

Classification automatique et stratégie d'annotation appliquées à un concept philosophique: la dimension psychologique du concept de LANGAGE dans l'œuvre de Bergson ¹

Jean Danis ¹, Jean-Guy Meunier ¹, Jean-François Chartier ¹, Motasem Alrahabi ², Jean-Pierre Desclés ²

¹ LANCI- Laboratoire d'Analyse Cognitive de l'Information-UQAM-Montréal-Canada

² Laboratoire LaLIC – Université Paris-Sorbonne

Résumé

Les outils d'assistance à la lecture et à l'analyse experte des textes sont de plus en plus demandés dans le domaine des humanités. En philosophie, ces processus de lecture et d'analyse s'effectuent dans bien des cas par l'entremise d'analyses conceptuelles. Les approches lexicométriques et macrotextuelles offrent des pistes intéressantes mais demeurent limitées lorsqu'il s'agit d'analyser les multiples facettes d'un concept philosophique. Nous présentons dans cet article une méthode de fouille textuelle qui combine à l'analyse d'une concordance un algorithme de classification non supervisée (k-moyennes) et une stratégie d'annotation assistée par la plate-forme sémantique EXCOM. Le traitement est appliqué au concept de LANGAGE dans le corpus d'Henri Bergson.

Abstract

Assisting tools for analysis and expert reading of texts are more and more demanded by researchers in the Humanities. In philosophy, these expert reading processes are often carried out by conceptual analysis. Macro-textual approaches have given appealing results but remain limited when it's time to analyze the multifaceted properties of philosophical concepts. We present in this article a computer-assisted text mining method that allows researchers to explore the multifaceted contexts of a philosophical concept. The method aims at integrating to standard clustering methods a computer-assisted annotation strategy with the use of the platform called "EXCOM". Our method is here applied to the concept of LANGUAGE in Henri Bergson's corpus.

Keywords: clustering method, semantic annotation, concordance, philosophical concept, excom platform, textmining

1. Introduction

Dans bien des cas, les chercheurs des humanités demeurent relativement insatisfaits des principaux outils d'analyse textuelle développés jusqu'à ce jour. On demande une assistance à l'interprétation fidèle à des pratiques disciplinaires ciblées et classiques (Rockwell, 2003;

¹ Ce travail a été réalisé dans le cadre de subventions du Conseil de recherches en sciences humaines du Canada (CRSH) et du Fonds québécois de la recherche sur la société et la culture (FQRSC) en collaboration avec plusieurs chercheurs du LANCI (Nicolas Payette) et du Laboratoire LaLIC (Dory Diebakhate).

Unsworth, 2005 ; Bradley, 2007). Tel est le cas de l'analyse conceptuelle pratiquée en philosophie. Effectuée pour confirmer des hypothèses ou pour mieux saisir un système de pensée, l'analyse conceptuelle s'effectue par une fouille approfondie des principaux concepts qui participent à la formation des doctrines. Peu d'outils ont jusqu'à maintenant été conçus pour assister ce type de pratique.

Plusieurs tenants de la sémantique textuelle ont développé des méthodologies d'assistance à l'analyse de l'évolution des concepts théoriques au sein des discours (Valette, 2003; Loiseau, 2003 ; Poudat, 2006). Bien qu'heuristiques d'un point de vue diachronique, ces approches à caractère linguistique demeurent dans une large mesure axées sur les composantes sémantico-sémiques (isotopiques) et lexicales d'une langue donnée. Lorsqu'il s'agit d'aborder les concepts dans une optique philosophique, l'analyse doit déborder de la dimension linguistico-structurale. Les propriétés des concepts philosophiques se présentent souvent avec une terminologie à « multiples entrées » (sémantique, métaphorique, logique, rhétorique...), (Dixsaut, 1985). Dans une perspective logico-cognitiviste, le contenu conceptuel découle également de plusieurs opérations cognitives telles le jugement, la schématisation, la conscience, les inférences, l'attitude propositionnelle, etc. (Rey, 1983; Brandom, 1994). Dans cette perspective, le contenu conceptuel relève de sources et d'opérations fortement liées au contexte qui, tantôt sont de nature linguistique (péri et paratexte), mais qui relèvent dans bien des cas des productions scientifiques et culturelles reliées au texte étudié.

L'analyse des propriétés génériques du contenu conceptuel exige d'être sensible à ces diverses sources et opérations qui, le plus souvent, ne sont décelables que par une exploration approfondie de la dimension micro-textuelle. Cette exploration implique des stratégies de catégorisation fine qui font constamment interagir l'analyse de la matérialité textuelle et l'analyse de composantes extralinguistiques (cognitive, discursive, doctrinale, etc.).

