



## **Analyse de données textuelles informatisée : l'articulation de trois méthodologies, avantages et limites**

### **Computer-Assisted Textual Data Analysis: The Combination of Three Methodologies, Benefits and Limitations**

Martine Paindorge, Jacques Kerneis et Valérie Fontanieu

Volume 11, numéro 1, novembre 2015

Sur le thème de l'analyse de données textuelles informatisée

URI : <https://id.erudit.org/iderudit/1035933ar>

DOI : <https://doi.org/10.7202/1035933ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Prise de parole

ISSN

1712-8307 (imprimé)

1918-7475 (numérique)

[Découvrir la revue](#)

Citer cet article

Paindorge, M., Kerneis, J. & Fontanieu, V. (2015). Analyse de données textuelles informatisée : l'articulation de trois méthodologies, avantages et limites.

*Nouvelles perspectives en sciences sociales*, 11(1), 65–92.

<https://doi.org/10.7202/1035933ar>

Résumé de l'article

Nous nous interrogeons sur les méthodologies à utiliser pour analyser le contenu des programmes et documents d'accompagnement publiés par le ministère de l'Éducation Nationale en France. Ce corpus textuel présente une forme particulière dans la mesure où les paragraphes de texte sont complétés de tableaux et graphismes variés. L'article présente une étude exploratoire et précise d'abord pourquoi et comment nous articulons deux méthodologies d'analyse logicielle (Alceste et Tropes) et une analyse dite « manuelle ». Les résultats obtenus indiquent des fonctions similaires et une complémentarité. Nous précisons enfin les conditions à respecter pour mener à bien ce travail, les avantages et les limites inhérentes.

# **Analyse de données textuelles informatisée: l'articulation de trois méthodologies, avantages et limites**

**MARTINE PAINDORGE**

Université de Lorraine, Laboratoire STEF-ENS Cachan-Ifé

**JACQUES KERNEIS**

Université de Bretagne Occidentale, Laboratoire CREAD

**VALÉRIE FONTANIEU**

École normale supérieure de Lyon, Institut Français de l'Éducation

## **Introduction**

**D**ans le cadre d'une recherche qui vise à comprendre les processus de sélection et de transformation des ressources par les enseignants (ReVEA<sup>1</sup>), nous nous interrogeons sur la ou les méthodologies à utiliser pour analyser le contenu des programmes et « ressources pour l'enseignement » publiées par le Ministère de l'Éducation nationale.

Le présent article a pour objectif de rendre compte de notre recherche méthodologique, élaborée avec trois expertises de chercheurs différentes. L'intention n'est pas ici de réaliser une comparaison de plusieurs logiciels mais de présenter une première approche exploratoire mise en œuvre pour étudier ce corpus. La

---

<sup>1</sup> Le projet ReVEA (Ressources Vivantes pour l'Enseignement et l'Apprentissage) est soutenu par l'Agence Nationale de la Recherche. Une description du projet est accessible à l'adresse internet <http://anr-revea.fr/>.

recherche présentée est menée à propos d'un domaine disciplinaire, les Sciences et Techniques Industrielles (STI), et est conduite dans la perspective d'une extension à d'autres enseignements.

Nous présenterons d'abord les caractéristiques du corpus retenu et les questions que soulève cette analyse. Puis nous décrirons la méthodologie et discuterons les résultats obtenus avant de préciser les avantages et les limites de ce travail.

## 1. Un corpus particulier

En France, l'enseignement fait l'objet d'une prescription qui se concrétise notamment par des programmes dans chaque discipline, publiés en tant que texte officiel. Pour accompagner la mise en œuvre de ces programmes, le ministère édite des documents dits d'accompagnement, dénommés parfois « ressources pour l'enseignement ». Ceux-ci n'ont pas le même statut réglementaire que les programmes mais revêtent un caractère semi-officiel, eu égard à leur origine. Les deux catégories de textes, programmes et documents d'accompagnement (que nous distinguerons ainsi dans la suite du texte) sont susceptibles, d'une part, d'indiquer des ressources aux enseignants et, d'autre part, d'influer sur la sélection qu'ils effectuent; aussi est-il décidé d'analyser ces documents<sup>2</sup>.

<sup>2</sup> Nous avons retenu, plus spécifiquement, les documents suivants : *Enseignements technologiques (transversaux et spécifiques des spécialités architecture et construction, énergies et environnement, innovation technologique et éco-conception, systèmes d'information et numérique) du cycle terminal de la série STI2D*, Ministère de l'Éducation nationale, de l'enseignement et de la recherche, 2011, <http://www.education.gouv.fr/cid55415/mene1104262a.html>, site consulté le 31 mars 2015; « Programmes du collège. Programmes de l'enseignement de technologie », *Bulletin officiel spécial n° 6 du 28 août 2008*, Ministère de l'Éducation nationale, 2008, [http://media.education.gouv.fr/file/special\\_6/53/1/Programme\\_technologie\\_33531.pdf](http://media.education.gouv.fr/file/special_6/53/1/Programme_technologie_33531.pdf), site consulté le 31 mars 2015; *Ressources pour le cycle terminal. Enseignements technologiques transversaux et enseignements spécifiques (série STI2D)*, Éduscol, Ministère de l'Éducation nationale (DGESCO), 2011, [http://media.eduscol.education.fr/file/STI2D/15/2/LyceeGT\\_Ressources\\_STI2D\\_T\\_Enseignement\\_Technologique\\_Specifiques\\_182152.pdf](http://media.eduscol.education.fr/file/STI2D/15/2/LyceeGT_Ressources_STI2D_T_Enseignement_Technologique_Specifiques_182152.pdf), site consulté le 31 mars 2015; *Technologie au collège. Ressources pour faire la classe en 6<sup>ème</sup>, 5<sup>ème</sup>, 4<sup>ème</sup> et 3<sup>ème</sup>*, Éduscol, Ministère de l'Éducation nationale (DGESCO), 2009, [http://cache.media.eduscol.education.fr/file/Programmes/16/6/College\\_Technologie\\_Ressources\\_114166.pdf](http://cache.media.eduscol.education.fr/file/Programmes/16/6/College_Technologie_Ressources_114166.pdf), site consulté le 31 mars 2015.

Nous cherchons à repérer quelles ressources sont indiquées dans les programmes et les documents qui les accompagnent. Il s'agit également d'identifier des éléments susceptibles d'indiquer une progressivité dans les prescriptions, mais aussi d'éventuelles spécificités d'un ou plusieurs enseignements.

L'esprit dans lequel ont été conçus le processus d'élaboration des programmes scolaires et le cadrage de leur rédaction semble perdurer jusqu'à aujourd'hui<sup>3</sup>. La description par Dominique Raulin<sup>4</sup> d'une commission chargée des programmes et des cahiers des charges pour la rédaction se retrouve encore en grande partie actuellement et explique certainement une ressemblance entre les disciplines quant à la forme des textes publiés par le ministère pour les contenus d'enseignement. Ces recommandations influent sur la forme des textes : les programmes se présentent généralement sous forme d'un texte comprenant des énumérations incluses sous forme de tableaux, les documents d'accompagnement comprennent des textes, des tableaux, mais aussi une part non négligeable de graphismes de différents types (schéma d'ensemble, représentation d'un processus, illustration de l'enchaînement de plusieurs apprentissages, etc.).

Les textes concernant les Sciences et Techniques Industrielles, publiés sous forme numérique par le ministère de l'Éducation nationale, constituent dix fichiers, soit environ 400 pages (150 000 mots environ). Dans cet article, nous nous limitons à l'analyse des données textuelles. Plusieurs méthodologies semblent envisageables, chacune présentant des intérêts, mais aussi des limites.

<sup>3</sup> La charte 2013 des programmes n'existait pas pour les textes que nous étudions dans l'article; elle ne contredit pas le texte de 1992. Voir à l'appui, respectivement : *Charte des programmes. Charte relative à l'élaboration, à la mise en œuvre et au suivi des programmes d'enseignement ainsi qu'aux modalités d'évaluation des élèves dans l'enseignement scolaire*, Conseil supérieur des programmes, Ministère de l'Éducation nationale, de l'enseignement et de la recherche, 2013, [http://cache.media.education.gouv.fr/file/04\\_Avril/37/5/charte\\_programme\\_csp\\_312375.pdf](http://cache.media.education.gouv.fr/file/04_Avril/37/5/charte_programme_csp_312375.pdf), site consulté le 8 juillet 2015 et « Charte des programmes », *Bulletin officiel de l'Éducation nationale des lois et règlements n° 8 du 20 février 1992*, Ministère de l'Éducation nationale, 1992, <http://artheque.ens-cachan.fr/items/show/5647>, site consulté le 31 mars 2015.

