L'analyse lexicométrique au service du corpus DECLICS2016

Griselda Drouet[(1)], Élisabeth Richard[(2)], Fabienne Moreau[(3)]

(1) Maître de conférences 7ème section – Université Rennes II
Laboratoire LIDILE (Linguistique Ingénierie Didactique des Langues) – EA 3874
(2) Professeure 7ème section – Université Rennes II
Laboratoire LIDILE (Linguistique Ingénierie Didactique des Langues) – EA 3874
(3) Maître de conférences 27ème section – Université Rennes II
Laboratoire LIDILE (Linguistique Ingénierie Didactique des Langues) – EA 3874

Revue *Éducation, Santé, Sociétés* Vol. 5, No 2 | pp. 107-131

Reçu le : 07/10/18 | Accepté le : 19/06/19

Résumé : Dans le cadre d'un projet interdisciplinaire, une analyse linguistique quantitative d'un corpus d'entretiens en santé est menée. Elle s'appuie sur l'enregistrement d'échanges oraux entre des médecins et/ou psychanalystes et leurs patients. L'objectif est de saisir ce qui se dit en conservant un regard profane (celui du linguiste) en matière médicale. L'analyse, outillée du logiciel *ad hoc* Iramuteq mettant en valeur des foyers lexicaux significativement représentatifs des discours produits, à partir d'analyses factorielles, est effectuée en deux étapes. Le corpus complet est d'abord exploré faisant ressortir les premiers foyers conceptuels communs à tous les locuteurs ; ensuite il est découpé en séparant les paroles des patients de celles des soignants (médecins et psychanalystes). Plusieurs foyers conceptuels sont extraits au sein d'une classification descendante hiérarchisant les mots contribuant le plus à définir le corpus DECLICS2016 dans son ensemble comme pour chaque locuteur-type. *In fine*, les enjeux mis au jour lors de ces échanges particuliers mènent à interroger la formation en santé publique pour améliorer l'écoute à l'hôpital.

Mots-clés : Linguistique, analyse de corpus, discours médicaux, lexicométrie, text-mining

Introduction

Le projet DECLICS [1] part du constat récent de difficultés rencontrées par le secteur hospitalier et ses acteurs face au besoin croissant d'écoute des patients. En effet, les pratiques des professionnels évoluant, les formations à l'éthique et au contact humain se révèlent indispensables à la formation des futurs praticiens. Or, ces formations n'existant que peu ou pas encore dans le domaine médical, il est nécessaire de les construire à partir de

1. La Région Auvergne Rhône-Alpes et l'Université Clermont Auvergne ont signé pour la période 2016-2021 une convention autour du projet DECLICS — *Dispositif d'Etudes CLiniques sur les Corpus Santé*. Cette collaboration scientifique engage des laboratoires de recherches en Sciences Humaines et Sociales et en Médecine pour fournir à la recherche universitaire des retombées sociétales et une application en santé.

recherches appliquées qui puissent mettre en avant les aspects relationnels et communicationnels les plus pertinents à l'entretien d'une relation soignant-patient efficace. L'aspect interdisciplinaire de notre étude, au croisement des sciences humaines et des problématiques médicales de santé publique, est ainsi attendu des communautés hospitalières. Le recours à l'analyse linguistique d'un corpus de données langagières recueillies *in situ* s'avère en ce sens tout à fait judicieux (Vergely et *al.*, 2009) L'apport de la linguistique de corpus (Léon, 2008), qui travaille exclusivement sur des paroles authentiques, se veut, dans tout champ de recherche, contributive sur des aspects à la fois quantitatifs et qualitatifs. La linguistique vise le quantitatif, car l'utilisation de logiciels de fouilles automatisées liées à des analyses statistiques et lexicométriques (Reinert, 1983 ; Ratineau, 2009 ; Zhai et Massung, 2016) permet de mesurer le poids de différentes tendances à l'œuvre dans un corpus. La linguistique vise aussi le qualitatif par les retours sur analyses qui cadrent le regard porté aux tendances émergentes tout au long du processus quantitatif, car le linguiste dispose d'une expérience en matière de consultations de nombreux corpus de référence. La facette quantitative de l'analyse permet, à partir de tendances chiffrées, d'interroger la spécificité du corpus. La finesse des résultats dépend des interprétations subséquentes que le linguiste pose sur ce qui pourrait paraître peu évident à la première lecture.

Cet article expose dans une première partie le cadrage conceptuel et méthodologique de l'étude, déployé à partir d'une étude antérieure montrant l'intérêt de relier la linguistique et la médecine. Dans une seconde partie, sont détaillées très précisément les analyses effectuées à partir des résultats obtenus en utilisant les logiciels de fouilles automatisés. Dans une troisième partie, nous ouvrons un espace à la discussion des interprétations afin de proposer des perspectives de recherche en matière de formation en santé publique.

1 Cadrage conceptuel et méthodologie d'analyse du corpus

1.1 Cadre théorique

Notre recherche s'ancre sur la problématique générale du projet DECLICS. Elle s'inscrit dans le prolongement de recherches précédentes portant sur l'utilité de la linguistique pour décrypter le discours médical (Haxaire et *al.*, 2006, 2010 ; Drouet et *al.*, 2017 ; Ingrand et *al.*, 2018a et b ; Richard et *al.*, 2018). Elle prend également source dans la lignée des recherches portant, en linguistique, sur l'étude des phénomènes syntaxiques et énonciatifs de la langue à l'oral (Richard 2018a ; Drouet 2019 ; Richard et Drouet, 2014, 2016) et de récurrences, répétitions et reformulations (Le Bot et *al.*, 2008 ; Schuwer et *al.*, 2008 ; Richard, 2014). Nous évoquerons en premier lieu en quoi les liens linguistique et médecine sont fructueux. Puis nous déclinerons nos appuis théoriques dans le champ de la linguistique.

1.1.1 L'apport du linguiste à la médecine

Nous avons déjà initié un croisement d'analyses entre corpus d'entretiens médicaux et analyse linguistique dans le champ de la prévention du cancer. Cette recherche appliquée comparait l'efficacité d'un dépistage personnalisé à la prévention classique non personnalisée (Ingrand et *al.*, 2018). Le dispositif testait l'intervention téléphonique réitérée d'une infirmière spécialisée dans le discours de prévention vers des patients atteints du cancer du colon (cas index) et auprès de sa fratrie afin d'inciter cette dernière (304 personnes

en tout) à choisir un dépistage par coloscopie. Ces interventions, enregistrées, transcrites étaient ensuite soumises à l'analyse croisée d'un chirurgien, d'un ingénieur médical, de sociologues et de linguistes. Ce corpus comportait les 98 transcriptions d'entretiens téléphoniques entre l'infirmière et le sondé, soit 3574 formes lexicales à analyser, représentant environ 120 000 occurrences effectives. La linguistique de corpus intervient à ce niveau et doit compter sur l'aide de logiciels pour sélectionner des tendances.

L'analyse outillée a permis d'isoler deux points de vue très différenciés portant sur le discours de prévention. Il s'est avéré que le discours de prévention semblait mieux assimilé au sein des discours émanant du groupe soumis à l'intervention personnalisée alors qu'il restait questionné ou rejeté dans le discours des témoins soumis à la prévention non personnalisée. Dans ce dernier groupe, a pu être démontré, grâce à l'étude discursive menée, que les discours étaient moins homogènes : distincts, multiples et polyphoniques, ils indiquaient que le locuteur ne s'était pas approprié les foyers conceptuels (mots clefs, intérêts, explications) permettant de s'approprier une conduite de prévention. En revanche, dans le discours des fratries du groupe intervention personnalisée, la diversité s'était réduite positivement : le discours du locuteur intégrait les notions de *dépistage/opération,* de *tests,* de *risques et* de *cancer* dans une même voix, c'est-à-dire au sein d'un discours intégré, argumenté et cohérent. L'observation de la circulation des voix énonciatives (Rabatel, 2005) dans les interactions, mise au jour par les linguistes a permis des échanges fructueux au plan interdisciplinaire (Ingrand et *al.*, 2018) [2].

1.1.2 Les études lexicométriques en linguistique

Si l'analyse lexicométrique n'est pas nouvelle en linguistique, elle s'avère, en revanche, de plus en plus outillée, et les logiciels d'appui à cet outillage, s'ils apportent une aide précieuse, supposent de maitriser ce qui se cache derrière les chiffres. Pour notre part, nous nous appuierons sur les bases théoriques de l'analyse lexicométrique du discours (Reinert, 1983 ; Ratineau 2009 ; Zhai et Massung, 2016) que nous couplerons avec les outils utilisés en analyse conversationnelle et interactionnelle intégrant depuis longtemps les aspects énonciatifs comme indicateurs privilégiés (Goffman, 1974 ; Brown et Lévinson, 1987 ; Traverso, 2001 ; Kerbrat-Orecchioni, 2005 ; Maingueneau, 2012 ; Rabatel, 2016). Cette double perspective permet de poser des interprétations sur les récurrences lexicales (lexicométrie) en tenant compte des aspects liés aux spécificités de l'échange. Les répétitions, pour exemple, permettent de mieux comprendre en quoi les discours tenus sont cohérents les uns avec les autres (Richard et Noailly, 2014). Les contradictions performatives ou mises en scène, au contraire, permettent de mieux comprendre où le discours se fait problématique (Richard et Drouet, 2017 ; Drouet 2019). De plus, on dispose de recherches croisant les aspects linguistique et médical (Haxaire et *al.*, 2003, 2005, 2010 ; Chauveau Thoumelin, 2014) qui ont montré l'importance de vérifier en quoi le discours était un marqueur de subjectivité à l'œuvre. Aucun discours n'est neutre, et le discours des médecins comme celui de leurs patients comportent des indices de prise en considération et d'interprétation de discours autres que chacun des locuteurs convoque pour le prendre à son tour en charge ou, au contraire, pour s'y opposer explicitement ou implicitement.