Nous présentons dans cet article une application d'une méthode de fouille textuelle qui vise à assister informatiquement la lecture experte réalisée en complémentarité aux approches classificatoires. La méthodologie permet d'assister le processus d'analyse conceptuelle en combinant à l'opération standard de classification de segments une annotation catégorielle automatique. Réalisée à l'aide de la plate-forme d'annotation sémantique EXCOM, cette méthodologie s'inscrit dans les recherches sur le développement de l'annotation automatique de catégories utilisées à des fins d'analyse experte. Nous présentons en premier lieu les fondements théoriques qui ont guidé notre méthodologie. Ceci sera suivi de l'application concrète de la stratégie expérimentée avec le concept de LANGAGE dans l'œuvre d'Henri Bergson.

2. Fondements méthodologiques

2.1. *Production des contextes d'inférences d'un concept philosophique: fondements d'une classification appliquée à la concordance*

Nous définissons le processus d'analyse d'un concept spécifique comme *une méthode systématique de parcours textuel qu'un spécialiste met en œuvre afin d'identifier, dans le texte, les manifestations explicites des types d'opérations cognitives catégorisantes (sémantique et logique) exprimées dans une forme linguistique canonique*. Techniquement, ce processus d'analyse vise à explorer en profondeur l'organisation interne de termes poly-occurents présents autour d'un mot-pôle donné. Plusieurs ont développé des approches offrant une lecture orientée vers les co-occurents de termes spécifiques (Martinez, 2003 ; Mayaffre, 2008 ; Brunet, 2008).

La méthode d'analyse conceptuelle des textes assistée par ordinateur (LACTAO), (Chartier et al., 2008; Meunier and Forest, 2009), emprunte une voie similaire, cette fois-ci en intégrant des méthodes courantes de classification textuelle (*text mining*) à l'analyse d'une concordance. Dans la présente expérience, l'algorithme de classification de type non hiérarchique « K-moyennes » est appliqué spécifiquement à l'ensemble des segments d'une concordance. La procédure permet de produire des classes de segments qui forment en eux-mêmes des sous-corpus. L'approche classificatoire est ainsi essentiellement vectorielle ². En regroupant des segments textuels d'une concordance selon des critères de similarité lexicale, l'opération produit des classes d'équivalence (Jain and Flynn, 1999 : 265).

Ces classes doivent cependant être soumises à une analyse plus sophistiquée. La simple classification, la description statistique et la représentation graphique des classes ne constituent pour l'analyste en philosophie que des étapes préliminaires ³. La lecture experte exige d'appliquer à ces classes des analyses de deuxième niveau. L'hypothèse qui nous guide dans ce deuxième niveau d'analyse est que les classes de segments issues d'une classification correspondent à diverses dimensions d'un concept, c'est-à-dire divers vecteurs de sens ou divers « champs signifiants ». D'un point de vue théorique, ces dimensions relèvent des *contextes d'inférences* associés au concept, à savoir les divers réseaux de propositions qui permettent d'établir ce que l'on infère à partir d'un concept donné ⁴.

2.2. Analyse d'une classe de contextes d'inférences d'un concept philosophique: fondements d'une annotation assistée informatiquement

L'analyse en profondeur du contenu des documents issus de la classification est généralement d'une ampleur considérable. Comme plusieurs l'ont souligné (Rockwell, 2003 ; Bradley, 2007), ceci constitue le cœur de ce qui est appelé l'interprétation du concept. On peut comprendre ce processus comme un ajout de *couches* peritextuelles et interprétatives aux segments de textes.

Un des enjeux de cette recherche est d'assister informatiquement ce travail d'ajout de couches que l'on considère ici comme étant diverses formes d'annotation de nature multicatégorielle (rhétorique, logique, discursive, linguistique, éditoriale, logique, etc.). Plusieurs plates-formes et logiciels ont été récemment développés afin de réaliser des multi-annotations de nature linguistique mais peu ont été jusqu'à maintenant réalisés pour assister une annotation tournée vers les aspects d'ordre sémantique, rhétorique ou logique.

Nous explorons dans cette recherche une assistance à l'annotation à l'aide de la plateforme sémantique « EXCOM » développée au LALIC (Djioua et al., 2006 ; Alrahabi and Desclés, 2009). Basée sur la méthode de l'« Exploration Contextuelle », EXCOM permet de procéder à des annotations automatiques de segments par l'analyse de marqueurs textuels dans leur contexte.

² La classification regroupe ici des segments de textes (vecteurs) en fonction de leur similarité. Ces vecteurs ont tous pour composante un mot pôle spécifique. Les textes sont donc regroupés en fonction des structures de termes similaires autour d'une unité lexicale qui peut être la lexicalisation d'un concept.

³ Aucun lecteur expert sérieux en philosophie ne se contentera de la production de classes de segments textuels. Tout arrêt à cette étape créera chez ce type de spécialistes une résistance à utiliser l'informatique comme assistance à l'analyse.

⁴ À titre d'exemple, dans le cas du concept d'évolution chez Darwin, le contexte d'inférences exprime ce que l'on peut inférer du concept : sa relation aux concepts de VARIATIONS LENTES, de DÉMOGRAPHIE, d'ENVIRONNEMENT, d'ESPÈCES, d'INDÉTERMINISME, etc. De ce contexte inférentiel se dégage différents types de raisonnements associés à une pratique scientifique déterminée; abduction, induction, comparaison, etc.