<sup>4</sup> Dominique Raulin, *Les programmes scolaires*, Paris, Retz, 2006, p. 179-181.

## 2. Quelle(s) méthodologie(s) choisir?

Le travail sur les textes officiels s'inscrit ici en didactique et la recherche bibliographique réalisée indique l'absence d'articles relatifs aux méthodologies d'analyse de tels types de texte dans les revues de didactique parues depuis 1990, période du développement des outils logiciels d'analyse textuelle. En revanche, si nous étendons la requête aux différents champs des sciences humaines, les articles portant sur l'analyse textuelle informatisée sont plus nombreux.

Un premier ensemble est composé de publications proposant diverses catégorisations des logiciels existants. Les classifications mettent en relation la ou les méthodes d'analyse sous-jacentes dans ces nouveaux outils et les usages possibles. Ainsi, Jacques Jenny montre la relation entre les modes d'approche pour l'analyse et les présupposés implicites, inscrits dans des paradigmes, qui président aux choix. L'auteur distingue alors des approches lexicométrique, socio-sémantique (analyse thématique), par réseaux de mots associés, l'analyse propositionnelle et prédicative des discours. Le même article catégorise une trentaine de logiciels, en leur associant des usages possibles. Jacques Jenny<sup>5</sup> complète sa bibliographie de 1997 sans remettre en question sa première classification<sup>6</sup>. Bernard Fallery et Florence Rodhain, après avoir énoncé des facteurs pour le choix d'un type d'analyse, ne distinguent, eux, que quatre grands types d'analyse de données textuelles, en prenant appui sur essentiellement quatre logiciels pour illustrer leur propos. L'analyse, pour ces auteurs, peut être lexicale (logiciel Alceste), linguistique (logiciel Tropes), cognitive (logiciel Decision Explorer), thématique (logiciel Nvivo)<sup>7</sup>.

<sup>5</sup> Jacques Jenny, *À propos des logiciels d'analyse textuelle pratiqués en France pour la recherche en sciences sociales*, 2002, <http://www.cmh.pro.ens.fr/bms/arcati/BMS54-Jenny-New.htm>, site consulté le 31 mars 2015.

<sup>6</sup> Jacques Jenny, *Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine. État des lieux et essai de classification*, 1997, [http://jacquesjenny.com/legs-sociologique/?page\\_id=1253](http://jacquesjenny.com/legs-sociologique/?page_id=1253), site consulté le 8 janvier 2015.

<sup>7</sup> Bernard Fallery et Florence Rodhain, *Quatre approches pour l'analyse de données textuelles: lexicale, linguistique, cognitive, thématique*, 2007, <http://>

Les catégorisations proposées par Christophe Lejeune mettent plutôt en relation le principe de construction des catégories d'analyse et l'intervention de l'utilisateur. Christophe Lejeune distingue trois familles d'outils logiciels (génération des catégories automatique, construction par l'utilisateur, proposition de registres)<sup>8</sup>. Le même auteur positionne sur deux axes (montrer-calculer et explorer-analyser) cinq familles de logiciels en fonction de leurs fonctionnalités (lexicométrie, concordances, automates, registres, outils réflexifs)<sup>9</sup>. Il propose une liste de soixante-trois produits, classés dans trois catégories : réflexive, registre, libre<sup>10</sup>.

Si classer les logiciels par catégorie rend plus facile l'élaboration d'un panorama, toute classification présente des limites et certains outils peuvent relever de plusieurs familles. Pour autant, ces classifications constituent une indication pour les chercheurs qui, pour un même corpus, peuvent souhaiter mobiliser les fonctionnalités de chaque outil dans une perspective de complémentarité. Ainsi, Didier Demazière met à l'épreuve quatre outils (Alceste, Calliope, Prospéro, et Trideux) pour analyser un corpus constitué de résumés proposés pour une conférence en sociologie<sup>11</sup>. Philippe Wanlin fait le point sur les atouts et désavantages de Lexico et de Nvivo, mais discute aussi les apports combinés d'une méthode logicielle et d'un traitement « manuel » des données<sup>12</sup>.

---

[www.strategie-aims.com/events/conferences/7-xvieme-conference-de-l-aims/communications/2078-quatre-approches-pour-lanalyse-de-donnees-textuelles-lexicale-linguistique-cognitive-thematique/download](http://www.strategie-aims.com/events/conferences/7-xvieme-conference-de-l-aims/communications/2078-quatre-approches-pour-lanalyse-de-donnees-textuelles-lexicale-linguistique-cognitive-thematique/download), site consulté le 31 mars 2015.

<sup>8</sup> Christophe Lejeune, « Au fil de l'interprétation, l'apport des registres aux logiciels d'analyse qualitative », *Swiss Journal of Sociology*, vol. 3, n° 34, 2008, p. 593-603.

<sup>9</sup> Christophe Lejeune, *Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative*, 2010, <http://hdl.handle.net/2268/61098>, site consulté le 31 mars 2015.

<sup>10</sup> Christophe Lejeune, *Méthodes qualitatives informatisées*, <http://www.squash.ulg.ac.be/logiciels>, site consulté le 23 septembre 2015.

<sup>11</sup> Didier Demazière, *Des logiciels d'analyse textuelle au service de l'imagination sociologique*, 2005, <http://bms.revues.org/978>, site consulté le 31 mars 2015.

<sup>12</sup> Philippe Wanlin, *L'analyse de contenu comme méthode d'analyse qualitative d'entretiens : une comparaison entre les traitements manuels et l'utilisation de logiciels*, 2007, [http://www.recherche-qualitative.qc.ca/documents/files/revue/hors\\_serie/hors\\_serie\\_v3/Wanlin2.pdf](http://www.recherche-qualitative.qc.ca/documents/files/revue/hors_serie/hors_serie_v3/Wanlin2.pdf), site consulté le 28 mars 2015.

Mathieu Brugidou et Michèle Moine réalisent une analyse en utilisant les logiciels Alceste et Tropes en fonction d'objectifs précis: Alceste doit leur permettre d'explorer les oppositions et les équivalences thématiques qui structurent les discours, Tropes est choisi, d'une part, pour analyser les catégories de mots avec l'intention de vérifier l'hypothèse d'une structure inégale du discours et, d'autre part, construire un lexique d'une notion<sup>13</sup>.

Il convient de remarquer que, dans toute la bibliographie consultée, le logiciel Alceste fait l'objet de publications régulières et nombreuses depuis sa création en 1986. Les chercheurs rendent compte d'usages variés des fonctions du logiciel, par exemple Nikos Kalampalikis en psychologie sociale<sup>14</sup>. Réalisé en économie et gestion, le travail de Dominique Peyrat-Guillard et Daniel Dufresne utilise ce logiciel avec l'hypothèse que l'analyse des référentiels devrait permettre de mettre en évidence une proximité de langage entre services<sup>15</sup>. Centrée sur l'utilisation de l'outil, Monique Dalud-Vincent pose la question de l'influence de la forme du corpus (ici la prise en compte ou non de l'interviewer ou de la correction par le transcripteur des propos de l'interviewé) sur les résultats obtenus avec le logiciel<sup>16</sup>. Monique Dalud-Vincent montre par ailleurs que le paramétrage du logiciel a une influence sur les résultats, et donc sur l'analyse produite<sup>17</sup>.

<sup>13</sup> Mathieu Brugidou et Michèle Moine, *Normes émergentes et stigmatisations. Une analyse comparative à partir des deux questions ouvertes sur les raisons de ne pas trier les déchets et de ne pas faire d'économie d'énergie*, 2010, [http://lexicométrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0885-0896\\_147-Brugidou.pdf](http://lexicométrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0885-0896_147-Brugidou.pdf), site consulté le 26 août 2015.