2. Les analyses ne sont pas restées cloisonnées, puisqu'elles ont été confrontées aux analyses épidémiologiques et sociologiques de nos collègues chercheurs, et se sont très clairement appuyées les unes des autres. Il s'agissait alors à proprement parler d'une étude interdisciplinaire, qui a mené à la publication de l'ouvrage collectif.

Lebart et Salem (88, p. 183) définissent la lexicométrie comme un : « ensemble de méthodes permettant d'opérer des réorganisations formelles de la séquence textuelle et des analyses statistiques portant sur le vocabulaire d'un corpus de textes ». Du point de vue technique, l'analyse lexicométrique est ainsi une méthodologie qui, s'appuyant sur la récurrence lexicale, définit les thématiques saillantes agissant au cœur des échanges comme foyers conceptuels structurants. Elle permet dans un second temps, en déployant des facteurs opposant ou distinguant certains termes clefs, de mieux cerner les subtilités, les marques et les indices de la parole (Reinert, 2004, 2007, voir infra). Ceux-ci permettent de caractériser le niveau de cohérence et d'ajustement entre interlocuteurs, ou au sein du discours d'un même locuteur.

Dans le cadre de cette étude, l'analyse lexicométrique sera mise au service des attendus du projet DECLICS, que l'on peut résumer par les questions suivantes : Qu'est-ce qui dans le corpus prouve que les acteurs (professionnels de santé et patients) s'entendent, s'écoutent mutuellement ? Peut-on dégager des éléments lexicométriques qui retracent une écoute effective (ou au contraire un manque d'écoute effective) ? Partant de ces problématiques communes au projet, l'analyse lexicométrique explore des pistes à développer ultérieurement dans le cadre interdisciplinaire.

1.2 Cadre méthodologique

Nous avons soumis à l'analyse lexicométrique huit enregistrements du corpus DECLICS2016, ce qui correspond à environ 260 minutes de temps de parole (4h20 min), soit 159 869 mots (2,15 Mo de données).

1.2.1 Les données : le corpus DECLICS2016

La spécificité de ce corpus réside dans le type d'interaction qu'il convoque : quatre entretiens médicaux entre patient(s) et médecin(s) d'une part, et quatre entretiens entre patient(s) et psychanalyste(s) de l'autre. C'est la diversité des voix des différents énonciateurs (Rabatel, 2005, 2016) que nous avons cherchée à analyser.

1.2.2 L'utilisation du logiciel Iramuteq

Nous avons mené une analyse lexicométrique quantitative réalisée avec l'outil Iramuteq développé par Ratineau (2004, 2007) qui propose une classification hiérarchique descendante (Reinert, 1983, 1986, 1991) des différentes occurrences lexicales employées dans le discours. Sans entrer dans les détails du fonctionnement statistique du logiciel, nous rappellerons que le logiciel Iramuteq est une interface permettant d'aboutir à des analyses multidimensionnelles de textes *via* le logiciel statistique R. Les résultats sont finalement affichés par l'interface (Loubère et Ratineau, 2013). La Classification Hiérarchique Descendante (voir Reinert, 2004, pour le détail) est décrite par Reinert en ces termes : *« Il s'agit, non pas de comparer les distributions statistiques des « mots » dans différents corpus, mais d'étudier la structure formelle de leurs cooccurrences dans les « énoncés » d'un corpus donné. »* (Reinert, 1993, p. 9).

L'usage d'Iramuteq facilite la présentation des données en proposant une visualisation des récurrences des associations des items lexicaux (co-occurrences) qui donnent un aperçu direct des objets de discours à différencier quantitativement : *« la co-occurrence des mots*

pleins appartenant à un même segment [qui] est une trace possible de l'acte d'énonciation, et donc de son contenu » (Reinert, 2007, p. 193). Le procédé de classification hiérarchise, en les regroupant, les segments du texte en fonction de la distribution différenciée de leur vocabulaire. Nous nous sommes appuyées sur le nombre x de « mondes lexicaux » extraits automatiquement, et les avons interprétés comme autant de « points de vue » actualisés par le discours comme définit par Reinert (93, p. 11) :

> « Un énoncé [qui] traduit [...] davantage un point de vue particulier plutôt qu'une représentation, le point de vue impliquant en son centre l'existence d'un "sujet" dans une certaine modalité du faire ou de l'être. Cette notion n'a donc rien d'absolu. Elle est relative à l'activité ou à l'état d'un sujet et ne renvoie pas obligatoirement à un système de références préétabli, celui-ci pouvant l'être ou non, reconstruit, imaginé... Notre hypothèse principale consiste justement à considérer le vocabulaire d'un énoncé particulier comme une trace pertinente de ce point de vue il est à la fois la trace d'un lieu référentiel et d'une activité cohérente du sujet-énonciateur. Nous appelons mondes lexicaux, les traces les plus prégnantes de ces activités dans le lexique. »

Malgré la facilité *a priori* de l'utilisation du logiciel, l'interprétation des résultats ne va pas de soi : elle s'effectue pas à pas. Après le scan qui permet un décompte global des occurrences, des formes et des hapax[3], on extrait ces éléments à partir de chaque locuteur type : on regroupe les paroles des patients, des médecins et des psychanalystes. L'ensemble de ces formes[4] représente la spécificité du corpus, ventilée sur chaque locuteur engagé dans l'entretien de soin. On traite les occurrences comme un découpage, des segments analysés. Les formes « pleines » ou « actives » (substantifs, verbes, adjectif, adverbes) représentent les mots lemmatisés qui sont ramenés à l'infinitif pour les verbes, au masculin singulier pour les noms et les adjectifs. Les mots-outils, nécessaires à la syntaxe (pronoms, déterminants, connecteurs, auxiliaires) sont pris en compte dans une analyse supplémentaire. Souvent négligés, ces mots outils représentent au sein de notre orientation argumentative et pragmatique des échanges ce qui se joue dans l'interaction et révèlent les postures énonciatives du locuteur (Richard, 2018b). Nous faisons de cette étape, le traitement des mots-outils, une étape obligatoire, que nous appliquons au processus de dégagement de sens de l'analyse lexicométrique globale.

1.3 Cadrage de la problématique

Qu'apprend-t-on dans ce corpus en matière d'écoute ? L'analyse linguistique est conduite selon une méthodologie objective portant exclusivement sur les données recueillies et transcrites *in extenso*. Aucune représentation ou hypothèse *a priori* n'est venue augmenter les données brutes, selon les méthodologies en usage dans la linguistique de corpus (Léon, 2008 ; Cori et *al.*, 2008). En revanche, l'étude conduite permet d'émettre des hypothèses, au fil de l'analyse, sur les contenus discursifs mis en exergue. De quoi parle ce corpus ? Les différents locuteurs parlent-ils tous de la même chose (et de la même manière) ? Pour répondre à cette problématique, spécifiquement appliquée au corpus DECLICS2016, nous

3. Un « hapax » est une forme qui n'apparaît qu'une seule fois dans le corpus.

4. Lebart et Salem (88 : 182) définissent la notion de « forme » de la manière suivante : « archétype correspondant aux occurrences identiques dans un corpus de textes, c'est à dire aux occurrences composées strictement des mêmes caractères non-délimiteurs d'occurrence. »

serons amenées à considérer la dimension dialogique du discours, les différentes postures et instances énonciatives, le « feuilleté énonciatif » (Rabatel, 2005, 2007), que nous dégagerons de l'étude lexicométrique. Le linguiste pourra alors, au final, donner à voir pour chaque acteur, en tant que locuteur type médecin, locuteur type patient, locuteur type psychanalyste, les figures de reprises du dit, d'écoute, ou au contraire d'absence de reprise, d'incohérence dans les entretiens visant le soin.

La visualisation de données linguistiques est nouvelle dans le champ de la formation médicale. La contribution entend montrer en quoi la présentation méthodique et minutieuse des items lexicométriques constitue une présentation visuelle et ainsi une initiation facilitée pour mieux comprendre (puis gérer auprès des professionnels de santé) ce qui est au cœur des malentendus potentiels dans les entretiens soignants-patients.

2 Analyses des données et résultats

2.1 Analyse du corpus global : premier regard sur l'ensemble des discours

Un premier scan rapide avec le logiciel nous donne le décompte suivant pour :

- le discours des patients : 32379 occurrences, 3132 formes, 1628 hapax,
- le discours des psychanalystes : 11388 occurrences, 1456 formes, 808 hapax,
- le discours des médecins : 11501 occurrences, 1603 formes, 879 hapax.

2.1.1 Deux mots représentatifs de l'échange : @mh, d'accord@

L'analyse préliminaire lancée fait apparaître un nuage de mots, hiérarchisant les emplois : la taille des items est représentative de leur fréquence dans le discours [fig1]. Ce rendu visuel de fréquence lexicale des items langagiers typifie le corpus global DECLICS2016.



Figure 1 – Nuage de mots du corpus global DECLICS2016.

On constate que les marqueurs discursifs « *mh* » et « *d'accord* » sont les éléments linguistiques les plus fréquents et situent ainsi le corpus global comme relevant d'une interaction langagière, au sens d'une conversation, alternant des validations (« *d'accord* ») et des marques moins explicites (« *mh* » comme marqueur d'accord phatique ou d'hésitation). DECLICS2016 est un corpus interactif, d'échange dynamique marqué par l'écoute (« *mh* ») et l'approbation (« *d'accord* »).

2.1.2 Mise au jour de deux univers généraux distincts

Une première analyse menée sur le corpus global décrit en les caractéristiques principales : de quoi parle-t-on pendant ces 4h20 d'enregistrement ? La Classification Hiérarchique Descendante, ou méthode Reinert, permet de regrouper des mondes lexicaux et de mettre en évidence les thématiques générales du corpus [fig2]. On met ainsi en évidence des mots qui apparaissent le plus souvent ensemble, des topics, ou univers généraux, qui se subdivisent en sous-classes. Qu'est-ce qui est cooccurrent, et pourquoi ?