L'annotation s'effectue selon différentes relations sémantico-discursives généralement utilisées dans une démarche de *fouille textuelle* ⁵.

La stratégie d'annotation utilise dans cette recherche le système EXCOM et la carte sémantique « causalité » (Garcia, 1998) afin de repérer et de regrouper au sein d'une classe les différentes expressions de causalité. Comme nous le verrons, les expressions de causalité identifiées permettent de mettre en lumière les liens d'influences entre les entités et phénomènes présents au sein des classes soumises à l'analyse.

La méthode d'analyse présentée ici (LACTAO) se déploie en deux grands moments: l'étape classificatoire, puis l'étape d'annotation. Nous présentons quelques éléments de l'étape classificatoire telle qu'appliquée au concept de langage chez Bergson. Nous présenterons par la suite de façon plus détaillée l'annotation d'un des sous-corpus issus de cette classification (classe 18, dimension « psychologique » du concept de « langage »).

3. La classification appliquée à la concordance « langage » dans le corpus de Bergson

Après avoir choisi un terme spécifique qui devient l'expression conceptuelle de départ de l'analyse (dans ce cas-ci, la lexie « langage »), une concordance est produite afin d'extraire tous les contextes de ce terme. L'opération classificatoire est par la suite appliquée à la concordance afin de produire les différentes classes de segments.

L'extraction des contextes de « langage » dans l'ensemble du corpus de Bergson ⁶ donne dans cette expérience particulière un sous-corpus (concordance) de 208 segments (contextes) formés de 5 phrases ⁷. L'application de l'algorithme « K-moyennes » a produit 20 classes contenant en moyenne 10 segments.

L'algorithme *K-moyennes* a la particularité de produire pour chaque classe un centre final (centroïde), celui-ci étant le vecteur prototype de la classe. À partir de ces vecteurs, une représentation graphique (figure 1) est produite pour illustrer les différents liens de proximité entre les classes de segments qui ont des vecteurs prototypes « similaires » ⁸. Les liens permettent de situer les classes de segments qui ont des vecteurs *proches* et ainsi de les relier en fonction de leurs régularités lexicales communes. Bien qu'il s'agisse d'une approche mathématique différente, la technique permet à l'instar d'une classification hiérarchique ascendante de dégager un premier niveau supérieur de regroupement de classes. Nous avons illustré à l'aide de trois cercles des regroupements de classes établis par ces liens de proximité (Fig. 1) ⁹. La

⁵ Par exemple, des relations identifiant la causalité : Quelle est la cause de tel fait ?, ou encore l'influence : Quel est le rôle ou la contribution de tel fait ? Qui a dit quoi ?, etc.).

⁶ La somme de ces ouvrages constitue l'œuvre philosophique de Bergson dans sa quasi totalité, soit les écrits allant de 1888 à 1932.

⁷ Le sous-corpus produit par la concordance contient 29 804 mots dont 2 652 uniques (types).

⁸ La proximité est ici traitée par la distance entre vecteurs prototypes des classes. Celle-ci est inversement proportionnelle à l'attraction (proximité) entre deux vecteurs prototypes. Cette distance est calculée par le « Normalized Dot Product ». Le calcul effectué à partir de la mesure de similarité cosinus se traduit formellement par : $1 - \frac{A \cdot B}{\|A\| \cdot \|B\|}$; ceci correspond à exprimer des distances par des normes comprises entre 0 et 1. Nous n'avons retenus ici que les plus petites distances entre classes (celles en deçà du seuil 0.70).

⁹ Le premier pôle (PSYCHOLOGIE et SENTIMENTS) illustre le regroupement des classes 05,17,02,12 et 01 qui ont toutes pour vecteur prototype le plus proche la classe 18. Les lemmes représentatifs de ces classes sont « âme », « sentiment », « amour », etc. Le deuxième pôle (ÉPISTÉMOLOGIE, SCIENCE et PHILOSOPHIE) illustre un regroupement entre les classes 09, 04,11, et 16, etc. Les lemmes représentatifs de ces classes sont

représentation permet ainsi d'instancier des parcours de lecture axés sur des regroupements de classes qui ont des vecteurs prototypes similaires à savoir, des régularités lexicales communes.