<sup>14</sup> Nikos Kalampalikis, « L'apport de la méthode Alceste dans l'étude des représentations sociales », dans Jean-Claude Abric (dir.), *Méthodes d'étude des représentations sociales*, Paris, Éditions Érès, 2003, p. 147-163.

<sup>15</sup> Dominique Peyrat-Guillard et Daniel Dufresne, *L'analyse exploratoire de référentiels de compétences avec Alceste : une aide à la lecture de l'analyste*, 2010, [http://lexicométrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0141-0150\\_019-Peyrat.pdf](http://lexicométrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0141-0150_019-Peyrat.pdf), site consulté le 26 août 2015.

<sup>16</sup> Monique Dalud-Vincent, « Les "choix" du sociologue avec Alceste. De la forme du corpus aux résultats obtenus », *Bulletin de méthodologie sociologique*, n° 105, 2010, p. 25-52.

<sup>17</sup> Monique Dalud-Vincent, « Les "choix" du sociologue avec Alceste. Du paramétrage des unités de contexte aux résultats obtenus », *Bulletin de méthodologie sociologique*, n° 107, 2010, p. 23-48.

Aussi, la quantité de logiciels, la diversité des usages d'un même produit, les possibilités d'allier plusieurs solutions pour un même corpus fait émerger le besoin de disposer de critères de choix. Mathieu Brugidou *et al.*<sup>18</sup> s'accordent avec Bernard Fallery et Florence Rodhain<sup>19</sup> pour prendre en considération la problématique de recherche, le profil de l'analyste, le type de corpus, le type de traitement en fonction des résultats attendus. Ils notent, également, que ces facteurs ne se situent pas au même niveau et sont en perpétuelle interaction.

Dans notre travail, l'intention est de rechercher dans les programmes et documents d'accompagnement si des ressources sont indiquées, si elles sont associées à une thématique, si elles sont spécifiques à une discipline, à un niveau d'enseignement. Nous ne cherchons pas à soumettre ensuite les résultats aux enseignants, comme l'ont fait Fabienne Venant et Jean-François Maheux<sup>20</sup>. Le logiciel Lexico, utilisé par les chercheurs québécois, fondé sur une méthode textométrique, ne paraît pas pertinent pour notre étude. Une liste exhaustive des ressources, de même qu'une liste de leur catégorisation peuvent être obtenues par une analyse sémantique que nous qualifierons de « manuelle », car réalisée sans aide logicielle. Mais cette méthode, outre le temps qu'elle nécessite, semble peu appropriée pour examiner les questions d'unité du champ disciplinaire. Le travail réalisé par Dominique Peyrat-Guillard et Daniel Dufresne<sup>21</sup> nous conduit à émettre l'hypothèse qu'une analyse statistique effectuée avec le logiciel Alceste est susceptible de dégager des groupes dont nous étudierons la proximité avec un niveau d'enseignement. De même, l'utilisation de Tropes par Mathieu Brugidou et Michèle Moine<sup>22</sup> pour étudier

<sup>18</sup> Mathieu Brugidou *et al.*, *Les facteurs de choix et d'utilisation de logiciels d'Analyse de Données Textuelles*, 2000, <http://lexicometrica.univ-paris3.fr/jadt/jadt2000/pdf/04/04.pdf>, site consulté le 31 mars 2015.

<sup>19</sup> Bernard Fallery et Florence Rodhain, *op.cit.*

<sup>20</sup> Fabienne Venant et Jean-François Maheux, *Réforme de l'enseignement au Québec : une visite guidée par la textométrie*, 2014, <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/54-JADT2014.pdf>, site consulté le 26 août 2015.

<sup>21</sup> Dominique Peyrat-Guillard et Daniel Dufresne, *op. cit.*

<sup>22</sup> Mathieu Brugidou et Michèle Moine, *op. cit.*

les catégories de mots et la création d'un lexique incite à mettre en œuvre le logiciel pour repérer les ressources, qui sont des substantifs, et créer un lexique autour de la notion de ressource.

Chacune des méthodes, basée ou non sur l'utilisation d'un logiciel, apporte des résultats de nature différente. Dans ces conditions, une méthodologie articulant l'utilisation de différents logiciels, associée à une analyse « manuelle », semble susceptible d'apporter des informations plus nombreuses et plus variées, lesquelles peuvent servir à valider les résultats d'une des méthodes, notamment la méthode « manuelle » pour laquelle la subjectivité du chercheur est perçue comme un écueil. Une telle méthodologie articulée, qui reste à construire, offre également la possibilité de croiser les premiers résultats obtenus afin d'émettre de nouvelles hypothèses. Nous décrivons ensuite notre processus d'élaboration.

### **3. Méthodologie**

Dans un premier temps, à partir d'un corpus commun, il s'agit de mettre en œuvre chaque méthodologie d'analyse. La première méthodologie utilise le logiciel Alceste (version 2012), la seconde, le logiciel Tropes (version 8.4), la troisième est une analyse « manuelle ». Chacune, prise individuellement, apporte des résultats qui sont repris pour une seconde phase. Ce second temps, basé sur la discussion entre chercheurs, est consacré à interroger les premiers résultats et à en produire de nouveaux en effectuant de nouveaux traitements.

Pour faciliter la lecture, nous présentons successivement les processus mis en œuvre et les résultats intermédiaires de la première phase avant de détailler les nouvelles actions qu'ils ont suscitées.

#### **3.1. Préparation d'un corpus commun**

Le travail présenté ici étant exploratoire, nous nous limitons aux programmes et documents d'accompagnement de deux enseignements étudiés dans le projet ReVEA, lesquels enseignements s'inscrivent dans le domaine de spécialité d'une des auteurs : la

technologie collège et les Sciences et Techniques Industrielles Développement Durable (STI2D), qui tous deux relèvent du domaine « Sciences et Techniques Industrielles ». Cette décision permet de travailler sur des textes de deux niveaux scolaires différents (collège et lycée) et, ainsi, sur des contenus ayant une logique propre tout en s'inscrivant dans une unité déclarée par les rédacteurs du ministère, l'unité des disciplines technologiques. Ainsi pourront être étudiées les questions de ressources spécifiques à une discipline ou à un niveau d'enseignement et les questions de ressources liées à une thématique. Les quatre documents sont publiés sur le site [educ.gouv](http://educ.gouv) et téléchargeables au format pdf. Pour la technologie collège, les programmes comprennent une introduction commune aux mathématiques, sciences de la vie et de la terre, sciences physiques et chimiques, et à la technologie. Cette partie est exclue car les questions de recherche ne portent pas sur l'interdisciplinarité; le programme de technologie n'est pris en compte qu'à partir de la page 9.

La première étape consiste en une préparation du corpus. Afin d'être utilisable par les deux logiciels, chaque fichier au format pdf est transformé en fichier au format txt, ce qui implique la suppression des graphismes et la suppression des retours à la ligne dus notamment au texte présent dans les tableaux. La lecture préalable de l'ensemble des textes conduit à traiter séparément les programmes, rédigés sous forme de tableaux avec des expressions courtes (par exemple une liste de compétences ou de niveaux à atteindre), et les documents d'accompagnement, présentant des textes beaucoup plus rédigés. Nous ne présentons ici que l'étude menée pour le fichier regroupant les documents d'accompagnement de technologie collège et enseignement STI2D de lycée. L'homogénéité du corpus (liée au cadrage de leur production), ainsi que le nombre de mots (70 000 environ) permettent de respecter les conditions de validité indiquées par Max Reinert et de ses collaborateurs<sup>23</sup>.