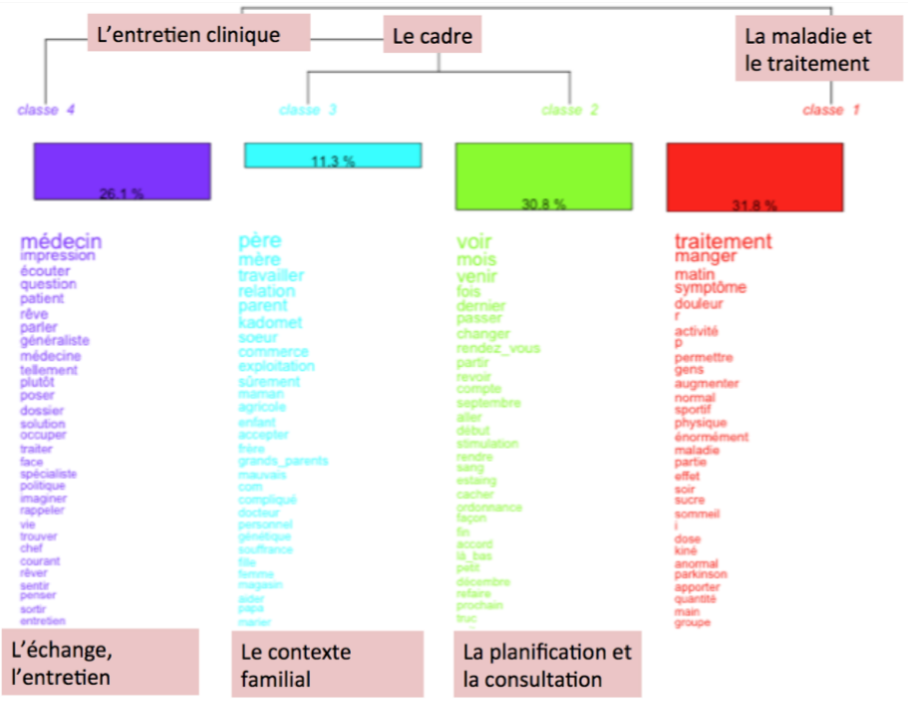


FIGURE 2 – Analyse CHD corpus global [5].

Légende : L'arborescence est directement fournie par le logiciel Iramuteq. Le dendrogramme donne la liste des formes les plus associées pour chaque classe. Les titres définissant chaque univers et/ou classe sont issus de notre interprétation linguistique.

De cette classification, deux grands univers de discours ressortent : l'univers 1 qui est composé d'une seule hyper-classe, la classe 1 (rouge) et l'univers 2 qui se subdivise en

deux hyper-classes : la classe 4 isolée (violette) et l'hyper-classe associant les classes 3 + 2 (respectivement bleue et verte).

La première remarque qui peut être faite est que ces deux univers ne sont pas équilibrés. En effet, quand l'univers 2 représente 68,2% du corpus, l'univers 1, lui, totalise seulement 31,8 % de ce dernier. Nous avons sous les yeux une répartition deux tiers / un tiers, qu'il convient alors de décrire plus précisément. L'objectif est de comprendre ce que regroupe chacune des classes mais aussi ce qui isole nettement 1 des trois autres classes et ce qui rapproche (2+3) de 4. C'est bien cette dynamique des classes que le logiciel permet de visualiser et qu'il convient ici d'interpréter [6].

— Univers 1 – Classe 1 (31,8% du corpus) : la maladie chronique.

La classe 1 (en rouge) occupe à elle seule 31, 8% du corpus. Les mots les plus représentatifs de cette classe sont : « *traitement, manger, matin, symptôme, douleur* ».

C'est la classe qui parle de la maladie et de son traitement. Le lexique est révélateur du corps malade « *douleur* (vs *souffrance* dans la classe 3), *symptômes, syndrome, parkinson, diabète, insuline, colopathie* » mais aussi du traitement « *dose, quantité, cinquante, dix, augmenter* » et de son régime « *manger, pain, patate, sucre* », le découpage du temps « *matin, soir, pendant* » structurant toute posologie « *pendant X jours, le matin* ». Le pronom le plus représentatif de cette classe est le pronom indéfini « *on* » : on fait alors l'hypothèse qu'il représente le corps médical, et permet d'en marquer la prise de position. La classe de la justification du traitement et le marqueur reformulatif « *c'est-à-dire* » qui apparaît également et distingue le discours représenté comme discours explicatif, justificatif.

Hypothèse univers 1 : C'est l'univers de la maladie et de son traitement.

— Univers 2 (68,2% du corpus) : l'entretien clinique.

L'univers 2 concentre l'entretien clinique. En effet, les mots les plus représentatifs de cet univers ont pour point commun de construire l'acte médical.

Hyper-classe : classe 4 (en violet). Elle représente à elle seule 26,1 % du corpus. C'est la classe de la communication du patient avec son soignant, le fond de l'entretien clinique, comme en attestent les mots qui apparaissent le plus fréquemment ensemble dans le corpus : on y trouve des pronoms qui désignent l'autre : « *celui-là, monsieur, ceux-là* (eux par rapport à moi) », des verbes de parole et d'expression des sensations/émotions : « *écouter, parler, poser, rêver, penser, répondre, sentir, imaginer, traiter, rappeler, espérer, ressentir* », ainsi que des adverbes d'intensité et de point de vue : « tellement, maximum, spécialement, justement, évidemment, malheureusement ».

Résultat : cette classe traite de l'échange langagier, de la communication entre les différents acteurs dans les entretiens. C'est ce(ux) dont on parle lors de l'entretien.

Hyper-classe 3-2 (turquoise et verte). Le cadre de l'entretien

6. Pour rappel, l'objectif de ces interprétations n'est pas de mener à des conclusions fermes mais plutôt d'émettre des hypothèses quant aux relations entre les univers lexicaux ainsi stabilisés dans le discours.

La classe 3 (en turquoise) isole un discours autour du contexte familial. Cette classe qui n'est pas très importante quantitativement (seulement 11,3% du corpus) stabilise très distinctement les mots de la famille des relations familiales « *père, mère, relation, parent, fille, sœur, enfant, frère, grands-parents, maman, génétique* » Un autre phénomène particulier dans ce corpus est que cette classe présente également toute une isotopie qui traite des métiers : « *commerce, exploitation, agricole, magasin* ». Les formes personnelles représentatives de la classe sont essentiellement des marques possessives de troisième personne « *sa* » ou « *leur* », mais aussi de première personne « *ma, mon* ». Entremêlée à ces différents lexèmes surgit une chaîne évaluative négative : des adjectifs évaluatifs : « *compliqué, mauvais* », des lexèmes « *souffrance, souffrir* », l'adverbe de négation temporel : « *jamais* » et des indéfinis de la quantité nulle : « *rien, aucune* ». On souligne néanmoins la présence du verbe « *accepter* » et d'adverbes notables : « *enfin, puis*, », cadratifs de la narration.

Résultat : c'est la classe de la narration de l'histoire familiale.

La classe 2 (en vert) est très importante puisqu'elle représente à elle seule 30,8% du corpus. Les mots les plus représentatifs sont « *voir, mois, venir, fois, dernier* ». C'est le discours qui s'occupe de la planification de la consultation. Le pronom de troisième personne domine, ainsi que les termes d'adresse : « *Madame* ». Les verbes sont les verbes de la consultation médicale : « *voir, venir, prescrire, revoir, refaire* » ; le lexique également puisque l'on y trouve les substantifs de l'administration médicale : « *rendez-vous, ordonnance, décembre* ».

Résultat : Cette classe est celle de la planification de la consultation. Elle dresse le cadre clinique dans lequel aura lieu l'entretien.

Le regroupement des classes 2 et 3, aussi petit soit-il, est extrêmement important comme en atteste son étalement dans le corpus (*cf.* l'analyse AFC ci-après). Le sujet abordé ici est la thématique de l'histoire familiale. Cette thématique est intrinsèquement liée à l'expression de la souffrance.

Hypothèse globale pour l'univers 2 : Les classes 3 et 2 sont reliées : il y a un lien entre la famille, l'histoire familiale et la planification médicale. Le patient est très présent, actif, dans ces deux classes, alors que dans la classe 4 (isolée), le patient semble désincarné, il est le réceptacle de la consultation clinique.

2.1.3 Interprétation : la consultation clinique au carrefour de tous les discours

L'analyse Factorielle des Correspondances (AFC) ci-dessous [fig3] nous amène à ajouter quelques remarques aux résultats et hypothèses formulés précédemment (2.1.2).

On constate tout d'abord que la représentation graphique met à l'écart la classe 1 (rouge), ce qui souligne que l'univers lexical qui la caractérise est très éloigné des trois autres classes, très distinctif. Cet isolement, couplé au fait que la classe n'est pas si importante quantitativement, montre que le discours de la maladie et de son traitement ne résume pas en soi l'entretien. L'univers 2 quant à lui se détache sur l'axe vertical. On remarque que les classes 2 et 3 (ici verte et bleue) se mêlent en partie, laissant la classe 4 plus isolée, comme le montrait aussi la CHD. La classe de la famille et celle de la consultation clinique présentent donc un lien qu'il faudra approfondir dans les analyses suivantes.