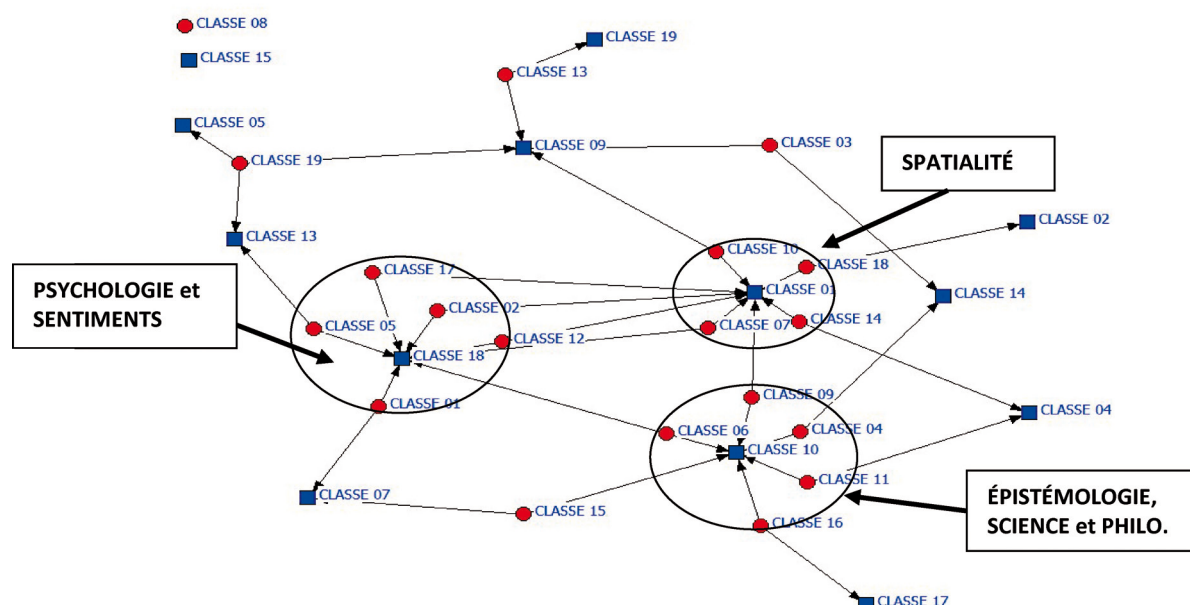


Figure 1 : Réseaux entre classes de segments contenant le mot-pôle « langage »; distances entre centroïdes de classes

Plusieurs stratégies existent pour décrire le contenu signifiant de ces réseaux et des classes qui en font partie. La recherche de mots ou de phrases caractéristiques par le calcul des spécificités (Lebart and Salem, 1994) constitue une des méthodes largement utilisées. Dans cette recherche, la stratégie choisie consiste plutôt à thématiser, c'est-à-dire à exprimer par un ou plusieurs termes ce contenu en ayant recours à l'extraction des lemmes qui ont la valeur *tf-idf* la plus élevée (*term frequency * inverse document frequency*), (Salton, 1989)¹⁰. L'interprétation synthétique non automatisée des lemmes représentatifs¹¹ de ces réseaux permet ainsi de dégager les principaux champs thématiques de la classification. Il apparaît alors que le concept de « langage » se déploie au sein de différentes thématiques spécifiques telles : « ÉPISTÉMOLOGIE, SCIENCE et PHILOSOPHIE », « PRAGMATISME », « SEMIOTIQUE », « TEMPORALITÉ », « SPATIALITÉ », « SENTIMENTS », etc.¹²

« philosophie », « acte », « connaissance », « action », « idée », etc.

¹⁰ Notons que l'indice *tf-idf* (*term frequency * inverse document frequency*) (Salton, 1989) est ici calculé à partir de la concordance, à savoir le sous-corpus extrait du corpus original. L'indice permet de retenir l'ensemble des termes d'une classe qui sont les plus représentatifs de celle-ci par rapport au contenu des autres classes. Le principe est ici le suivant : un mot du lexique d'une classe sera d'autant meilleur pour caractériser le contenu signifiant et le contexte d'inférence du concept étudié s'il est très fréquent dans cette classe et rare dans les autres classes (Forest and Meunier, 2004; Forest, 2006).

¹¹ Les unités les plus représentatives sont ici retenues de la façon suivante : parmi les termes de chaque pôle lexico-thématique (réseau de classes), nous retenons les 8 lemmes ayant la valeur *tf-idf* la plus élevée.

¹² Dans la classe 10 et les classes qui lui sont reliées, le concept de « langage » est associé aux dimensions « ÉPISTÉMOLOGIE », « SCIENCE » et « PHILOSOPHIE », le concept s'y déploie par des segments ayant pour lemmes spécifiques : « philosophie », « action », « idée », « forme », « vue », « représentation », « connaissance », etc. La classe 18 forme quant à elle un autre pôle lexico-thématique, cette fois-ci marqué par les dimensions

Ce type d'analyse de la dimension conceptuelle est intéressant. Le lecteur expert exigera cependant une interprétation encore plus fine. Ici intervient notre stratégie d'annotation *assistée* informatiquement. Nous présentons ce type d'assistance réalisée à l'aide de la carte « causalité » de la plate-forme EXCOM. La stratégie d'annotation est appliquée dans cette recherche à une classe spécifique : la classe 18 (langage, dimension « PSYCHOLOGIE » et « SENTIMENTS »).