<sup>23</sup> Max Reinert, *Alceste, un outil d'aide pour l'analyse du discours. Notice simplifiée (de la version de base commune aux versions 4.x)*, <http://www.ling.uqam.ca/forum/satoman/images/AlcesteNoticeSimplifiee.pdf>, site consulté le

## 3.2. L'analyse logicielle réalisée avec le logiciel Alceste

### 3.2.1. Quelques éléments sur le fonctionnement d'Alceste

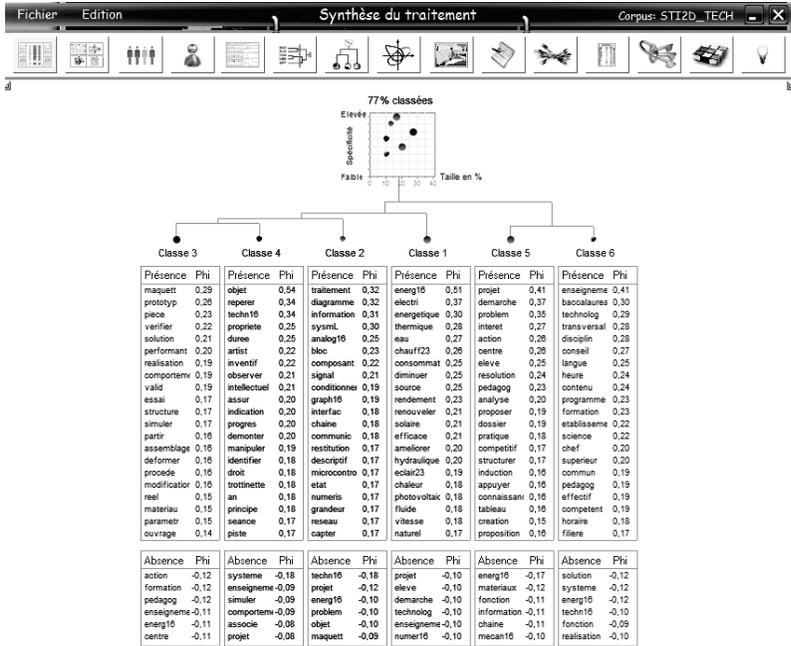
Nous rappelons maintenant plusieurs informations relatives au logiciel, utiles pour l'interprétation des résultats. Alceste<sup>24</sup> crée d'abord un lexique du corpus en recensant les mots sous leur forme lemmatisée (6 063 formes distinctes dans le corpus). Donc, un mot avec deux sens distincts selon le contexte sera comptabilisé sous le même lemme si l'on n'agit pas pour rectifier le lexique. Le logiciel procède ensuite à une segmentation du texte en tenant compte de la ponctuation ainsi que d'un nombre suffisant de formes analysables par segment. Ces segments sont appelés les u.c.e. (unités de contexte élémentaires). Suit un classement des u.c.e. selon les cooccurrences de mots. Ce classement est la résultante de deux partitions obtenues à partir de deux longueurs de segments distinctes, établies en nombre de formes que contiennent ces segments (il s'agit des paramètres B21 et B22 dans la suite). C'est ainsi qu'un certain nombre d'u.c.e. se trouvent, au final, non classées. Le premier résultat affiché par Alceste, appelé classes par le logiciel et considéré comme des mondes lexicaux par les chercheurs, est constitué des formes contenues dans les u.c.e., classées selon l'ordre décroissant de leur lien avec la classe (voir la figure 1). La méthode est entièrement statistique et n'opère pas sur le sens des mots. Par conséquent, une forme d'occurrence forte peut ne pas apparaître en raison de l'absence de lien à un monde lexical en particulier, lorsque le mot est important dans le corpus, mais pas lié à une thématique en particulier. De même, des notions intéressantes qualitativement n'apparaîtront pas si elles sont peu fréquemment évoquées dans le corpus (c'est le cas notamment des hapax, au nombre de 2 683). Ce traitement automatique s'effectue sans *a priori* pour l'obtention des mondes lexicaux; seule l'interprétation d'un monde lexical peut devenir éventuellement subjective.

---

24 juillet 2015; Max Reinart *et al.*, *Alceste, logiciel de données textuelles (non daté)*, <http://www.image-zafar.com/Logiciel.html>, site consulté le 22 juillet 2015.

<sup>24</sup> Alceste, comme Analyse des Lexèmes Co-occurents dans un Ensemble de Segments de Texte.

Figure 1  
Exemple des classes affichées par le logiciel Alceste



### 3.2.2. Le processus mis en œuvre lors de l'utilisation d'Alceste

Nous considérons que deux variables sont susceptibles d'influer : le niveau d'enseignement et le type de texte (nommé ici ressources). Ces variables caractéristiques des textes sont introduites par une ligne étoilée (voir le tableau 1).

## Tableau 1

## Exemple de préparation du corpus pour Alceste

\*\*\*\* \*rep\_1 \*Niv\_STI2D \*Typ\_ress

*Ce document d'accompagnement a pour objectif d'aider les équipes enseignantes intervenant dans le cycle terminal du baccalauréat STI2D à organiser concrètement la formation. Il ne reprend donc...*

Comme cela est évoqué dans les travaux de Monique Dalud<sup>25</sup>, des essais successifs de paramétrages sur des longueurs (B21 et B22) des segments à classer permettent d'observer les similitudes et les différences de formes obtenues dans les classes. Il s'agit là de tenter de s'adapter à la densité des textes, en prenant en compte un nombre de mots plus ou moins important pour exprimer les idées, et, de manière concomitante, de tester la stabilité des résultats. Chaque paramétrage conduit à un pourcentage d'u.c.e. classées et un nombre de classes différents. Le tableau 2 indique les six premières formes des différentes classes obtenues avec trois paramétrages différents (l'identifiant 121 correspond au paramétrage par défaut).

<sup>25</sup> Monique Dalud-Vincent, *op.cit.*

Tableau 2  
Analyses successives avec Alceste

Traitement 121 (par défaut)	Classe 3		Classe 4	Classe 2	Classe 1	
B21 = 10 B22 = 12 86 % u.c.e. classées	Materiaux Maquett Solution Procédé Piece Objet		Traitement Diagramme Information Sysml Analog Bloc	Energ Electri Energetique Thermique Eau Chauff	Projet Enseignement Pedagog Technologie Demarche Formation	
Traitement 122	Classe 3		Classe 2	Classe 1	Classe 4	Classe 5
B21 = 8 B22 = 10 85 % u.c.e. classées	Materiaux Solution Maquett Procédé Objet Piece		Traitement Information Analog Diagramme Composant Grandeur	Energ Electri Energetique Thermique Eau Chauff	Enseignement Formation Transversal Pedagog Technologie Competent	Projet Demarche Problem Etape Resolution Creation
Traitement 123	Classe 3	Classe 4	Classe 2	Classe 1	Classe 6	Classe 5
B21 = 12 B22 = 14 77 % u.c.e. classées	Maquett Prototyp Piece Verifier Solution Performant	Objet Reperer Techn Propriete Duree Artist	Traitement Diagramme Information Sysml Analog Bloc	Energ Electri Energetique Thermique Eau Chauff	Enseignement Baccalaureat Technolog Transversal Disciplin Conseil	Projet Demarche Problem Interet Action Centre
Phi discipline	STI2D 0,06	Techno 0,58	STI2D 0,15	STI2D 0,10	STI2D 0,11	STI2D 0,06

Dans le traitement 123, les classes 3 et 4 forment une sous-division de la classe 3 du traitement par défaut (et de la classe 3 du traitement 122) et les classes 5 et 6, une sous-division de la classe 1 du traitement par défaut.

Le traitement 123, que nous retenons, semble apporter plus de précisions, bien que le pourcentage des u.c.e. classées ait diminué, mais reste acceptable. Il permet d'obtenir six mondes lexicaux que nous examinerons au début de la seconde phase.

### 3. 3. L'analyse logicielle réalisée avec le logiciel Tropes

#### 3.3.1. Quelques éléments sur le fonctionnement de Tropes

Pierre Molette<sup>26</sup>, le manuel de référence Tropes<sup>27</sup> et l'aide en ligne<sup>28</sup> du logiciel apportent des informations sur les méthodes utilisées par le logiciel. Ces informations permettront de mieux comprendre et de nuancer les résultats obtenus. Pierre Molette indique que le logiciel procède à des analyses morphosyntaxique et sémantique avant d'effectuer des statistiques<sup>29</sup>. L'analyse morphosyntaxique conduit à repérer les différentes catégories grammaticales et à lemmatiser. Mais à la différence d'Alceste, qui conserve les lemmes obtenus pour ensuite étudier les cooccurrences, Tropes procède à un regroupement en classes d'équivalents qui associent les références (noms communs ou noms propres) qui apparaissent fréquemment dans le texte et qui possèdent une signification voisine. Ces classes d'équivalents sont regroupées de plus en plus largement en Références utilisées, puis Univers de référence 2, eux-mêmes regroupés en Univers de référence 1. C'est donc un résultat basé sur le regroupement à plusieurs niveaux que fournit Tropes, comme l'illustre Pierre Molette<sup>30</sup>.