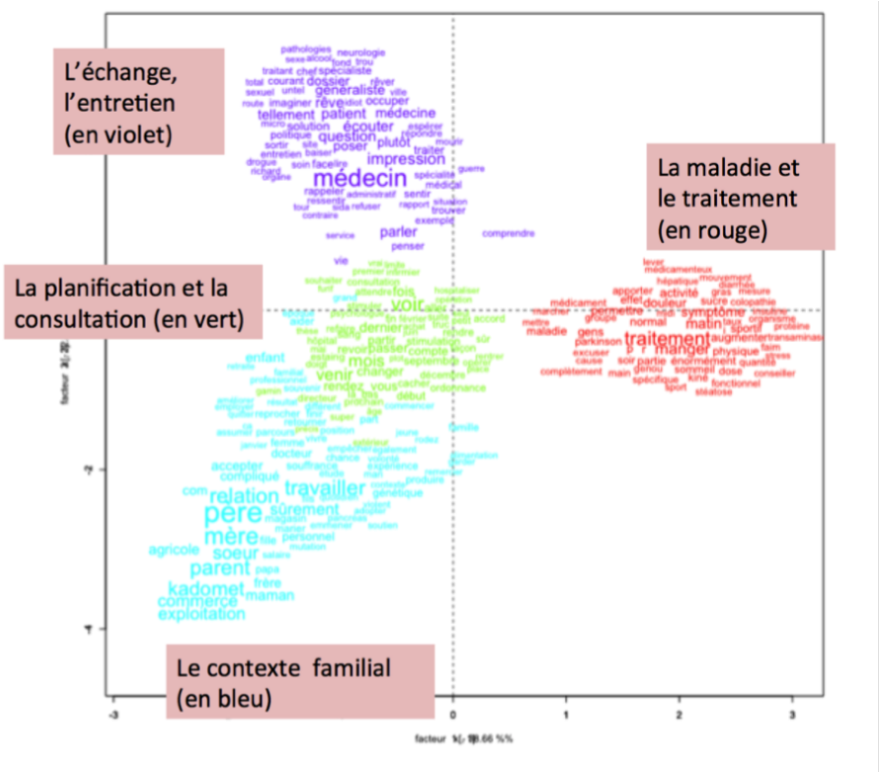


FIGURE 3 – Analyse Factorielle des Correspondances (AFC) appliquée au corpus global.

Légende : Les formes sont colorées selon les classes obtenues par la méthode Reinert. Les thématiques présentées dans les encadrés relèvent de notre interprétation linguistique.

Ce qui nous paraît le plus intéressant ici, c'est de constater la position centrale de la classe 2 : c'est la classe de la consultation clinique, celle qui est au carrefour de tous les discours. Même si la classe est faible quantitativement, l'analyse factorielle révèle sa position centrale, ramassée sur elle-même, dense et semblant faire le lien, à différents degrés, avec toutes les autres classes. C'est le point de rencontre clinique : la consultation en elle-même.

2.2 Focus sur le discours des patients

2.2.1 L'hôpital et la vie à l'extérieur de l'hôpital

L'analyse des classifications montre [fig4] un discours qui se construit sur deux pôles, dont l'un n'est constitué que d'une seule classe (la classe 4 en violet), isolée de l'autre pôle qui stabilise, quant à lui, trois autres classes.

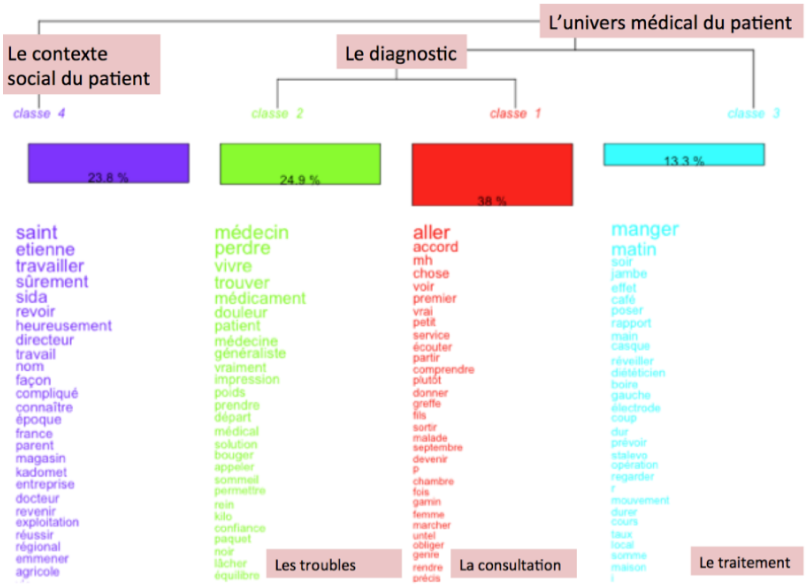


FIGURE 4 – Analyse CHD corpus patients.

— Univers 1 (23,8%) : une seule classe, la classe 4 : la cartographie sociale du patient.

La classe 4 englobe le contexte social du patient, la vie à l'extérieur de l'hôpital.

On y trouve des lexèmes de lieu : « *saint Etienne, régional, France, Londres, auvergne* », le lexique de la maladie : « *sida, kadomet, vie, mort, résultat, papier, docteur, souci, traitement, parkinson, réanimation, pompier* », le lexique du milieu professionnel : « *directeur, travail, magasin, entreprise, exploitation, agricole, commerce, étude, milieu, interne, avocat, retraite, école* » ainsi que le lexique de la famille : « *parent, relation, père, personnel, domaine, contexte* ». Apparaissent également des lexèmes positivement connotés : « *occasion, fête, courage, aide, chance, super* ».

De nombreux verbes exprimant des actions de la vie ordinaire étoffent cette classe : « *travailler, réussir, cacher, revoir, exister, habiter, imaginer, gêner* ». Des verbes d'état : « *paraître, décéder, mourir* », des verbes de mouvement : « *revenir, emmener, apporter, venir* », des verbes à valeur dynamique : « *aider, reprocher, conseiller, suffire, accepter, connaître, transmettre* » et des verbes qui parlent de transformation : « *évoluer, changer, laisser, améliorer, rester* ». On trouve également quelques adjectifs qualificatifs : « *compliqué, grand, jeune, facile* » *ainsi que des adverbes :* « *sûrement, heureusement, également, peut-être* » et beaucoup de chiffres et de dates : « *quatre-vingt-quinze seize, deux, cinq, quatorze, mille vingt-cinq, quatre-vingt quinze, cent* ».

Hypothèse univers 1 : cet univers parle de la vie du patient, dans tous les domaines : c'est la classe du contexte social, professionnel, personnel du patient. C'est une sorte de carte d'identité du patient.

— Univers 2 (76,2%) : l'hôpital : deux hyper-classes.

Hyper-classe de la classe 3 seule (13,3%) : le traitement et ses effets secondaires.

Dans cette classe, le malade parle lui-même (on y trouve le pronom « me ») de son corps affecté par la maladie et du traitement *per se*, ainsi que de ses effets secondaires. On y trouve le lexique de l'alimentation : « *manger, café, diététicien, boire* », le lexique du traitement : « *stalevo, électrode, casque, dose, anesthésie, piquer, opération machine* », le lexique du corps : « *jambe, main, genou, cervical, bras, force, mouvement* », couplé au lexique des effets indésirables : « *effet, coup dur, tremblement, trou, problème* ». Les adjectifs sont évaluatifs et connotés négativement : « *anormal, important* ». Les marques temporelles sont nombreuses : « *matin, soir, lundi, midi, lever, dimanche, mercredi, lendemain, hier, heure, après* » et les quantités « *du (partitif), quarante-sept, trente-cinq* », deux isotopies prototypiques de la posologie du traitement.

Hyper-classe (2+1) (62,9%) : le diagnostic.

Classe 2 (24,9%) : les troubles, la maladie.

Dans cette classe, apparaissent très majoritairement les lexèmes de la maladie : « *médecin* (vs « *docteur* » dans la classe précédente), *médicament, douleur, patient, médecine, généraliste, sommeil, victime, septicémie, crise, urgence, infirmier, soin, aide-soignant, maladie* ». Cette classe stabilise également le lexique du corps du patient : « *poids, rein, kilo, musculaire, sang, peau, corps, tête, estomac, dos, cœur* ». Se singularisent parmi le lexique précédent, quelques lexèmes qui parlent de sentiments : « *confiance, question, angoisse* ». De nombreux verbes relatifs aux soins médicaux marquent cette classe : « *souffrir, sentir, traiter, subir, soigner, shooter* ». Mais sont également présents des verbes qui parlent de la vie ordinaire : « *perdre, vivre, trouver, prendre, bouger, appeler, lâcher, permettre, demander, tenir, reprendre, tomber, protéger, baiser, reconnaître* ». Les adjectifs de cette classe font tous partis de l'isotopie de la maladie : « *médical, équilibré, politique, hypocondriaque, pénible* ». On remarque la présence d'un seul adverbe très significatif, relatif aux troubles du patient : « *malheureusement* ».

Classe 1 (38%) : la consultation.

La classe 1 stabilise la thématique de la consultation dans son sens administratif. On y trouve le lexique des soins médicaux : « *greffe, service, chambre, rendez-vous, stimulation, consultation, examen, courrier* » couplé au lexique de la maladie : « *malade, antécédent, hépatite* ». L'isotopie de la famille n'est pas tout à fait absente non plus : « *petit fils, gamin, femme, untel, mari, fille, sœur* ».

De très nombreux verbes apparaissent dans cette classe : certains ont pour point commun d'exprimer la notion d'échange : « *voir, écouter, comprendre, donner (rv), discuter, consulter, chercher, parler, envoyer, écrire* », d'autres parlent de mouvements et de changements d'état : « *hospitaliser, marcher, sortir, partir, devenir, rendre, obliger, modifier, mettre, retourner, préférer, rentrer, commencer, essayer* ».

Cette classe est marquée par la personne 3 : « *elle, leurs* ».

Résultat : c'est la classe de la consultation médicale, de l'acte administratif auquel doit se rendre le patient.