4. Dimensions « psychologie et sentiments » du concept de **LANGAGE** chez Bergson : analyse et annotation détaillées des segments de la classe 18

4.1. Carte sémantique de la « causalité » (Garcia, 1998)

Inspirée du modèle cognitif de la Dynamique des forces de Leonard Talmy (1998), la carte sémantique CAUSALITÉ (Garcia, 1998) est une forme d'« ontologie » linguistique qui permet d'identifier 25 types de relations causales spécifiques (*créer, empêcher, faciliter, pousser-à, etc.*). Ces relations, dont la signification est décrite par des « schèmes » sémantico-cognitifs (Desclés, 1990), expriment de façon générique l'influence « causale » d'une situation sur une autre. Outre le fait d'aider à la structuration de la terminologie de la classe, les relations causales identifiées par l'annotation sont utilisées pour cerner comment l'auteur perçoit et structure les différents phénomènes et entités qui forment le contenu conceptuel de la classe.

4.2. Annotation des relations causales : résultats de l'annotation de la classe 18 (« langage », PSYCHOLOGIE et SENTIMENTS)

Dans la classe 18, le concept de « langage » se déploie dans un réseau de propositions ayant pour unités représentatives les lemmes « état », « sentiment », « mot », « idée », « extérieur », « haine », « homme » et « juxtaposer ». Nous présentons à titre d'exemple un des segments caractéristiques de son contenu. Les unités représentatives selon la valeur *tf-idf* sont en caractère gras.

Cependant le langage désigne ces **états** par les mêmes **mots** chez tous les **hommes** ; aussi n'a-t-il pu fixer que l'aspect objectif et impersonnel de l'amour, de la **haine**, et des mille **sentiments** qui agitent l'**âme**. Nous jugeons du talent d'un romancier à la puissance avec laquelle il tire du domaine public, où le **LANGAGE** les avait ainsi fait descendre, des **sentiments** et des **idées** auxquels il essaie de rendre, par une multiplicité de détails qui se **juxtaposent**, leur primitive et vivante individualité. Mais de même qu'on pourra intercaler indéfiniment des points entre deux positions d'un mobile sans jamais combler l'espace parcouru, ainsi, par cela seul que nous parlons, par cela seul que nous associons des **idées** les unes aux autres et que ces **idées** se **juxtaposent** au lieu de se pénétrer, nous échouons à traduire entièrement ce que notre **âme** ressent : la pensée demeure incommensurable avec le langage. Seg. 17, Bergson, DI1888.

Le traitement d'EXCOM s'effectue sur les différents segments de la classe 18 en identifiant premièrement les unités linguistiques destinées à être annotées. La plate-forme segmente préalablement les différents contextes (segments) de la classe en « espaces de recherche » par l'entremise des signes de ponctuations fortes (point des phrases, point d'interrogation, etc.). Les unités linguistiques annotées dans la classe 18 sont essentiellement des phrases. La plate-forme a au total identifié et catégorisé 21 phrases du contenu textuel de la classe (section 4.2.1). Ces segments ont reçu via EXCOM différentes *étiquettes* de catégories « CAUSALITÉ » du fait qu'ils présentent des entités linguistiques et des relations quiinstancient les règles prédéfinies

« SENTIMENTS » et « PSYCHOLOGIE » ; le concept s'y déploie par des segments contenant les lemmes « amour », « sentiment », « impression », « état », « moi »... Le concept de « langage » se déploie dans la classe 01 et les classes qui lui sont reliées par la thématique de la SPATIALITÉ, etc.

au sein de la carte de la plate-forme. Nous avons surligné dans les segments certains des indices observables utilisés par EXCOM dans les processus d'identification et de catégorisation des différentes relations causales¹³. L'opération d'annotation n'a ici identifié que deux types de relations, soit « l'influence » et la « modalité ». La première forme se rapporte essentiellement à l'influence (*l'effet*) d'une situation sur une autre, la deuxième porte sur la partie du déroulement d'une situation en cours (*tenter de, achever de, entamer...*). Les segments annotés ont été manuellement regroupés en fonction de leurs différents types d'annotations et des domaines auxquels ils se rapportent. Nous présentons en premier lieu les annotations qui se rapportent aux processus psychiques, ceci sera suivi de celles qui relèvent d'une théorie de la connaissance. L'espace de cet article ne nous permet de présenter qu'un échantillon des annotations produites par la plate-forme.

4.2.1. Descriptions d'une causalité générale entre phénomènes psychiques¹⁴

4.2.1.1. Segments exprimant la suspension (l'arrêt) d'un processus psychique en cours

CAUSALITÉ\INFLUENCE\ arrêter d'entretenir\.

Mais à mesure que l'on creuse au-dessous de cette surface, à mesure que le moi redevient lui-même, à mesure aussi ses états de conscience cessent de se juxtaposer pour se pénétrer, se fondre ensemble, et se teindre chacun de la coloration de tous les autres. Seg. 15. BERGSON 1888.

CAUSALITÉ\INFLUENCE\ arrêter d'entretenir\.

