottant

À l'intérieur de ce bloc, des données peuvent également être identifiées comme appartenant à un champ particulier, tout en étant déjà incluses dans un autre champ. Ceci donne l'impression que ces éléments constituent des portions de texte flottantes.

Une autre particularité de notre balisage est de dissocier le balisage *logique*, à savoir le repérage des champs informationnels, du balisage *physique* qui considère les critères typographiques du document informatisé. Par le biais d'une feuille de style associée au document .xml contenant le texte balisé, il est alors possible de restituer l'aspect physique du texte rétroconverti. Il est par ailleurs intéressant de mentionner que le balisage physique du document permet, ainsi que cela a pu être mis en application pour le *Tlfi* (Dendien, Pierrel 2003), d'effectuer un balisage logique semi-automatique du document.

2.2. L'apport "positif" de l'Informatique

L'aspect "positif" de la déconstruction que nous proposons grâce au balisage *souple* ou *flottant* se trouve quant à lui illustré par les possibilités d'interrogation du document qui découlent de ce processus de balisage.

Dans une entreprise telle que la rétroconversion d'un dictionnaire, les capacités d'interrogation des données et d'investigation au cœur du texte vont entièrement dépendre de la richesse du balisage appliqué. Pour notre entreprise d'informatisation des articles du dictionnaire *Grammaire & Littérature* de l'*Encyclopédie Méthodique*, la puissance d'interrogation des données repose sur un savant mélange de technologies : 1) le choix d'une Définition de Type de Document (DTD) à la fois rigoureuse, souple et riche, et 2) le choix des outils de développement.

¹ La *Prosonomasie* est une figure de Rhétorique par laquelle on fait allusion à la ressemblance du son qui se trouve entre différents noms ou différents mots.

La DTD TEI existant pour les dictionnaires n'étant pas pleinement satisfaisante pour la rétroconversion des dictionnaires anciens (Cf. Rey 2004), et qui plus est pour des dictionnaires anciens de nature encyclopédique, nous avons élaboré notre propre DTD.

Sans pour autant détailler celle-ci, nous pouvons mentionner le fait que cette dernière concilie à la fois le repérage d'éléments traditionnellement identifiés dans la microstructure classique d'un dictionnaire (information grammaticale, information étymologique, définition, etc.) et des éléments supplémentaires (titres d'ouvrages, noms de personnes, noms de personnages, etc.).

Afin d'exploiter au mieux cette DTD, nous avons eu recours à un ensemble d'outils de développement à la fois novateurs, gratuits et permettant de livrer un exécutable.

Notre choix s'est porté sur une Interface de Programmation (API, Application Programming Interface) et plus précisément sur la représentation hiérarchique en mémoire du document XML sous forme d'arbre fournie par l'API DOM (Document Object Model).

Parmi les langages permettant d'exploiter l'API DOM, nous avons retenu le langage de programmation C++ en association avec les bibliothèques QT.

De tous ces choix méthodologiques a pu émerger un outil d'interrogation de notre corpus de travail : le logiciel CorpXML¹.

Conformément à notre souci de proposer un outil qui ne soit pas spécifique à notre seul corpus, le logiciel CorpXML ne s'appuie pas directement sur notre DTD et reconstitue de manière dynamique la structure de tous les documents qui lui sont proposés en lecture.

Deux grands types de recherche sont intégrés au logiciel CorpXML. Une recherche dite "simple" permet d'effectuer une interrogation sur le corpus étudié à partir de mots-clés.

La figure 4 fournit ainsi un exemple de recherche visant à identifier tous les articles comportant à la fois le mot "air" et le mot "nasal" :