#### 3.3.2. Le processus mis en œuvre lors de l'utilisation de Tropes

Parmi les nombreux résultats affichés après l'analyse automatique par le logiciel, nous nous concentrons sur ceux qui sont susceptibles d'apporter des éléments de réponse à nos questions. Ainsi, par exemple, nous ne sélectionnons que la liste des substantifs, avec l'hypothèse qu'elle indiquera des ressources, et nous excluons

<sup>26</sup> Pierre Molette, *De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste*, 2009, <http://www.tropes.fr/PierreMolette-CommunicationColloquePsychoTarbesJuin2009.pdf>, site consulté le 24 juillet 2015.

<sup>27</sup> *Tropes, version 8.0, manuel de référence*, 2011, <http://www.tropes.fr/ManuelDeTropesVF800a.pdf>, site consulté le 20 février 2015.

<sup>28</sup> Nous avons consulté plus particulièrement dans le menu Aide le chemin Aide/Index/Analyse (notions).

<sup>29</sup> Pierre Molette, *op. cit.*, p. 4-6.

<sup>30</sup> Pierre Molette, *ibid.*, p. 11.

ce qui concerne les autres catégories grammaticales ainsi que le style du discours. Cette liste des substantifs est exportée vers un fichier tableur afin de faciliter la comparaison avec les résultats des autres méthodologies (voir le tableau 3).

Tableau 3

Extraction de la liste des substantifs proposée par Tropes (les dix premiers)

<b>Terme</b>	<b>Occurrence</b>
système	622
énergie	313
projet	298
solution	295
information	226
réalisation	226
élève	218
matériau	212
fonction	211
objet	204

Dans cette liste, le terme « système » apparaît comme étant le plus fréquent.

D'autres fonctionnalités sont disponibles, elles seront activées en fonction de la discussion des résultats des trois analyses.

### 3. 4. L'analyse qualifiée de « manuelle »

#### 3.4.1. *Quelques éléments sur l'analyse « manuelle »*

L'analyse réalisée prend appui sur les techniques décrites dans les ouvrages de Laurence Bardin<sup>31</sup> et d'Annick Weil-Barais<sup>32</sup>. Le chercheur définit des descripteurs susceptibles d'apporter des éléments de réponse aux questions de recherche, lesquels sont rassemblés dans une grille d'analyse renseignée ensuite grâce à une lecture analytique du texte. La grille renseignée comprend les extraits de texte sélectionnés en fonction des descripteurs (ce qui permet d'apporter la preuve) et des éléments pour l'interprétation. Aussi, même si des traces de l'analyse sont conservées

<sup>31</sup> Laurence Bardin, *L'analyse de contenu*, Paris, Presses Universitaires de France, 1977.

<sup>32</sup> Annick Weil-Barais, *Les méthodes en psychologie*, Paris, Bréal, 1997.

pour pouvoir être reprises par un autre chercheur, toutes les étapes comportent un risque lié à la subjectivité de l'être humain.

### 3.4.2. *Le processus mis en œuvre lors de l'analyse « manuelle »*

L'analyse sémantique est réalisée en utilisant des descripteurs comme « à partir de », « prend appui », mais la compréhension du texte par le chercheur permet aussi de sélectionner d'autres extraits de texte. Pour chaque sélection, la page est repérée, l'extrait jugé significatif est recopié et ce qui est considéré comme la ressource est souligné. On obtient ainsi une liste de ressources. Les informations recueillies sont ensuite catégorisées, soit par fonction (représentation, simulation, communication), soit par nature (schéma, base de données, logiciel, matériel, ressources humaines, modèle). La catégorisation n'est pas établie *a priori*, elle correspond à un regroupement *a posteriori*, en cohérence avec d'autres travaux déjà réalisés et avec les questions de recherche (voir les tableaux 4 et 5).

#### Tableau 4

Processus et résultats obtenus par l'analyse « manuelle »

*p.13 : Études de dossiers technologiques s'appuyant sur des documentations techniques sélectionnées  
Le phasage des opérations est traité à partir du planning général de réalisation d'une construction.  
L'impact carbone est abordé au travers des FDES et de logiciels spécifiques d'aide à la décision.*

*p.14 : L'utilisation d'outils de simulation numérique  
Les études de dossiers technologiques peuvent prendre appui sur des études comparées ou sur des opérations de réhabilitation.  
Les études de dossiers technologiques*

**Extrait de l'analyse sémantique, les ressources repérées sont soulignées.**

Tableau 5

Catégories de ressources obtenues après analyse manuelle du fichier « Bulletin officiel STI2D »

Outil de représentation	représentations numériques, représentations des répartitions et de l'évolution des grandeurs énergétiques (diagramme, vidéo, image), représentations associées au codage de l'information : variables, encapsulation des données
Outil de simulation	bibliothèques, maquette numérique fournie, logiciel adapté pour simuler le comportement structurel, outils de simulation numérique, logiciel de simulation du comportement énergétique, logiciel de simulation mécanique, progiciels de simulation des procédés
Outil de communication	outils d'échanges, de suivi et d'animation (ENT, revues de projet, cartes mentales, flux opérationnels) ainsi que ceux de formalisation et de synthèses (bases de données, bilan d'expérience) en vue d'une exploitation collaborative
Graphismes	diagrammes adaptés SysML, graphes de flux d'énergie (schéma cinématique, schéma électrique, schéma fluide), schéma architectural (mécanique, énergétique, informationnel), diagrammes états-transitions simple
Bases de données	bases de données de brevets, bases de données de coûts, bases de connaissances numériques
Logiciel	progiciel traitant uniquement des impacts environnementaux, logiciels spécifiques d'aide à la décision, bibliothèques graphiques, modéleur numérique, maquettes volumiques, progiciels volumiques intégrant un module de traitement du comportement dynamique des systèmes
Matériel	interface homme-système, constituants ou composants d'une chaîne d'énergie, systèmes didactiques simples, matériel grand public communicant, réseau local de type Ethernet, modem-routeur ADSL, produit industriel ou d'un système, objets manufacturés, maquettes instrumentées, systèmes instrumentés, maquettes didactisées, dispositifs expérimentaux simples
Ressources humaines	visites de sites industriels, de conférences
Modèle	modèle de présentation «torseur des actions mécaniques»
Document technique/ document de référence	notice technique d'un système, FDES (fiches de déclaration environnementale et sanitaires), documentations techniques, formulaires pour la détermination des pertes de charges des réseaux fluidiques, spécifications normalisées, normes en vigueur
Documents produits par élève, professeur	relevés sur terrain (photos, topographique, échantillon), dossiers technologiques

Après cette première phase individuelle commence l'analyse croisée.

### 3.5. L'analyse croisée

Les résultats obtenus précédemment sont d'abord analysés, puis discutés. De nouvelles opérations, combinant les informations issues du premier traitement et les fonctionnalités offertes par chaque méthode, sont réalisées.

### 3.5.1. *Discussion sur les premiers résultats*

Nous ne commentons pas ici tous les résultats obtenus, mais, pour illustrer l'analyse croisée, nous présentons les principaux points qui ont suscité des traitements complémentaires.

Le logiciel Alceste, avec le paramétrage retenu, propose six mondes lexicaux définis par des formes lemmatisées, classées en fonction de leur importance statistique. Ainsi, la classe 1 rassemble des formes liées à l'énergie, aux sources d'énergie tandis que la classe 2 regroupe des formes liées à la notion de système, avec un point de vue descriptif (diagramme, chaîne, bloc...). Il convient de noter la présence de la forme « SysML » qui correspond au langage de description d'un système automatisé. Cette observation est remarquable quand on sait que cet outil a été introduit pour la première fois dans les nouveaux programmes de STI2D et fut l'objet de demandes de formation par les enseignants de cette discipline. Il est décidé d'explorer davantage cette forme. La classe 3 évoque la démarche technologique, comprenant différentes phases (recherche de solutions, prototypage, validation...) tandis que la classe 4 comprend des formes reliées au collège (trottinette, démonter, observer...). Les phi d'association des variables étoilées aux classes confirment que la classe 4 est particulièrement liée à la technologie au niveau du collège, les autres n'étant pas particulièrement liées à l'une ou l'autre des disciplines (très faiblement à STI2D). Toujours dans la classe 4, il faut remarquer comme une opposition entre « objet », fortement lié à la classe, et « système », significativement absent; cela appelle de nouveaux traitements. La classe 5 renvoie à la démarche de projet et la classe 6 liste des formes évoquant la mise en œuvre, d'un point de vue administratif.