*Et, à vrai dire, pour la commodité du **LANGAGE**, il [le moi psychique] a tout intérêt à ne pas rétablir la confusion là où règne l'ordre, et à ne point troubler cet ingénieux arrangement d'états en quelque sorte impersonnels par lequel il a cessé de former « un empire dans un empire ».* Seg. 12. BERGSON 1888.

CAUSALITÉ\INFLUENCE\ bloquer \.

*Cependant le **LANGAGE** désigne ces états [de conscience individuelle profonde] par les mêmes **mots** chez tous les hommes ; aussi n'a-t-il pu fixer que l'aspect objectif et impersonnel de l'amour, de la haine, et des mille sentiments qui agitent l'âme.* Seg. 15. BERGSON 1888.

CAUSALITÉ\INFLUENCE \ bloquer\.

De même que la durée fuyante de notre moi se fixe par sa projection dans l'espace homogène, ainsi nos impressions sans cesse changeantes, s'enroulant autour de l'objet extérieur qui en est cause, en adoptent les contours précis et l'immobilité. Seg. 7. BERGSON 1888.

4.2.1.2. Segments exprimant la création de nouvelles situations

CAUSALITÉ\INFLUENCE\ créer\.

*Et pour nous induire à tenter le même effort sur nous-mêmes, ils [les créateurs] s'ingénieront à nous faire voir quelque chose de ce qu'ils auront vu : par des arrangements rythmés de mots, qui arrivent ainsi à s'organiser ensemble et à s'animer d'une vie originale, ils nous disent, ou plutôt ils nous suggèrent, des choses que le **LANGAGE** n'était pas fait pour exprimer.* Seg. 207. BERGSON 1900.

CAUSALITÉ\INFLUENCE\ créer\.

Ainsi se forme un second moi qui recouvre le premier, un moi dont l'existence a des moments distincts, dont les états se détachent les uns des autres et s'expriment, sans peine par des mots. Seg. 12. BERGSON 1888.

CAUSALITÉ\INFLUENCE\ créer\.

¹³ Les unités surlignées ne représentent ici qu'une partie des indices utilisés par la plate-forme pour catégoriser les segments. EXCOM se base en fait sur des règles plus complexes (présence, absence, position et/ou ordre des indices dans le contexte). L'espace de cet article ne nous permet pas de présenter la totalité de ces règles.

¹⁴ Les propositions surlignées correspondent aux principaux indices retenus par EXCOM pour caractériser le type de relation causale exprimée dans le segment. Par exemple, l'expression « **se forme un second moi** » est pour EXCOM une expression linguistique de la relation de causalité créative (créer), c'est-à-dire « il se forme » (causalement) : il se crée.

Le mot, qui ne note de la chose que sa fonction la plus commune et son aspect banal, s'insinue entre elle et nous, et en masquerait la forme à nos yeux si cette forme ne se dissimulait déjà derrière les besoins qui ont créé le mot lui-même. BERGSON 1900.

CAUSALITÉ\INFLUENCE\ créer\.

*Au lieu d'une vie intérieure dont les phases successives, chacune unique en son genre, sont incommensurables avec le **LANGAGE**, nous obtiendrons un moi recomposable artificiellement, et des états psychiques simples qui s'agrègent et se désagrègent comme font, pour former des mots, les lettres de l'alphabet.* BERGSON 1888.

4.2.1.3. Segments portant sur la modalité d'une action en cours (tendance à la réalisation d'un processus)

CAUSALITÉ\MODALITE\ tenter de\.

*Nous tendons instinctivement à solidifier nos impressions, pour les exprimer par le **LANGAGE**.* BERGSON 1888.

CAUSALITÉ\MODALITE\ tenter de\.

*Nous jugeons du talent d'un romancier à la puissance avec laquelle il tire du domaine public, où le **LANGAGE** les avait ainsi fait descendre, des sentiments et des idées auxquels il essaie de rendre, par une multiplicité de détails qui se juxtaposent, leur primitive et vivante individualité.* BERGSON 1888.

4.2.2. Description causale, théorie de la connaissance et influence du langage

CAUSALITÉ\INFLUENCE\ influencer\.

*Dès lors, nous nous rapprocherons plus encore de la réalité en convenant de désigner par les signes invariables **X** et **Y**, (...) les deux directions différentes que notre imagination leur assigne pour la plus grande commodité du **LANGAGE**.* BERGSON 1888.

CAUSALITE\INFLUENCE\se-soumettre-à\.

*Mais de même qu'on pourra intercaler indéfiniment des points entre deux positions d'un mobile sans jamais combler l'espace parcouru, ainsi, par cela seul que nous parlons, par cela seul que nous associons des idées les unes aux autres et que ces idées se juxtaposent au lieu de se pénétrer, nous échouons à traduire entièrement ce que notre âme ressent : la pensée demeure incommensurable avec le **LANGAGE**.* BERGSON 1888.