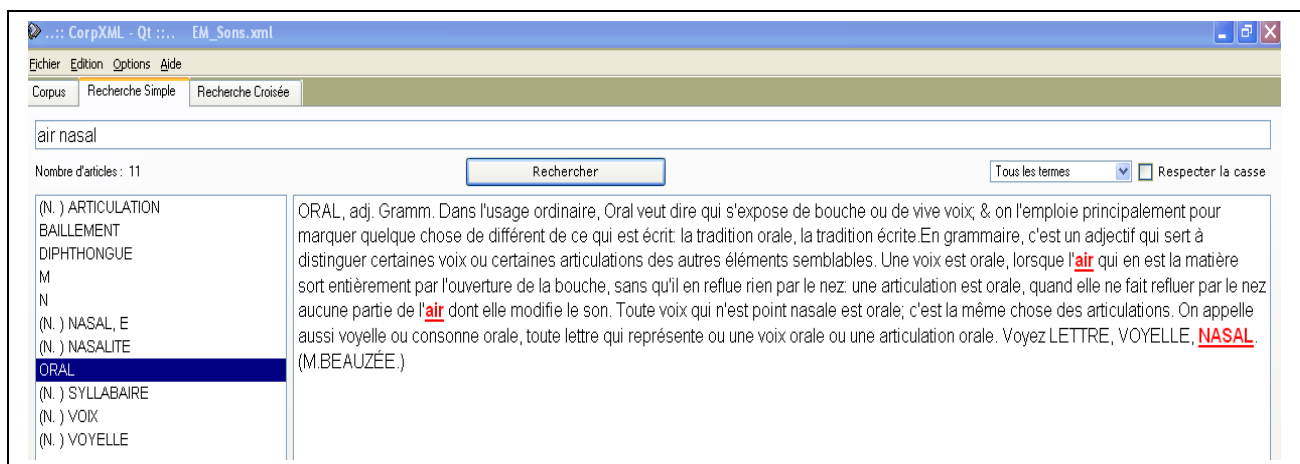


Figure 4. Recherche simple du logiciel CorpXML.

Dans la colonne de gauche s'affiche la liste des articles contenant les mots recherchés, tandis que la partie droite de l'écran permet de lire l'article sélectionné par l'utilisateur au sein de cette liste. Les mots ayant fait l'objet de la recherche sont quant à eux mis en évidence par un procédé visuel combinant un soulignement et une coloration en rouge.

Le second type de recherche, la recherche dite "croisée", repose plus particulièrement sur les balises XML qui jalonnent le texte. Elle permet de croiser de manière assez fine plusieurs critères de recherche.

En ce qui concerne notre corpus d'étude, il est par exemple possible, ainsi que l'illustre la figure 5 ci-dessous, d'obtenir très facilement l'ensemble des articles rétroconvertis constituant des **Substantifs** appartenant au domaine de la **Grammaire** (noté "Gramm") et comportant des portions de texte en **Latin** :

¹ Cet outil gratuit est téléchargeable à l'adresse internet suivante : <http://www.up.univmrs.fr/delic/perso/rey/methodique/index.htm>

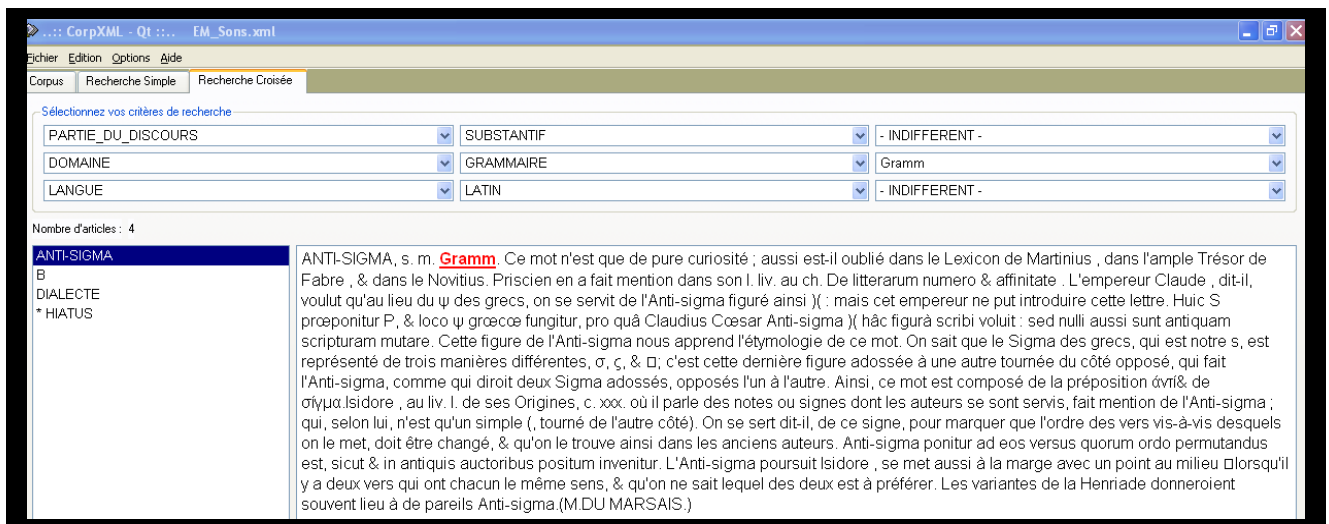


Figure 5. Recherche croisée du logiciel CorpXML.

La colonne de gauche de cette figure fournit la liste des articles correspondant à nos critères après l'application de nos filtres successifs. Dans la partie droite de l'interface, on visualise l'article sélectionné dans la liste obtenue et les portions de texte ayant éventuellement été spécifiées par les critères de recherche sont mises en valeur par le procédé graphique déjà mentionné plus haut.

Conclusion

Ainsi que nous l'avons montré ici, la solution du balisage *souple* ou *flottant* pour la rétroconversion des dictionnaires anciens constitue le point d'intersection de deux visions de l'informatisation des données.