Le logiciel Tropes permet d'obtenir une liste de 993 substantifs classés par nombre d'occurrences. Les premiers termes sont à rapprocher des formes proposées dans les mondes lexicaux d'Alceste, l'ensemble des substantifs est à comparer avec la liste obtenue par l'analyse « manuelle ». Parmi les substantifs, la référence « système » est la plus fréquente (622 occurrences). Tropes propose également une listes des Références (qui per-

mettent de répondre à la question « Sur quoi porte le texte? »). Parmi les dix Références les plus fréquentes figurent « système », « énergie », « projet », « solution », « information », « fonction », « réalisation », « élève », « matériau », « objet ».

Au cours de l'analyse « manuelle », le chercheur remarque la présence dans les documents « ressources pour l'enseignement STI2D », à différents moments, de listes de normes indiquées comme des ressources pour l'enseignant. Cette importance accordée aux normes est en cohérence avec leur présence explicite pour la première fois dans les programmes. Or, le terme n'apparaît pas significativement dans les premières analyses logicielles. Cette différence appelle aussi de nouveaux traitements.

### 3.5.2. De nouvelles investigations

La présence inattendue, mais jugée significative par le logiciel, de la forme « SysML » dans la classe 2 proposée par Alceste a conduit à de nouvelles opérations récapitulées dans le tableau 6.

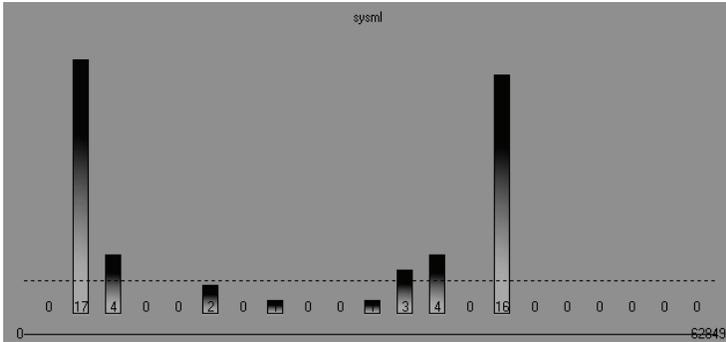
Tableau 6

Traitement complémentaire Tropes	Traitement complémentaire analyse manuelle	Traitement complémentaire Alceste
<ul style="list-style-type: none"> <li>- n'existe pas comme substantif</li> <li>- repérage grâce à la fonction «rechercher»</li> <li>- répartition en rafale ponctuelle, pas de présence au collège</li> </ul>	<ul style="list-style-type: none"> <li>- cité parmi les outils de description d'un système automatisé</li> <li>- uniquement cité dans les documents de STI2D</li> <li>- beaucoup de graphismes afférents</li> </ul>	<ul style="list-style-type: none"> <li>- le concordancier indique une présence forte en classe 2, une absence significative pour les classes 1 et 3</li> </ul>

L'absence de « SysML » parmi les substantifs de Tropes n'est pas étonnante car il s'agit d'un acronyme (Systems Modeling Language). Le graphe de répartition de Tropes (voir la figure 2) confirme l'analyse « manuelle » (présence uniquement en STI2D, indiquant là une spécificité du lycée), la répartition de la forme SysML dans les classes proposées par Alceste montre la spécificité de l'outil SysML pour la notion de système.

Figure 2

Graphe de répartition de « SysML » dans Tropes



La seconde investigation vise à documenter l'absence de mise en avant de « norme » par les logiciels alors que l'analyse sémantique en indiquait une certaine importance (voir le tableau 7).

Tableau 7

Traitement complémentaire Tropes	Traitement complémentaire analyse manuelle	Traitement complémentaire Alceste
<ul style="list-style-type: none"> <li>- scénario autour de «norme», regroupe les termes «standard», «normalisation», «norme»</li> <li>- liste possible de toutes les normes citées dans le texte</li> <li>- graphe de répartition indique deux rafales précises, toujours en STI2D</li> <li>- quatre liaisons avec «ressources»</li> </ul>	<ul style="list-style-type: none"> <li>- la liste repérée au cours de l'analyse individuelle sert de référence pour comparer avec les résultats Tropes</li> </ul>	<ul style="list-style-type: none"> <li>- la forme lemmatisée «norme», de fréquence d'occurrence égale à 68, n'est pas classée</li> </ul>

L'absence de « norme » parmi les formes significatives proposées par Alceste s'explique par le principe de fonctionnement du logiciel : son apparition dans le texte ne lui confère pas un lien plus fort avec une classe qu'avec les autres. Tropes permet d'affiner les résultats de l'analyse sémantique individuelle : quatre occurrences sont repérées. À chaque fois, le mot « ressources » est placé en aval (acté), il s'agit des expressions « normes et ressources ». Cette association tend à préciser le positionnement du prescripteur qui considère les normes comme une ressource, mais une ressource parmi d'autres.

Enfin, les formes « objet » et « système » ont attiré l'attention par leur présence antagoniste dans la classe 4 et le nombre d'occurrences de « système » repéré dans Tropes (648). Cette prégnance n'avait pas été repérée au fil de l'analyse « manuelle ».

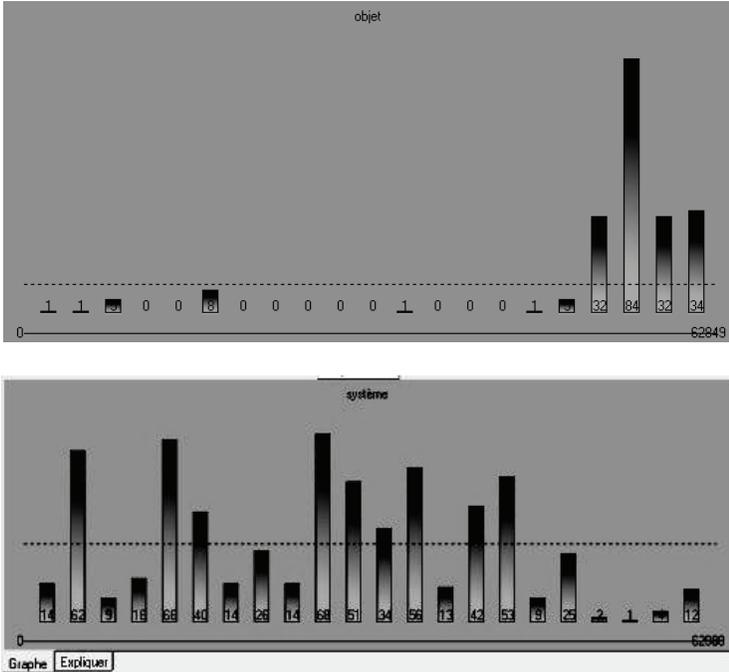
Tableau 8

Traitement complémentaire Tropes	Traitement complémentaire analyse manuelle	Traitement complémentaire Alceste
- répartition chronologique de «système» et «objet»	- relecture «ciblée» tendant à repérer des indications sur l'importance d'«objet» au collège et de «système» au lycée, dans une perspective de progressivité	- objet: 212 occurrences - «système» : 648 (distinct de «system» (comme pour System Modeling Language : 10 occurrences)

La recherche du nombre d'occurrences (avec la même fiabilité pour les deux logiciels) n'est intéressante que si nous retournons sur les fichiers distincts qui ont constitué le fichier de travail présenté ici (il rassemble les documents d'accompagnement de technologie et de STI2D). Le terme « système » est bien le nom le plus fréquent dans le document d'accompagnement relatif aux STI2D, celui d' « objet », le plus fréquent pour le collège. La répartition chronologique de Tropes confirme ce résultat. En ce sens, les logiciels font émerger un résultat très intéressant quant à la question de l'unité des disciplines technologiques, résultat non repérable par l'analyse « manuelle ».