Hypothèse univers 2 : L'univers 2 stabilise la sphère médicale du patient, des troubles à la consultation, du diagnostic posé au traitement.

Conclusion : L'analyse CHD montre clairement deux pôles : l'intérieur de l'hôpital et l'extérieur de l'hôpital (la vie ordinaire). Ces deux pôles semblent bien cloisonnés dans le discours du patient.

2.2.2 Interprétation : la maladie et la consultation au cœur du discours médical

L'analyse factorielle [fig5] montre que le cœur du discours est composé de la classe de la maladie et de celle de la consultation (verte et rouge).

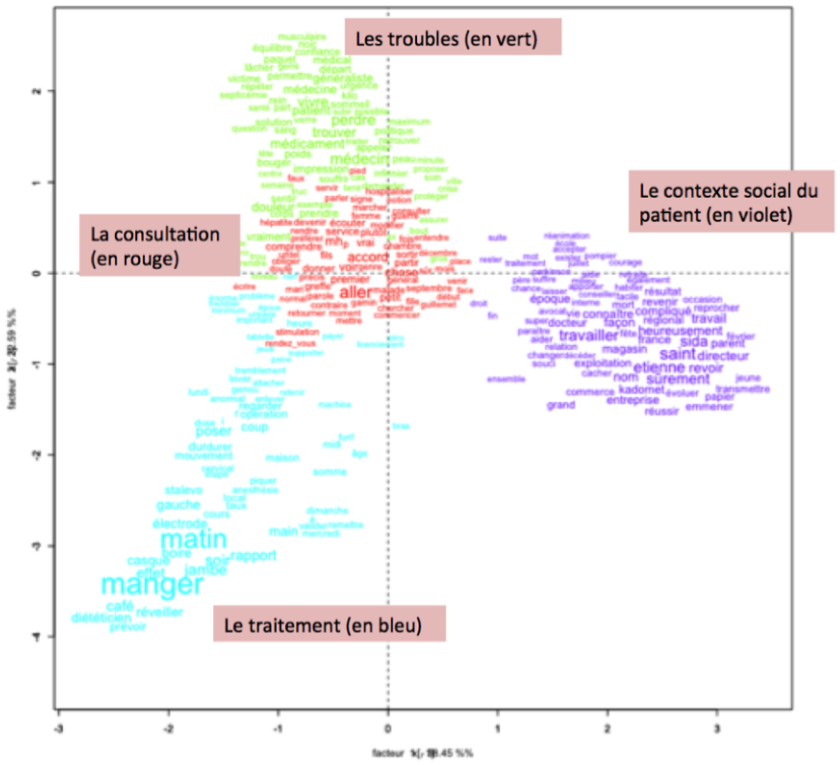


FIGURE 5 – Analyse factorielle corpus patients.

Ces deux thématiques sont encadrées par deux autres classes plus isolées : le traitement (en bleu), que l'on voit ici tisser un lien vers les deux classes précédemment évoquées, et plus complètement exclue encore, la classe violette qui englobe la thématique du contexte social du patient, sa vie en dehors de l'hôpital.

L'analyse factorielle confirme ainsi les tendances des quatre classes différenciées du dendrogramme [fig4] en plaçant cependant la maladie et la consultation au cœur de l'échange patients-soignants (médecins/psychanalystes).

2.3 Focus sur le discours des médecins

En recoupant le corpus, nous avons ensuite concentré notre attention sur le seul discours des médecins. Cette analyse est faite en vue de la comparaison avec le discours des seuls psychanalystes.

2.3.1 Un discours partitionné et chronométré

L'analyse des classes révèle d'emblée une particularité que ne présentaient pas les classifications précédentes. Le résultat montre un dendrogramme parfaitement équilibré [fig6], avec 4 classes d'environ 25 % chacune, symétriquement réparties en deux univers.

L'égale répartition des quatre classes illustre un discours parfaitement chronométré, témoignage de la maîtrise du temps de la consultation par le soignant.

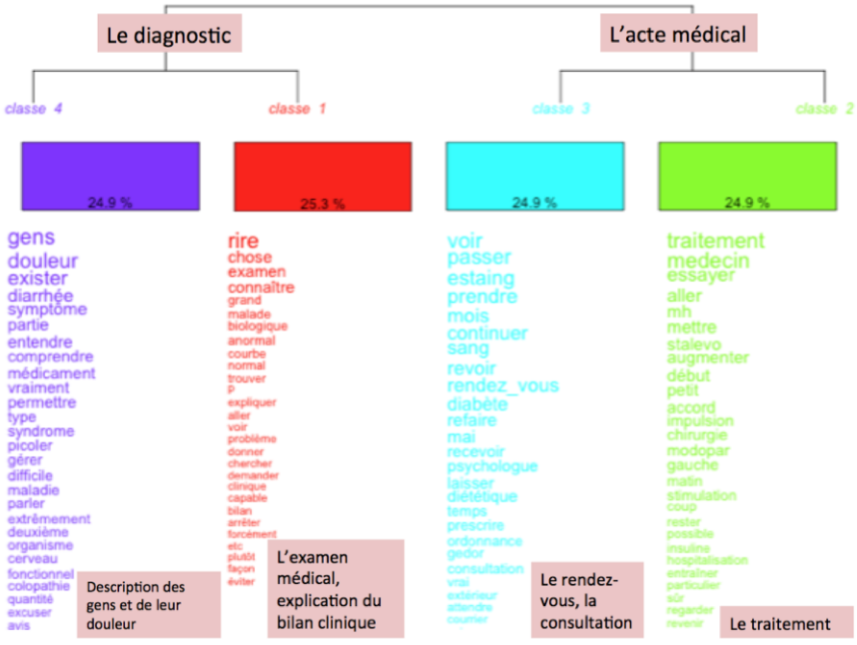


FIGURE 6 – Analyse CHD corpus médecins.

— L'univers 1 stabilise la moitié du corpus (50,2%). C'est l'univers du diagnostic.

La classe 4 (en violet) présente un quart du discours et parle de la description des « gens » (des patients) et de leur douleur.

Apparaissent ici des noms typiques de la maladie : « *douleur, diarrhée, symptôme, médicament, syndrome, maladie, organisme, colopathie, cerveau, quantité, avis mécanisme, expérience, signe, rapport, situation, régime, humain, diagnostic, victime, médecin* ». On trouve aussi des marques de la personne 5 : « *vous, vos* » et celles de la personne 3 et 6 « elle », « ils » qui désignent les « gens » On est dans un discours qui tend à objecti-

ver la douleur. Les adjectifs sont descriptifs : « *spécifique, diabétique, fonctionnel* ». Les verbes marquent l'observation de la douleur : « *entendre, comprendre, permettre, gérer, excuser, travailler, devenir, laisser, accepter, occuper* ». Enfin les adverbes sont intensifs : « *vraiment, extrêmement* ».

Résultat : la classe 4 fait une description des gens et de leur douleur mais il ne s'agit pas de la souffrance vécue et ressentie effectivement par le patient, ici le médecin objective la douleur, la met à distance : c'est la douleur clinique, décrite, non-incarnée.

La classe 1 (en rouge, 25,3%) stabilise des lexèmes de l'explication du bilan clinique. Les verbes en sont tout à fait représentatifs : « *connaître, expliquer, voir, demander, arrêter, éviter* ». Les noms le sont également : « *examen, malade, courbe, problème, médecin, maladie, bilan clinique* ». Les adjectifs sont descriptifs : « *malade, grand, biologique, anormal, normal* ». Les adverbes y sont explicatifs : « *forcément, c'est à dire, parce que* ».

Le pronom « *je* » subjectivise le discours, c'est la voix du médecin, trouve aussi le pronom indéfini « *les autres* » qui singularise d'autant plus la voix du médecin.

Les restrictifs sont intéressants aussi car ils sont très présents : « *rien, moins, sauf, non, ne* »

Résultat : Le discours est descriptif. Le médecin expose son diagnostic. C'est le discours du clinicien.

Hypothèse univers 1 : ce premier univers du discours représente le temps de parole du médecin consacré à l'échange avec le patient. Le médecin prend le temps de parler avec le patient, d'expliquer le bilan clinique. C'est le temps de l'entretien. Mais cet entretien est marqué par une certaine distance : le discours est technique, désincarné, le médecin est le clinicien et il expose son analyse de façon objective, savante.

— Univers 2 (49,8%) Le rendez-vous médical

La classe 3 (en bleu) représente 24,9 % du corpus. C'est la classe du rendez-vous médical. Y apparaissent les lexèmes typiques de la consultation, plutôt administratifs : « *rendez-vous, ordonnance, consultation, courrier, mesure exemple, sang, diabète, gedor, poids* ». Trois lexèmes se singularisent : « *aide, extérieur, psychanalyste* » renvoyant à la sphère extérieure à la clinique. Les adjectifs présents caractérisent la classe du médical : « *diététique, toxique (analyse ?) complémentaire* ». On trouve également des lexèmes temporels et locatifs spécifiques à la prise de rendez-vous : « *mois mai temps journée trois six cinq après ici* ». Les verbes sont nombreux et certains d'entre eux sont préfixé en « re » marquant l'itération : « *voir passer, prendre, continuer revoir, refaire, recevoir, laisser, prescrire, attendre, suivre, proposer, changer* ». Ce sont tous des verbes qui parlent du suivi de la consultation. L'adverbe : « *justement* » marque cette classe. Les pronoms sont englobants : « *nous, vous, je* ». Au « *je* » du médecin, au « *vous* » du patient s'ajoute le « *nous* » de la rencontre entre les deux.

Résultat : cette classe parle du rendez-vous médical, de la consultation et de son suivi.

La classe 2 (24,9%) est descriptive, c'est le début du traitement.