4.3. Interprétation des résultats

4.3.1. La classe 18

Le sous-corpus produit à partir des segments annotés guide le lecteur expert dans un parcours interprétatif. Dans l'exemple présenté ici, le parcours consiste à « compiler » de manière synthétique les divers segments exprimant les relations d'influences entre les entités (concepts) au sein d'un champ donné. Les segments de ce sous-corpus, une fois regroupés, expriment la *récurrence* de processus semblables entre des situations qui impliquent les principaux concepts de la classe. Le parcours permet ainsi de cerner la façon particulière dont Bergson *organise* un champ conceptuel autour du concept de langage.

Ainsi, dans l'exemple donné, on voit comment le langage interagit avec les processus psychiques (« moi », « impressions »), les sentiments (« amour », « haine », « sentiments »...) et le vécu conscientiel en général (« âme », « conscience »).

La psyché est dans la classe 18 caractérisée par un continuum indivisible, un processus dynamique formé d'« états de conscience », « fondus ensembles », « fuyants » et « sans cesse variables ». Les différentes « phases successives » de la conscience représentent des processus continus qui sont interrompus par le langage ou différents processus psychiques. Les segments annotés par une relation causale d'influence tournée vers l'ARRÊT et le BLOQUAGE d'un processus révèlent que *par* ou *pour* le langage :

- *notre moi se fixe (...)*
- *[le moi] a tout intérêt à ne pas rétablir la confusion là où règne l'ordre (...)*
- *à mesure qu'il [le moi] redevient lui-même, à mesure aussi ses états de conscience cessent de se juxtaposer pour se pénétrer, se fondre ensemble*
- *[le moi] tend instinctivement à solidifier les impressions (...).*

Le langage quant à lui :

- *(...) ne peut fixer que l'aspect objectif et impersonnel de l'amour, de la haine, et des mille sentiments qui agitent l'âme*
- *[fait que] nous échouons à traduire entièrement ce que notre âme ressent.*

Par l'entremise du langage, différents processus psychiques vont aussi engendrer de nouveaux états, de nouvelles situations contrastant avec des stades antérieurs. Les segments annotés par une relation causale d'influence axée sur la *CRÉATION* de nouvelles situations révèlent que le langage, par l'entremise des mots, engendre :

- Un moi recomposable artificiellement, et des états psychiques simples qui s'agrègent et se désagrègent (...)
- Un moi dont l'existence a des moments distincts, dont les états se détachent les uns des autres et s'expriment, sans peine par des mots.

On voit comment le contexte d'inférences de la classe marqué par les concepts « sentiment », « état », « âme », « haine », « juxtaposer », ... se voit davantage explicité par un parcours de lecture orienté vers différentes relations entre les entités. Rappelons que l'annotation des relations causales ne cible qu'une des dimensions spécifiques exprimées par le contexte d'inférences de la classe. Une interprétation plus contrôlée sollicitant l'expertise du chercheur demeure nécessaire afin d'explorer des dimensions qui auraient pu être occultées par l'opération d'annotation ou l'analyse lexicale. La tendance générale dégagée par l'annotation sur la causalité révèle que le langage pour Bergson relève davantage du « statique », du « limitatif » et par le fait même du « contraignant »¹⁵. Une lecture et analyse approfondie des segments de la classe indique que cette interprétation doit néanmoins être nuancée. Le caractère contraignant et limitatif du langage est en effet ici nuancé par des passages comme : « Il est présumable que, sans le langage, [l'intelligence] aurait été rivée aux objets matériels qu'elle avait intérêt à considérer », ou encore : « Le LANGAGE a beaucoup contribué à la [l'intelligence] libérer. Le mot, fait pour aller d'une chose à une autre, est, en effet essentiellement déplaçable et libre. »

4.3.2. Ouverture : vers une annotation multicatégorielle

Le sous-corpus¹⁶ produit ici par le regroupement des segments annotés ne représente qu'une dimension *annotative* parmi tant d'autres. D'autres parcours de lecture et d'annotation peuvent être instanciés au sein de la même classe par l'utilisation des autres cartes sémantiques d'EXCOM (rhétorique, hypothèses, expérimentation...). Ces autres parcours permettent de combiner pour un même segment textuel différentes catégories d'annotation et offrent ainsi la possibilité d'effectuer plusieurs niveaux de description du contenu textuel.

¹⁵ Le concept de langage est dans la classe 18 associé aux qualificatifs « grossier », « banal », « impersonnel », « simples », etc.

¹⁶ Ces annotations forment un sous-corpus qui, rappelons-le, est le sous-corpus d'un autre sous-corpus (la concordance).