Figure 3

Graphe de répartition de « système » et de « objet » avec Tropes



À l'issue de la présentation de la méthodologie, les notes prises tout au long de la mise en œuvre permettent de revenir sur les questions posées dans cet article.

#### 4. Avantages et limites de l'articulation de trois méthodes d'analyse textuelle

Sans surprise, et déjà notée dans d'autres articles, il existe une grande différence de temps de mise en œuvre entre une analyse utilisant des logiciels et une analyse « manuelle », d'une vingtaine d'heures pour l'analyse « manuelle » à moins d'une heure pour les premières analyses logicielles. De plus, les deux logiciels offrent une pluralité de visualisations possibles des résultats, ce qui n'est pas le cas pour l'analyse « manuelle », qui est d'abord qualitative; une présentation des résultats obtenus en utilisant des graphismes

nécessite un travail supplémentaire. Mais plutôt que de comparer, au risque d'opposer, méthode logicielle et méthode « manuelle », il s'agit d'examiner l'intérêt d'une méthodologie articulée.

#### 4.1. Une méthode articulée fructueuse

Davantage que l'addition ou la juxtaposition des résultats obtenus par chaque méthode, c'est l'utilisation des résultats d'une méthode pour un nouveau traitement par une autre méthode qui enrichit l'analyse.

Il semble que la complémentarité s'appuie d'abord sur les résultats produits par le logiciel Alceste ou par l'analyse « manuelle ». La catégorisation standard proposée par Tropes s'avère moins fructueuse qu'une entrée par les formes repérées dans une classe d'Alceste (pensons au cas de « SysML ») ou par l'analyse « manuelle » (pensons au cas de « norme »). Une fois ces résultats obtenus, Tropes permet de préciser à quel endroit du texte se trouvent les mots d'une classe et si leur présence est répartie dans l'ensemble du texte (la fonction « graphe de répartition » existe aussi sur Alceste, mais les résultats sont plus lisibles avec Tropes). De plus, il est possible, avec Alceste, d'étudier les relations de proximité entre les termes d'une même classe grâce à la visualisation offerte par le dendrogramme obtenu pour une classe donnée (classement complémentaire des formes d'une classe). Par ailleurs, l'analyse « manuelle », qui a abouti à la production d'une liste exhaustive de ressources potentielles, est enrichie par le classement par fréquence et par la répartition chronologique, lesquels permettent de nuancer cette liste.

#### 4.2. Des conditions à respecter

Le retour réflexif sur la mise en œuvre de la méthodologie permet d'en repérer plusieurs conditions. La première concerne la nécessaire étanchéité entre les différentes méthodes; vient ensuite un moment de confrontation entre les résultats obtenus par chacune. Sans cette différenciation des phases, la tentation, voire le risque d'enchaîner des traitements en surnombre n'est pas à négliger. Il s'agit donc de trouver la juste adéquation entre les possibilités

offertes par les logiciels et les résultats recherchés, leur pertinence. En effet, les nombreuses fonctions disponibles, de même que le temps court de traitement incitent à multiplier les opérations, sans valeur ajoutée évidente. Par exemple, avec Alceste, les tests de paramétrages sont à cibler; pour Tropes, de nombreuses fonctions n'ont pas été utilisées (découpage du texte en propositions remarquables ou en sections à exploiter séparément, classement par type de mot).

Une bonne maîtrise du ou des logiciels est également nécessaire, tout autant que l'est la connaissance du domaine dans lequel la recherche se situe, pour permettre une interprétation juste des résultats, l'émission de nouvelles hypothèses de travail ou celle de nouvelles requêtes. Le choix des fonctions à utiliser s'effectue en fonction de l'expérience du chercheur, qui se base sur des analyses antérieurement faites. Bien que la méthode puisse être utilisée par un chercheur seul, l'interactivité apparaît fructueuse pour échanger sur la variation de paramètres et les résultats générés par ces décisions, de même que sur l'interprétation des résultats. La complémentarité des chercheurs a facilité l'étude présentée ici.

Par ailleurs, au-delà de l'investissement en temps pour l'apprentissage d'un produit, investissement non négligeable, le coût constitue un frein à l'exploration d'autres logiciels. Bien souvent, les chercheurs doivent utiliser le matériel déjà existant dans le laboratoire.

## 5. Perspectives

Cette première étude exploratoire renforce le projet d'extension d'analyse à tous les programmes et documents d'accompagnement concernant les Sciences et Techniques Industrielles, voire aux autres champs disciplinaires étudiés dans le projet de recherche ReVEA. Eu égard à la taille de ce nouveau corpus, l'analyse « manuelle », notamment pour repérer la liste des ressources, semble difficile à envisager. En prenant appui sur une comparaison entre les listes obtenues par les différentes analyses (travail non présenté ici), il reste à concevoir une méthodologie

commençant par une analyse textuelle logicielle. À partir d'une liste des substantifs et de leur fréquence, le chercheur, en se basant sur son expertise et en ayant recours au concordancier et aux graphes de répartition, peut obtenir une liste des différentes ressources suggérées par les textes institutionnels, à confirmer par une analyse « manuelle ». Pour la question de la ou des spécificités des disciplines, la production de mondes lexicaux apparaît comme un résultat intéressant à interpréter, interprétation qui peut s'enrichir d'opérations complémentaires.

Dans cet article, nous avons choisi de ne présenter que l'analyse textuelle réalisée. Or, au début de notre texte, nous avons attiré l'attention sur les particularités de notre corpus, qui comprend des données de nature variée. Il reste à expérimenter une méthodologie qui articule une analyse sémiotique et les premiers résultats présentés. D'autres fonctionnalités des logiciels sont à explorer, des méthodologies d'analyse « manuelle » des graphismes et de leurs liens avec le texte sont à élaborer.

Par ailleurs, le logiciel Iramuteq procède au même type d'analyse qu'Alceste mais avec un algorithme modifié, conduisant à des maximisations produisant des résultats quelque peu différents, avec cependant une stabilité des mondes lexicaux. C'est ce que montrent Pierre Ratinaud et Pascal Marchand dans la comparaison des résultats obtenus avec les deux logiciels sur des textes de format particulier, les télégrammes<sup>33</sup>. Comme pour tous les logiciels libres, la formation de l'utilisateur n'est pas prise en charge par les concepteurs, à la différence des logiciels marchands. Mais l'expérience acquise avec Alceste est potentiellement transférable, il reste à tester le fonctionnement d'Iramuteq pour des documents comme les programmes et documents d'accompagnement, ces corpus particuliers mais fréquemment convoqués en didactique.

---

<sup>33</sup> Pierre Ratinaud et Pascal Marchand, *Application de la méthode Alceste aux « gros » corpus et stabilité des « mondes lexicaux » : analyse du CableGate avec Iramuteq*, 2012, <http://lexicomtrica.univ-paris3.fr/jadt/jadt2012/Communications/Ratinaud,%20Pierre%20et%20al.%20-%20Application%20de%20la%20methode%20Alceste.pdf>, site consulté le 31 mars 2015.

## Bibliographie

### Sources générales

Bardin, Laurence, *L'analyse de contenu*, Paris, Presses Universitaires de France, 1977.

<http://anr-revea.fr/>

Raulin, Dominique, *Les programmes scolaires*, Paris, Retz, 2006.

Weil-Barais, Annick, *Les méthodes en psychologie*, Paris, Bréal, 1998.

### Sources concernant le choix des méthodes et des logiciels

Brugidou, Mathieu *et al.*, *Les facteurs de choix et d'utilisation de logiciels d'Analyse de Données Textuelles*, 2000, <http://lexicometrica.univ-paris3.fr/jadt/jadt2000/pdf/04/04.pdf>, site consulté le 31 mars 2015.

Brugidou, Mathieu et Michèle Moine, *Normes émergentes et stigmatisations. Une analyse comparative à partir des deux questions ouvertes sur les raisons de ne pas trier les déchets et de ne pas faire d'économie d'énergie*, 2010, [http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0885-0896\\_147-Brugidou.pdf](http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0885-0896_147-Brugidou.pdf), site consulté le 26 août 2015.