On y trouve un grand nombre de noms de médicaments comparativement aux autres classes : « *stalevo, modopar, metformine, toviaz, insuline* », des noms relatifs au démarrage du traitement : « *traitement, médecin, début, impulsion, chirurgie, stimulation, hospitalisation, élément, besoin, tremblement* ». Des quantités liées à la posologie : « *soixante-quinze, un quatre cinquante du* (partitif) » ainsi que des lexèmes temporels concernant la prise du traitement : « *matin, semaine, date, heure, dernier* ». Les verbes parlent d'expérimentation : « *essayer, aller, mettre, augmenter, rester, entrainer, regarder, marcher, revenir, rendre compte, garder, contrôler, arriver croire* ». Les adjectifs sont qualifiants : « *petit, possible, particulier, grave* ». Conjonction, préposition, et adverbes construisent le cadre temporel de cette classe : « *quand après déjà finalement* ». Le pronom utilisé dans cette classe est particulièrement intéressant, il s'agit ici du pronom « *on* » inclusif.

Résultat : le médecin s'implique dans cette partie du discours, comme en témoigne l'utilisation toute particulière du pronom « *on* » qui rassemble le médecin et son patient. C'est la classe qui parle de l'expérimentation du traitement.

Hypothèse univers 2 : Le médecin en action. Dans cette classe, le discours du médecin impose moins de distance par rapport au patient, c'est celle de l'expérimentation technique, celle du terrain.

L'analyse des classes révèle une partition du discours du médecin en quatre temps bien définis et également répartis, peut-être imposée par le format de la consultation – un temps donné précieux et compté pour le médecin. Le « plan » de la consultation est prédéterminé, réglé, l'objectif du médecin étant l'efficacité dans la communication.

2.3.2 Interprétation : la maladie et son traitement au cœur du sujet

L'analyse factorielle [fig7] révèle une configuration qui n'est pas aussi cloisonnée que l'on aurait pu attendre par rapport au dendrogramme [fig6].

En effet, si les classes violette et rouge sont nettement séparées (la description de douleur n'est pas confondue avec le bilan clinique), les classes verte et bleue en revanche sont véritablement intriquées l'une dans l'autre.

Hypothèse : la description de la douleur et le bilan clinique sont deux discours bien distincts tandis que l'acte médical mêle administratif et prescriptif.

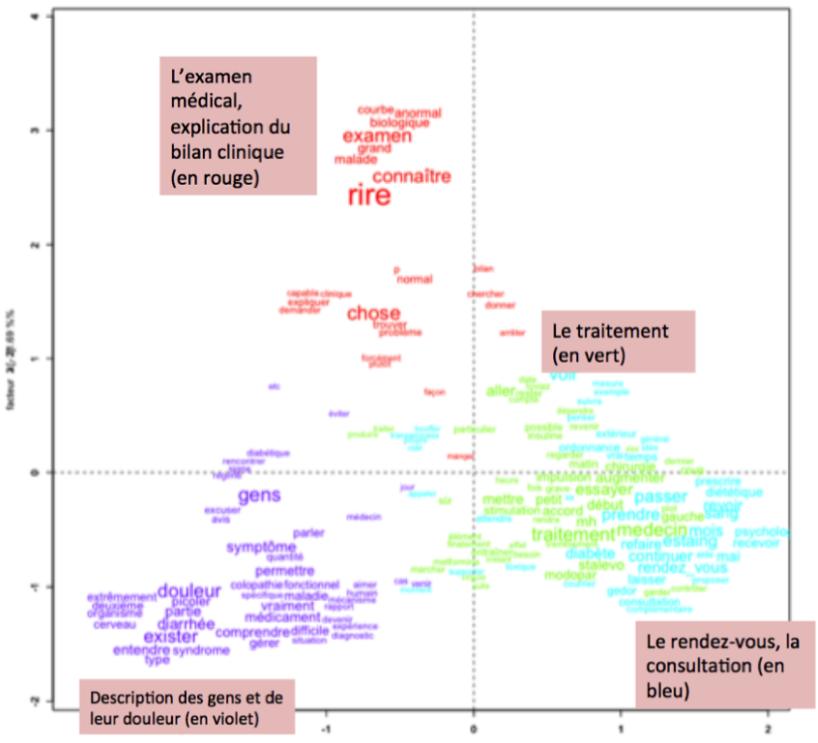


FIGURE 7 – Analyse Factorielle corpus médecins.

2.4 Focus sur le discours des psychanalystes

2.4.1 Parler avec le patient

On saisit ici une différence majeure avec l'analyse du discours précédent, celui des médecins. Les classes sont plus nombreuses, asymétriques et reliées entre elles en subdivisions [fig8]. Le discours est plus diffus dans l'entretien, moins partitionné. On observe deux univers déséquilibrés : l'un regroupant presque 83% du corpus, et l'autre seulement 17%.

— L'univers 1 stabilise une seule classe, la classe 5 (17, 2%), c'est la classe de l'analyse psychologique.

On y trouve les lexèmes : « *nuit, rêve, différence, colère, ordre, corps, chose, effet, médicament, groupe, limite, nom, exemple* » associés à l'interpellation co-occurrente du patient (« *tongul*[7] »). Les verbes parlent du ressenti physique et psychique : « *tomber, rappeler, mourir, réveiller, appeler, rêver, ressentir, sentir* » ainsi que de l'analyse en elle-même : « *revoir, continuer revenir, commencer* ». Les adverbes sont évaluatifs : « *finalement, beaucoup, souvent, juste* ». Le pronom de personne 1(auto-désignation) est dominant et la personne est mise en position réflexive : « *me* ».

7. Nom anonyme désignant la patiente.

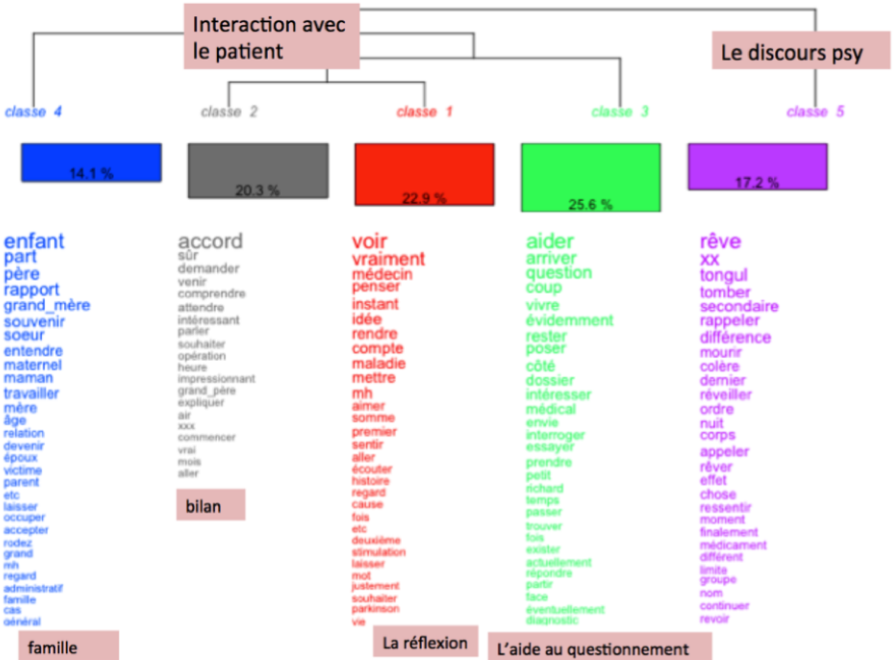


FIGURE 8 – Analyse CHD corpus psychanalystes.

Hypothèse univers 1 : c'est la classe de l'analyse, la matière du psychanalyste.

— L'univers 2 (82,9%)

Il est constitué de 4 classes reliées entre elles selon le schéma : {(1+2) + 3} + 4.

Classe 1 (22,9%) : c'est la classe de l'observation analytique.

Les verbes ici sont des verbes de réflexion et d'observation : « *voir, penser, rendre compte, sentir, écouter, mettre* » + le verbe d'appréciation « *aimer* ». Les noms parlent de la maladie : « *médecin, maladie, idée, histoire, regard, cause, stimulation* ». Les adverbes sont intensifs : « *vraiment, trop* » et on relève des marqueurs explicatifs : « *c'est à dire, hein* ». On trouve également des marques de restriction : « *ne, non, plus, que* ». La personne 3 est dominante : « *lui, leur* ».

Résultat : La classe 1 est au centre de l'échange avec le patient, c'est la classe de l'observation, le soignant qui analyse son patient.

Classe 2 (20,3%) : le bilan.

Les verbes sont très nombreux dans cette classe, on y trouve des verbes de communication comme : « *demander, comprendre, venir, parler, expliquer, attendre, souhaiter, commencer, vouloir* ». Assez peu de lexèmes avec seulement : « *opération, heure, grand-père, air* » qui ne sont pas très significatifs. Les participes présents à valeur adjectivale : « *impressionnant, intéressant* » ponctuent l'échange. Les adverbes marquent l'échange

et l'approbation : « *bien, encore, après, donc, oui, très* ». Les pronoms de la personne 1 sont dominants : « *moi, j(e), me* ».

Résultat : La classe 2 stabilise le cadre de l'analyse, l'échange dans la consultation.

Classe 3 (25,6%) : l'aide au questionnement, la remise en question.

Les verbes de cette classe sont des verbes de l'existentiel et de l'interrogation : « *aider (y ?) arriver, vivre, rester, poser, intéresser, interroger, essayer, prendre (le temps), passer, trouver, exister, répondre, partir, connaître, perdre* ». Les lexèmes parlent du questionnement du diagnostic médical : « *question, coup, côté, dossier, médical, envie, fois, homme, face, psychanalyste, diagnostic, médicament* ». Les adverbes de manière et de temps : « *évidemment actuellement éventuellement maintenant* » sont nombreux et accompagnés du marqueur d'atténuation : « *un petit peu* ». Les pronoms sont réflexifs : « *vous, vous-même, se (le réfléchi)* ».