5. Conclusion

Étant fortement liés au contexte, les différents processus discursifs et cognitifs qui participent au travail de conceptualisation doivent être analysés par une exploration détaillée de la dimension micro-textuelle. La classification de segments textuels ne constitue qu'une première étape à cette exploration et doit être complétée par des parcours de lecture qui incluent des stratégies de catégorisation reliées à l'expertise du chercheur. Ce processus d'interprétation généralement fait à la main peut être *assisté* par différentes stratégies d'annotation automatique. Nous avons expérimenté dans cette recherche une assistance à l'annotation catégorielle d'un sous-corpus de segments issu d'une opération de classification, le tout afin d'assister le plus finement possible une démarche d'analyse conceptuelle. L'identification par EXCOM des relations causales présentes au sein d'une classe spécifique de segments a permis dans cette recherche d'approfondir la dimension « psychologique » du concept de LANGAGE chez Bergson. Relevant des schèmes sémantico-cognitifs, ces relations causales ont permis de mettre en lumière la façon particulière dont Bergson *perçoit* et *structure* l'influence du langage sur les différents processus de la psyché. La présente recherche démontre que l'annotation automatique d'EXCOM peut permettre d'utiliser les avantages de la catégorisation automatique tout en demeurant fidèle aux exigences herméneutiques de la démarche interprétative du chercheur.

Terminons en rappelant que la stratégie d'annotation présentée dans cet article n'adopte qu'un angle d'entrée (la causalité) et qu'elle ne s'applique qu'à une dimension du concept de « langage » (la psychologie). Le concept de « langage » se déploie dans l'œuvre de Bergson au sein de plusieurs axes conceptuels qui mériteraient d'être explorés par la même stratégie développée dans la présente recherche.

Références

- Alrahabi M. et Desclès J.-P. (2009). EXCOM: Plate-forme d'annotation sémantique de textes multilingues. Laboratoire LaLIC. In *TALN 2009*, Senlis, 24-26 juin.
- Bradley J. (2007). Thinking about interpretation: Pliny and scholarship in the humanities. *Literary and Linguistic Computing*, 23(3) : 263-279.
- Brandom R. (1994). *Making it Explicit*. Cambridge: Harvard University Press.
- Brunet E. (2008). Les séquences (suite). In *Actes des JADT 2008*.
- Chartier J.-F., Meunier J.-G., Danis J. and Jendoubi M. (2008). Le travail conceptuel collectif : une analyse assistée par ordinateur du concept d'ACCOMMODEMENT RAISONNABLE dans les journaux québécois. In *Actes des JADT 2008*, p. 297-307.
- Desclès J.-P. (1990). *Langages applicatifs, langues naturelles et cognition*. Paris : Hermès.
- Dixsaut M. (1985). *Le naturel philosophe. Essai sur les dialogues de Platon*. Paris : Les belles lettres-Vrin.
- Djioua B. and Desclès J.-P. (2006). EXCOM : an automatic annotation engine for semantic information. In *Actes de FLAIRS 2006*, Floride.
- Garcia D. (1998). *Analyse automatique des textes pour l'organisation causale des actions. Réalisation du système informatique COATIS*. Thèse de doctorat, Université de Paris-Sorbonne.
- Jain A. and Flynn P.J. (1999). Data Clustering : A Review. *ACM Computing Surveys*, 31(3) : 264-323.
- Lebart L. and Salem A. (1994). *Statistiques textuelles*. Paris : Dunod.
- Loiseau S. (2003). Philosophical discourse from autonomy to engagement: Deleuze commentator of

- Spinoza. In Fløttum, K. and Rastier, F., editors, *Academic discourse. Multidisciplinary approaches*, Oslo: Novus: pp. 36-54.
- Martinez W. (2003). *Contribution à une méthodologie d'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de Doctorat en Sciences du Langage, Université de la Sorbone nouvelle-Paris 3.
- Mayaffre D. (2008). Quand « travail », « famille », « patrie » co-occurrent dans le discours de Nicolas Sarkozy. Étude de cas et réflexion théorique sur la co-occurrence. In *Actes JADT2008*.
- Meunier J.-G. and Forest, D. (2009). Lecture et analyse conceptuelle assistée par ordinateur : premières expériences. In Desclés, J.-P. and Le Priol, F., Dir. publ., *Annotation automatique et recherche d'informations*. Paris : Hermès.
- Poudat C (2006). *Typologie des concepts de linguistique : évaluation et élaboration en corpus de critères discriminants*. *Corpus en Lettres et Sciences sociales*, XI, 2: 91-101.
- Rey G. (1983). Concepts and stereotypes. *Cognition*, 15: 237-62.
- Rockwell G. (2003). What is text analysis, really?. *Literary and Linguistic Computing*, 18(2) : 209-219.
- Salton G. (1989). *Automatic Text ProcessinM*. Boston (MA) : Addison-Wesley.
- Talmy L. (1988). Force Dynamics in Language and Cognition. *Cognitive Science*, 12: 49-100.
- Unsworth, J. (2005) *What is Humanities Computing, and What is Not?*. In Braungart, G., Eibl, K. and Jannidis, F., editors, *Jahrbüch für Computerphilologie* 4, Paderborn: Mentis.
- Valette M. (2003). La genèse textuelle des concepts scientifiques. Étude sémantique sur l'œuvre du linguiste Gustave Guillaume. *Cahiers de lexicologie*, 89 : 125-142.