Dans un premier temps, cette solution constitue en effet une réponse à la volonté du lexicographe de ne pas "corrompre" la structure "molle" du dictionnaire encyclopédique ancien en cherchant à tout prix à retrouver au sein de ce dernier la rigueur structurelle de nos dictionnaires modernes.

Dans un deuxième temps, cette même solution répond au souci de l'informaticien d'utiliser une technologie permettant à la fois de respecter la souplesse du texte informatisé et d'en autoriser une interrogation à la fois riche et en conformité avec le courant actuel d'informatisation des documents.

Testé avec succès sur un corpus issu d'une encyclopédie du siècle des Lumières, le balisage *souple* ou *flottant* semble constituer une alternative intéressante aux solutions d'informatisation des dictionnaires anciens déjà existantes, et mérite certainement d'être mis en application sur d'autres ouvrages.

BIBLIOGRAPHIE

BEAUZÉE, N. & MARMONTEL, J-F. (1782-1784-1786). *Encyclopédie Méthodique. Grammaire & Littérature*, Paris (chez Panckoucke), Liège (chez Plomteux). 3 vol.

CARON, P., DAGENAIS, L., GONFROY, G. 1992. Le programme d'informatisation du Dictionnaire critique de la langue française de l'abbé Jean-François Féraud (1787), in T.R. Wooldridge (éd.), *Historical Dictionary Databases, CCH Working Papers*, 2, pp. 87-103.

DENDIEN, J. & PIERREL, J-M. 2003. Le trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence, *TAL*, Volume 44 - n°2/2003, 28 p.

Le Dictionnaire de l'Académie française : histoire et nuances de la langue française (1694- 1935). 2000. éditions Redon.

PANCKOUCKE, Ch-J. (1782-1832). *Encyclopédie méthodique ou par ordre de matières par une société de gens de lettres, de savants et d'artistes ; précédée d'un Vocabulaire universel, servant de Table pour tout l'Ouvrage, ornée des Portraits de MM. Diderot et d'Alembert, premiers Éditeurs de l'Encyclopédie*, Paris, Panckoucke.

REY, C. 2004. *Analyse et informatisation des articles traitant de l'étude des sons dans le dictionnaire Grammaire & Littérature de Nicolas Beauzée et Jean-François Marmontel, issu de l'Encyclopédie Méthodique*, Thèse de doctorat, Aix-en-Provence.

REY, C. & ZAOUI, C. 2004. Le balisage XML "ciblé" : une nouvelle approche dans l'informatisation des corpus, in *Actes de la conférence internationale sur la Fouille de Texte (CIFT'04)*, dans le cadre de la semaine du Document Numérique, La Rochelle, 22-24 juin 2004, pp. 121-133.

RICHELET, C-P. 1680. *Dictionnaire françois, Dictionnaires des XVIe et XVIIe siècles*, 1998, cd-rom pc, version 1.0, Champion électronique.

VÉRONIS, J. & IDE, N. 1996. Encodage des dictionnaires électroniques : problèmes et propositions de la TEI, in D. Piotrowsky (éd.), *Lexicographie et informatique - Autour de l'informatisation du Trésor de la Langue Française*, Actes du Colloque International de Nancy (29, 30, 31 mai 1995), Paris, Didier Erudition, pp. 239-261.

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

WIONET, C. et TUTIN, A. 1998. Informatisation du Dictionnaire Universel de Furetière revu par Basnage de Bauval (1702) : premier bilan, in *Actes du colloque-atelier international DictA1998* organisé par le Groupe d'Études sur l'Histoire de la Langue Française (GEHLF) et la Société Internationale d'Études Historiques et Linguistiques des Dictionnaires Anciens (SIEHLDA), Université de Limoges, 19-20 novembre 1998.

WOOLDRIDGE, T.R. 1998. L'informatisation du Dictionnaire de l'Académie française (DAF), in *Actes du colloque-atelier international DictA1998* organisé par le Groupe d'Études sur l'Histoire de la Langue Française (GEHLF) et la Société Internationale d'Études Historiques et Linguistiques des Dictionnaires Anciens (SIEHLDA), Université de Limoges, 19-20 novembre 1998.

WOOLDRIDGE, T.R. & LEROY-TURCAN, I. 1996. Les mots-clefs métalinguistiques comme outil d'interrogation structurante des dictionnaires anciens, in A. Clas, P. Thoiron & H. Béjoint (éds.), *Lexicomatique et dictionnaires*, Beyrouth, FMA & Montréal, AUPELFUREF, pp. 307-16.

WOOLDRIDGE, T.R. 1994. Projet d'informatisation du Dictionnaire de l'Académie (1694-1935), in B. Quemada & J. Pruvost (éds.), *Actes du Colloque international Le Dictionnaire de l'Académie française et la lexicographie institutionnelle européenne*, Institut de France, novembre 1994, Paris, Champion, pp. 309-20.

WORLD WIDE WEB CONSORTIUM. Extensible Markup Language (XML) : <http://www.w3.org/XML/>.