Résultat : cette classe du discours esquisse le *modus operandi* du psychanalyste : poser et faire se poser les questions qui aident. Le discours du psychanalyste est réflexif et stabilise différentes thématiques tournant autour du questionnement et de la réflexion, voire de la remise en question du patient.

Classe 4 (14,1%) : la famille.

Cette classe illustre le fait que la famille n'est jamais bien loin du patient, avec un lexique très fourni pour une si petite classe. On y trouve « *enfant père rapport grand-mère souvenir sœur maternel maman mère âge relation époux parent* » + le mot « *victime* ». Quelques verbes se dégagent : « *entendre, laisser, accepter, travailler, devenir* ». Ils peuvent marquer le « *travail* » que le patient est susceptible de faire sur le lien familial. Les déterminants possessifs de la personne 5 désignent le patient : « *votre, vos* » que l'on trouve ici couplés aux pronoms de personne 3 : « *elle, ils* » qui peuvent désigner le reste de la famille, les autres. La préposition « avec » met en relation, et la préposition « chez » désigne un lieu, peut-être lié à la notion du foyer.

Résultat : c'est la classe du discours sur la famille, les liens familiaux et les relations. C'est peut-être aussi celle du regard de la famille sur le patient (« *victime* »).

Hypothèse univers 2 : c'est le discours de la réflexion, l'interrogation de soi, le psychanalyste qui accompagne le patient dans la maïeutique de son auto-analyse.

2.4.2 Interprétation : l'analyse au cœur de la consultation du psychanalyste

L'AFC montre que les différentes classes sont plutôt mélangées, mises à part celle de la famille et celle de l'analyse psy qui se singularisent et s'isolent [fig9].

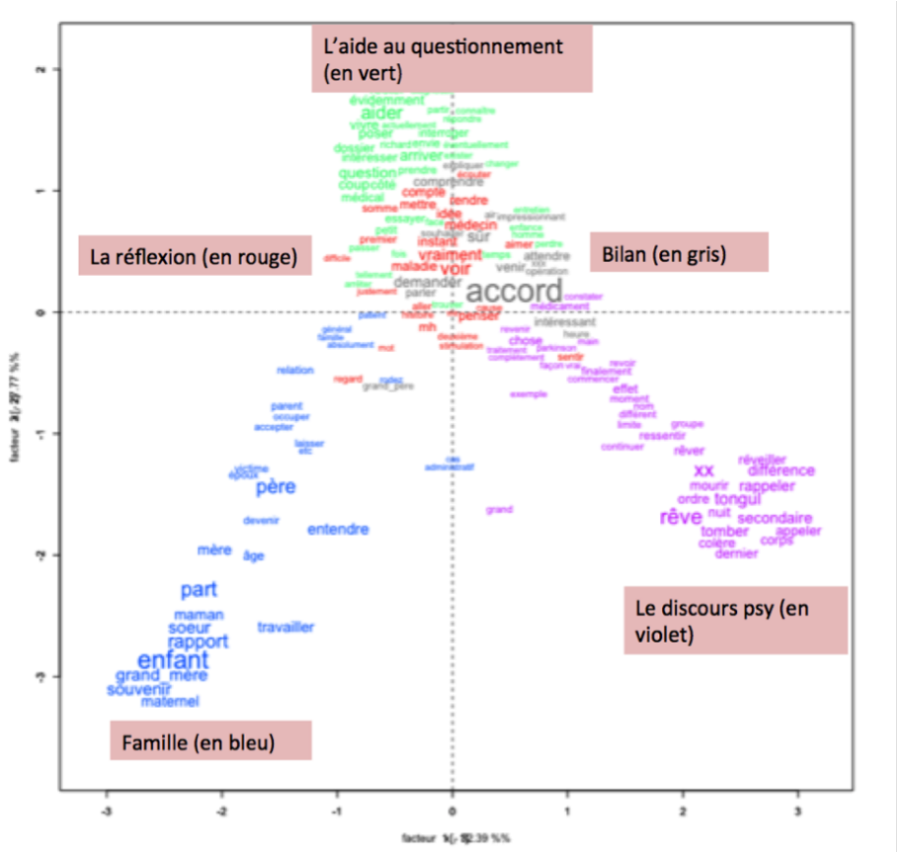


FIGURE 9 – Analyse factorielle corpus psychanalystes.

Les classes 1, 2 et 3 sont intriquées, comme emmêlées. C'est le noyau central de la séance du psychanalyste mettant en scène le questionnement du spécialiste et l'autoréflexion du patient sur fond de bilan personnel communiqué.

3 Discussion et perspectives

L'étude suivie du corpus permet de dégager des hypothèses interprétatives, et de profiler des éléments de réflexion quant à la formation en santé. Nous proposons d'abord une synthèse, puis discutons des hypothèses interprétatives que nous avons posées.

3.1 Synthèse

Les analyses lexicométriques successives proposées mettent en valeur les objets de discours spécifiques au corpus DECLICS2016 et cernent les spécificités du discours de chacun des locuteurs-type : médecin, psychanalyste, patient. Le tableau suivant [tab1] propose une synthèse.

FIGURE 10 – Synthèse des analyses menées sur le corpus DECLICS2016.

	Corpus global	Corpus patients	Corpus médecins	Corpus psychanalystes
Univers 1	La maladie chronique La maladie et de son traitement, le discours y est représenté comme un discours explicatif, justificatif.	Le contexte social du patient La vie à l'extérieur de l'hôpital.	Le diagnostic Le temps de parole du médecin consacré à l'échange avec le patient. Le médecin prend le temps de parler avec le patient, d'expliquer le bilan clinique. Cet échange est marqué par une certaine distance : le discours est technique, désincarné, le médecin est le clinicien et il expose son analyse de façon objective, savante.	L'analyse La matière du psychanalyste.
Univers 2	L'entretien clinique Le lien entre la famille, l'histoire familiale et la planification médicale. Le patient est très présent, actif, dans ces deux classes, alors qu'ailleurs le patient semble désincarné, il est le réceptacle de la consultation clinique.	La sphère médicale du patient Des troubles à la consultation, du diagnostic posé au traitement.	Le rendez-vous médical Le médecin en action. Dans cette classe, le discours du médecin impose moins de distance par rapport au patient, c'est celle de l'expérimentation technique, celle du terrain.	Le discours de la réflexion L'interrogation de soi, le psychanalyste qui accompagne le patient dans la maïeutique de son auto-analyse.

Le discours du patient est clairement cloisonné en deux pôles distincts : il y a (le discours de) l'hôpital et (le discours de) la vie en dehors de l'hôpital. Il semble que ces deux pôles ne se mélangent pas.

Le discours du médecin est un discours prédéterminé en fonction d'objectifs précis (*le diagnostic, le traitement, le prochain rendez-vous*) qui requièrent d'être efficace dans un temps limité. L'analyse montre qu'il suit un schéma discursif stable. Le médecin est le savant mais aussi le « technicien » du corps ; en somme, c'est le discours du spécialiste clinicien en action.

Le psychanalyste est également un spécialiste mais il mène son entretien différemment. Les classes qui apparaissent dans le discours sont plus nombreuses, sous-spécifiant le discours propre à la psychanalyse.

Si chaque locuteur-type (patient, médecin, psychanalyste) présente des objets de discours différents, il s'avère que patients et médecins présentent des classes de discours similaires, bien que la répartition soit différente selon leur statut. Le discours du psychanalyste reste quant à lui très distinctif : il faudra des analyses complémentaires pour cerner comment il est, même partiellement, intégré dans le discours du patient.

3.2 Formation en santé publique et perspectives pour la recherche

Au niveau de la formation des professionnels de santé, cette recherche pourrait se situer dans le développement ultérieur d'une didactique de l'interaction médecin/patient. Plusieurs scénarios seraient exploitables. Chaque médecin, confronté aux familles lexicales qui les représentent, peut se voir proposer de réagir, vérifiant, jusqu'à quel point il partage ce portrait dressé à partir du lexique employé. Une extraction de séquences assez révélatrices des emplois majeurs repérés grâce à notre classification descendante et présenté à plusieurs médecins en hétéro-confrontation vérifierait comment ils appréhendent leur posture professionnelle. Les discours de patients sont aussi émaillés de spécificités (syntaxiques et lexicales) sur lesquelles le professionnel de santé peut avoir intérêt à réfléchir : le médecin s'appuie-t-il sur une représentation mentale proche ou distante de ce qui est là mis à jour ?

Si la rencontre du professionnel de santé et du patient appelle impérativement la maîtrise de spécificités discursives, que nous avons commencé à mettre au jour par le biais de nos analyses, comment forme-t-on un locuteur compétent dans les formations médicales ? S'il appartient au locuteur expert (le professionnel de santé) de maîtriser les spécificités interactionnelles et conversationnelles de l'échange, peut-il à la lecture des analyses, progresser dans son expertise ? Que pense le médecin de la proximité qu'il entretient avec le patient, face au discours du psychanalyste, semblant plus décalé ? Que pense réellement partager le médecin avec son patient ?

Enfin quelques curiosités révélées par notre étude pourraient donner lieu à des échanges avec les soignants : ainsi la figure 7, du focus médecin pourrait permettre d'interroger les médecins sur la non perméabilité entre bilan clinique et douleur du patient, voire sur une meilleure façon de concilier les quatre foyers de préoccupations (bilan clinique, douleur, rendez-vous et traitement) au sein d'une consultation *idéale*.