Demazière, Didier, *Des logiciels d'analyse textuelle au service de l'imagination sociologique*, 2005, <http://bms.revues.org/978>, site consulté le 31 mars 2015.

Fallery, Bernard et Florence Rodhain, *Quatre approches pour l'analyse de données textuelles : lexicale, linguistique, cognitive, thématique*, 2007, <http://www.strategie-aims.com/events/conferences/7-xvieme-conference-de-l-aims/communications/2078-quatre-approches-pour-l-analyse-de-donnees-textuelles-lexicale-linguistique-cognitive-thematique/download>, site consulté le 31 mars 2015.

Jenny, Jacques, *Méthodes et pratiques formalisées d'analyse de contenu et de discours dans la recherche sociologique française contemporaine. État des lieux et essai de classification*, 1997, [http://jacquesjenny.com/legs-sociologique/?page\\_id=1253](http://jacquesjenny.com/legs-sociologique/?page_id=1253), site consulté le 8 janvier 2015.

Jenny, Jacques, *À propos des logiciels d'analyse textuelle pratiqués en France pour la recherche en sciences sociales*, 2002, <http://www.cmh.pro.ens.fr/bms/arcati/BMS54-Jenny-New.htm>, site consulté le 31 mars 2015.

Lejeune, Christophe, « Au fil de l'interprétation, l'apport des registres aux logiciels d'analyse qualitative », *Swiss Journal of Sociology*, vol. 3, n° 34, 2008, p. 593-603.

- Lejeune, Christophe, *Méthodes qualitatives informatisées*, <http://www.squash.ulg.ac.be/logiciels>, site consulté le 23 septembre 2015.
- Lejeune, Christophe, *Montrer, calculer, explorer, analyser. Ce que l'informatique fait (faire) à l'analyse qualitative*, 2010, <http://hdl.handle.net/2268/61098>, site consulté le 31 mars 2015.
- Peyrat-Guillard, Dominique et Daniel Dufresne, *L'analyse exploratoire de référentiels de compétences avec Alceste : une aide à la lecture de l'analyste*, 2010, [http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0141-0150\\_019-Peyrat.pdf](http://lexicometrica.univ-paris3.fr/jadt/jadt2010/allegati/JADT-2010-0141-0150_019-Peyrat.pdf), site consulté le 26 août 2015.
- Venant, Fabienne et Jean-François Maheux, *Réforme de l'enseignement au Québec : une visite guidée par la textométrie*, 2014, <http://lexicometrica.univ-paris3.fr/jadt/jadt2014/01-ACTES/54-JADT2014.pdf>, site consulté le 26 août 2015.
- Wanlin, Philippe, *L'analyse de contenu comme méthode d'analyse qualitative d'entretiens : une comparaison entre les traitements manuels et l'utilisation de logiciels*, 2007, [http://www.recherche-qualitative.qc.ca/documents/files/revue/hors\\_serie/hors\\_serie\\_v3/Wanlin2.pdf](http://www.recherche-qualitative.qc.ca/documents/files/revue/hors_serie/hors_serie_v3/Wanlin2.pdf), site consulté le 28 mars 2015.

## Sources concernant le logiciel Alceste

- Dalud-Vincent, Monique, « Les “choix” du sociologue avec Alceste. De la forme du corpus aux résultats obtenus », *Bulletin de méthodologie sociologique*, n° 105, 2010, p. 25-52.
- Dalud-Vincent, Monique, « Les “choix” du sociologue avec Alceste. Du paramétrage des unités de contexte aux résultats obtenus », *Bulletin de méthodologie sociologique*, n° 107, 2010, p. 23-48.
- Kalampalakis, Nikos, « L'apport de la méthode Alceste dans l'étude des représentations sociales » dans Jean-Claude Abric (dir.), *Méthodes d'étude des représentations sociales*, Paris, Éditions Érès, 2003, p. 147-163.
- Ratineau, Philippe et Pascal Marchand, *Application de la méthode Alceste aux « gros » corpus et stabilité des « mondes lexicaux » : analyse du CableGate avec Iramutec*, 2012, <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Ratinaud,%20Pierre%20et%20al.%20-%20Application%20de%20la%20methode%20Alceste.pdf>, site consulté le 31 mars 2015.
- Reinert, Max, *Alceste, un outil d'aide pour l'analyse du discours. Notice simplifiée (de la version de base commune aux versions 4.x)*, <http://www.ling.uqam.ca/forum/satoman/images/AlcesteNoticeSimplifiee.pdf>, site consulté le 24 juillet 2015.
- Reinert, Max *et al.*, *Alceste, logiciel de données textuelles (non daté)*, <http://www.image-zafar.com/Logiciel.html>, site consulté le 22 juillet 2015.

## Sources concernant le logiciel Tropes

- Molette, Pierre, *De l'APD à Tropes : comment un outil d'analyse de contenu peut évoluer en logiciel de classification sémantique généraliste*, 2009, <http://www.tropes.fr/PierreMolette-CommunicationColloquePsychoTarbesJuin2009.pdf>, site consulté le 24 juillet 2015.
- Tropes, version 8.0, manuel de référence*, 2011, <http://www.tropes.fr/ManuelDeTropesVF800a.pdf>, site consulté le 20 février 2015.

## Textes officiels

- « Charte des programmes », *Bulletin officiel de l'Éducation nationale des lois et règlements n° 8 du 20 février 1992*, Ministère de l'Éducation nationale, 1992, <http://artheque.ens-cachan.fr/items/show/5647>, site consulté le 31 mars 2015.
- Charte des programmes. Charte relative à l'élaboration, à la mise en œuvre et au suivi des programmes d'enseignement ainsi qu'aux modalités d'évaluation des élèves dans l'enseignement scolaire*, Conseil supérieur des programmes, Ministère de l'Éducation nationale, de l'enseignement et de la recherche, 2013, [http://cache.media.education.gouv.fr/file/04\\_Avril/37/5/charte\\_programme\\_csp\\_312375.pdf](http://cache.media.education.gouv.fr/file/04_Avril/37/5/charte_programme_csp_312375.pdf), consulté le 8 juillet 2015.
- Enseignements technologiques (transversaux et spécifiques des spécialités architecture et construction, énergies et environnement, innovation technologique et éco-conception, systèmes d'information et numérique) du cycle terminal de la série STI2D*, Ministère de l'Éducation nationale, de l'enseignement et de la recherche, 2011, <http://www.education.gouv.fr/cid55415/mene1104262a.html>, site consulté le 31 mars 2015.
- « Programmes du collège. Programmes de l'enseignement de technologie », *Bulletin officiel spécial n° 6 du 28 août 2008*, Ministère de l'Éducation nationale, 2008, [http://media.education.gouv.fr/file/special\\_6/53/1/Programme\\_technologie\\_33531.pdf](http://media.education.gouv.fr/file/special_6/53/1/Programme_technologie_33531.pdf), site consulté le 31 mars 2015.
- Ressources pour le cycle terminal. Enseignements technologiques transversaux et enseignements spécifiques (série STI2D)*, Éduscol, Ministère de l'Éducation nationale (DGESCO), 2011, [http://media.eduscol.education.fr/file/STI2D/15/2/LyceesGT\\_Ressources\\_STI2D\\_T\\_Enseignement\\_Technologique\\_Specifiques\\_182152.pdf](http://media.eduscol.education.fr/file/STI2D/15/2/LyceesGT_Ressources_STI2D_T_Enseignement_Technologique_Specifiques_182152.pdf), site consulté le 31 mars 2015.
- Technologie au collège. Ressources pour faire la classe en 6<sup>ème</sup>, 5<sup>ème</sup>, 4<sup>ème</sup> et 3<sup>ème</sup>*, Éduscol, Ministère de l'Éducation nationale (DGESCO), 2009, [http://cache.media.eduscol.education.fr/file/Programmes/16/6/College\\_Technologie\\_Ressources\\_114166.pdf](http://cache.media.eduscol.education.fr/file/Programmes/16/6/College_Technologie_Ressources_114166.pdf), site consulté le 31 mars 2015.