Au plan de la recherche, la prochaine étape de notre étude sera notamment de comparer le discours du médecin à celui du psychanalyste en dépassant l'analyse lexicométrique pour mieux intégrer la dimension interactive, manifestement différenciée, avec le patient. A l'évidence, les mécanismes d'écoute diffèrent et les approches du médecin et du psychanalyste n'usent pas des mêmes procédés. Il faudra mettre au jour de nouveaux indicateurs pour caractériser plus avant chaque type de locuteur. Nous serons alors en mesure de mettre plus en évidence ce qui marque l'appropriation du discours médical par le patient, ou encore de pointer les possibles situations langagières qui impliquent des mécanismes de protection de la face (« Face Threatening Acts », Goffman, 1974) pour vérifier qui se protège de quoi exactement. Nous pourrions également étudier, en situation, la notion de « politeness » (Brown et Levinson, 1987) pour vérifier jusqu'à quel point l'entretien

médical reste ancré dans un cadre, un scénario et investit peu alors le détour ou la variation. Ces deux notions sont importantes pour comprendre les enjeux de ces échanges, à savoir : ménager sa propre face en tant que locuteur et ménager celle d'autrui en tant qu'interlocuteur.

4 Conclusion

L'analyse linguistique du corpus DECLICS2016 met au jour des régularités comme des irrégularités saillantes dans les discours des différents interactants. Elle permet également de faire des hypothèses sur la circularité de ces discours, qui s'interpénètrent plus ou moins selon les acteurs. Ainsi, nous constatons que, tant chez le patient que chez le médecin, les discours sont le plus souvent cloisonnés. Au contraire, une dynamique circulaire semble être à l'œuvre dans le discours du psychanalyste. La parole gagnerait-elle à s'émanciper de ces barrières pour créer des liens entre les différentes thématiques ? Les analyses de corpus médicaux menées par le biais de logiciels lexicométriques peuvent permettre aux futurs professionnels de santé de s'initier à l'observation de ce que se dit très précisément dans ces échanges. Ce type d'observations peut se révéler très riche dans la pratique professionnelle en santé, tant pour faire un retour de ce qui se passe lors de ces entretiens spécialisés que pour améliorer la fluidité des consultations cliniques. Il conviendra de poursuivre plus avant notre recherche dans le cadre du projet DECLICS afin de mieux s'ajuster aux besoins spécifiques des professionnels comme des patients.

Nous remercions ici tous les relecteurs attentifs de cet article : nos relecteurs anonymes, ainsi qu'Emmanuèle Auriac-Slusarczyk et Mylène Blasco, pour leurs précieuses suggestions et leurs conseils avisés qui ont contribué à la rédaction de cet article

Références bibliographiques

Chauveau Thoumelin, P., et Grabar, N. (2014). La subjectivité dans le discours médical : sur les traces de l'incertitude et des émotions, *EGC, vol. RNTI-E-26*, 455-466.

Cori, M, David, S. et Léon, J. (2008). Construction des faits en linguistique : la place des corpus, *Langage,* 171.

Brown, P. et Lévinson, S. (1987). *Politeness : some universals in language usage,* London : Cambridge University Press.

Drouet, G. (2019). *La contradiction en discours, une mise en scène énonciative singulière.* Paris : Classiques Garnier.

Drouet, G., Richard, E. et Moreau, F. (2017). *A cross-over medical-sociological-linguistic study : when discourse analysis supports medical researches*. Communication présentée au DiscourseNet Congress #2, Interdisciplinary Discourse Studies, Theory and Practice #DNC2, Université de Warwick.

Goffman, E. (1974). *Les Rites d'interaction*, Paris : Les Éditions de Minuit.

Haxaire, C., Richard, E., Bodénez, P. et Locquet, C. (2010). Appréhension et gestion des traitements psychotropes par les généralistes (Bretagne Occidentale), *Revue Sociologie Santé,* 31, 331-348.

Haxaire, C., Richard, E., Dumitru-Lahaye, C., Genest, Ph., Bodénez, P. et Bail Ph. (2005). *Représentations de la santé mentale et de la souffrance psychique par les médecins généralistes (de Finistère et des Côtes d'Armor)*, Rapport MIRE/DRESS, convention MIRE 22/02.

Haxaire, C., Bodénez, P., Richard, E., Terrien, K., Bail, Ph., Locquet, C. et Noailly, M. (2003). *Gestion différentielle des dépendances par les médecins généralistes de Bretagne occidentale*, Rapport MILT / INSERM / CNRS, projet 99D 4.

Ingrand, I., Gaussot, L., Richard, E., Drouet, G., Moreau, F. et Ingrand, P. (sous presse). Information sur le risque familial de cancer colorectal, regards croisés sur l'efficacité d'une intervention personnalisée, *Revue Santé publique.*

Ingrand, I., Gaussot, L. et Richard, E. (éds) (2018). *Pour une approche personnalisée du dépistage familial du cancer colorectal. Une recherche interdisciplinaire : sciences médicales, sociales et humaines*. Limoges : Presse Universitaire de Limoges.

Kerbrat-Orecchioni, C. (2005). *Le discours en interaction*. Paris : Colin.

Maingueneau, D. (2012). *Les phrases sans texte*. Paris : Colin.

Le Bart L. et Salem, A. (1988). *Analyse Statistique des données textuelles*. Paris : Bordas.

Le Bot M.-C., Schuwer M. et Richard, E. (2008). *La reformulation. Marqueurs linguistiques – Stratégies énonciatives*. Rennes : Presses Universitaires de Rennes.

Schuwer, M., Le Bot, M.-C. et Richard, E. (2008). *Pragmatique de la reformulation – Types de discours – Interactions didactiques*. Rennes : Presses Universitaires de Rennes.

Léon, J. (2008). Aux sources de la « Corpus Linguistics » : Firth et la London School, *Langage,* 171, 12-33.

Rabatel, A. (2016). Diversité des points de vue et mobilité empathique. *In* M. Colas-Blaise, L. Perrin, et G.M. Tore, *Diversité des points de vue et mobilité empathique* (pp. 135-150). Limoges : Ed. Lambert-Lucas.

Rabatel, A. (2005). La part de l'énonciateur dans la construction interactionnelle des points de vue, *Marges linguistiques,* 9, 115-136.

Ratinaud, P. et Marchand, P. (2012). Recherche improbable d'une homogène diversité : le débat sur l'identité nationale, *Langages,* 187, 93-107.

Real Del Sarte, O. (1984). Éléments pour une analyse du discours médical en prison, *Déviance et société, 8*(2), 167-180.

Reinert, M. (2007). Postures énonciatives et mondes lexicaux stabilisés en analyse statistique de discours, *Langage et société,* 121-122, 189-202.

Reinert, M. (2004). La méthode d'analyse exploratoire des données textuelles "Alceste" et le problème de l'analyse de contenu, *Les Cahiers de Jérico-st*, 4, 79-90.

Reinert, M. (1993). Les "mondes lexicaux" et leur "logique" à travers l'analyse statistique d'un corpus de récits de cauchemars, *Langage et société*, 66, 5-39.

Reinert, M. (1983). Une méthode de classification descendante hiérarchique : application à l'analyse lexicale par contexte, *Les Cahiers de l'Analyse des Données,* 4, 471-484.

Richard, E. (2018a). *Des organisations dynamiques de l'oral*. Berne : Peter Lang.

Richard, E. (2018b). Posture(s) en débat, débat de posture(s) dans le cadre pluridisciplinaire. Communication présentée au colloque *Hommage à Alain Rabatel*, *Les postures énonciatives : autour des propositions d'Alain Rabatel*, Toulon.

Richard, E. (2014). à propos de répétition : entre continuité et rupture, in Rabatel A. et Magri V. (éds) *La répétition lexicale : approche discursive et pragmatique*, *SEMEN*, 38, 95-114.

Richard, E. et Drouet, G. (2016). Confirmer pour mieux détourner : marqueurs d'acceptation et modalités de transition, *Testi e linguaggi, 10/216,* Salerno : Carocci editore, 131-142.

Richard, E. et Drouet, G. (2017). *En même temps* : un marqueur double ? De la simultanéité temporelle à la contradiction énonciative. *In* F. Lefeuvre et G. Dostie (éds), *Lexique, grammaire, discours. Les marqueurs discursifs* (pp. 159-172). Paris : Champion.

Richard É., Noailly M. (2014). Mécanique et dynamique de la répétition en discours : figures de rhétorique et usages spontanés. *In* J. Delaplace (éd.), *L'art de répéter* (pp. 271-281). Rennes : Presses Universitaires de Rennes.

Richard, E., Drouet, G., et Moreau, F. (2018). L'appropriation des discours. *In* I. Ingrand, L. Gaussot et E. Richard (éds), *Pour une approche personnalisée du dépistage familial du cancer colorectal. Une recherche interdisciplinaire : sciences médicales, sociales et humaines* (pp. 119-151). Limoges : Presse Universitaire de Limoges.

Traverso, V. (2001). Analyse de consultations médicales en présence d'un intermédiaire linguistique non professionnel, *actes du VIIIe congrès de l'ARIC*, Université de Genève, 24-28 septembre.

Zhai, C. et Massung, S. (2016). *Text Data Management and Analysis : a Practical Introduction to Information Retrieval and Text Mining*. New York : Morgan & Claypool.

Références sur le Web

Ratinaux P. (2009). « IRaMuTeQ : Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires ». [En ligne : http ://www.iramuteq.org. (05-09-2018)]

Reinert, M. (2000). La tresse du sens et la méthode "Alceste". Applications aux "Rêveries du promeneur solitaire", 5e Journées internationales d'Analyse Statistique des Données Textuelles, [en ligne : http ://lexicometrica.univ-paris3].

Vergely P., Condamines A., Fabre C., *et al.* (2009). Analyse linguistique des interactions Patient/Médecin, Actes éducatifs et de soins, entre éthique et gouvernance, *Actes du colloque international* (Felix C., Tardif J., éd.), Nice 4-5 juin 2009, [en ligne : http ://revel.unice.fr/symposia/actedusoin/index.html ?id=750. (05-09-2018)]