

Caractérisation des échanges entre patients et médecins : approche outillée d'un corpus de consultations médicales

Ludovic Tanguy, Cécile Fabre, Lydia-Mai Ho-Dac et
Josette Rebeyrolle
CLLE-ERSS Université de Toulouse & CNRS

Résumé : Nous présentons une étude fondée sur un corpus de transcriptions de consultations médicales, dans le cadre d'un projet interdisciplinaire qui explore la question des inégalités sociales de santé. L'objet de cet article est de montrer comment, en tant que linguistes familiers du traitement outillé des corpus, nous avons choisi d'aborder ce matériau qui fait l'objet de questionnements disciplinaires complémentaires, et quels éléments de caractérisation spécifiques nous sommes en mesure d'apporter en réponse à une demande émanant de la sphère médicale.

Abstract: This paper presents a study based on a corpus of transcribed doctor-patient interactions, which is part of an interdisciplinary project that addresses the issue of social inequalities in health. It presents the specificity of our approach compared to other ways of handling this data, and reports some results based on the automatic analysis of the corpus.

Mots clés : Consultations médicales transcrites, corpus d'interactions, linguistique outillée, interdisciplinarité

Key words: Transcribed medical encounters, corpus of interactions, automatic analysis of corpora, interdisciplinary study

Introduction

Nous présentons un travail mené à partir d'un corpus de transcriptions de consultations médicales, dans le cadre d'un projet interdisciplinaire relevant du champ de la santé publique.

Plus précisément, l'objectif du projet INTERMEDE¹ est de comprendre dans quelle mesure les interactions qui se produisent entre le patient et le médecin généraliste dans le cadre de la consultation sont le reflet de certaines inégalités sociales de santé. Il s'appuie sur un corpus de consultations, complété par des entretiens et des questionnaires auprès du patient et du médecin. Au terme de ce projet qui nous a amenés à interagir avec des épidémiologistes, des sociologues et des psychologues, nous proposons de montrer comment, en tant que linguistes familiers des méthodes de la linguistique de corpus, nous avons choisi d'aborder ce corpus de consultations qui fait l'objet de questionnements disciplinaires complémentaires. Il nous paraît en effet utile de cerner la contribution spécifique de la linguistique dans un projet où se rencontrent des pratiques scientifiques différentes qui interrogent ensemble des données langagières.

Afin de présenter la spécificité de notre approche, nous commençons par une brève synthèse des principaux types de travaux qui ont pris pour objet les consultations médicales, et nous présentons les orientations méthodologiques de nos partenaires dans le projet. C'est dans ce contexte que nous déterminons ensuite les particularités de notre approche de ce matériau. La deuxième section décrit le corpus de consultations et les variables que nous avons considérées. Nous présentons enfin les analyses que nous avons effectuées, ainsi que les principaux résultats et apports méthodologiques.

¹ Projet INTERMEDE (2008-2011), financé par l'Institut de Recherche en Santé Publique (IRESP), coordonné par Thierry Lang (CHU Toulouse-Purpan et Inserm U558), intitulé « Interactions médecin/patient en médecine générale et inégalités sociales de santé : analyses interdisciplinaires ». Outre l'Inserm, les participants au projet sont, à Toulouse, les laboratoires LISST CIEU (UMR 5193), SOI-PRISMH (EA 3690), LERASS (EA 827), CLLE-ERSS (UMR 5263), et à Nantes, le PIMESP (CHU Nantes).

1. Les corpus de consultations médicales : diversité des modes d'investigation

Depuis plusieurs décennies, le milieu médical s'est intéressé à la façon dont l'acte de soin peut être éclairé par la prise en compte des échanges entre un médecin et un patient. De nombreux travaux de recherche ont ainsi été basés sur le recueil et l'analyse de ces interactions, et traitent majoritairement des données en langue anglaise² (Ainsworth-Vaughn 2001). Ces études s'inscrivent généralement dans deux paradigmes qui se distinguent par la méthodologie mise en œuvre pour analyser ce matériau.

1.1. Approches quantitatives

On regroupe sous cette première catégorie deux types d'études qui se différencient essentiellement par la nature et la quantité d'information mobilisée.

Les premières cherchent à mettre en relation directe des caractéristiques langagières simples avec des variables situationnelles, en appliquant des tests statistiques. Des études, comme celle de Skelton & Hobbs (1999), concernent par exemple la variation du nombre de mots et de questions posées en fonction du sexe du médecin. D'autres études s'intéressent au partage de certains mots de vocabulaire, au rôle des interruptions ou des rires, etc. Un deuxième type de travaux, plus répandu et à plus large spectre, analyse le déroulement des consultations en se fondant sur un codage spécifique et systématique de chaque énoncé du médecin et du patient. Le modèle standard actuel de codage est le RIAS (*Roter Interaction Analysis System*) (Roter & McNeilis 2003). Plusieurs dizaines de catégories spécifiques aux consultations

² A notre connaissance, très peu de travaux ont été menés sur les consultations en langue française. Plus qu'une tradition scientifique, il faut sans doute voir là une conséquence des politiques publiques, de nombreuses études et campagnes de recueil de données se justifiant, dans certains pays, par un système centralisé et/ou la nécessité d'évaluer des réformes importantes de la pratique médicale.

médicales permettent de coder le contenu thématique, les actes de langage et en règle générale la posture interactionnelle de chaque locuteur. Les étiquettes disponibles sont par exemple : « expression de l'inquiétude », « encouragement, optimisme », « apport d'information sur le traitement », etc. Cette annotation exclusivement manuelle est ensuite exploitée par le biais de méthodes quantitatives visant à établir les liens entre la fréquence et les enchaînements de certaines catégories et les informations externes disponibles. Cette approche permet par exemple de mesurer l'impact de certaines attitudes interactionnelles sur la satisfaction et/ou la compréhension du patient.

Dans le cadre du projet INTERMEDE, deux de nos partenaires ont appliqué une démarche quantitative au corpus de consultations, selon deux modalités : une approche outillée et statistique centrée sur l'étude des classes lexicales dans la tradition lexicométrique (Marchand *et al.* 2009) et une analyse fondée sur un étiquetage manuel préalable de séquences thématiques (ex : alimentation, raison de la visite, plainte, etc.), permettant de calculer leur répartition et leur enchaînement dans la communication (Génolini *et al.* à paraître). Dans les deux cas, des corrélations sont ensuite calculées avec des variables externes (cf. 2.2). Nos partenaires épidémiologistes, spécialistes de la quantification, n'ont par contre pas utilisé les consultations qui ne sont pas considérées comme directement exploitables par leurs méthodes. Ils ont donc exploité les questionnaires réalisés à l'issue de consultations (degré de concordance dans les réponses entre le médecin et le patient, degré de satisfaction de chaque acteur à l'issue de la consultation, etc.) (Schieber *et al.* à paraître).

1.2. Approches qualitatives

A l'opposé de ces approches statistiques qui ont tendance à s'éloigner rapidement du matériau langagier lui-même (comme le souligne notamment Waitzkin (1990)), le second type d'études des consultations s'inspire davantage des pratiques

propres aux sciences humaines et sociales. Heritage & Meynard (2006) les regroupent sous le terme de « micro-analyse », et les situent dans la sphère disciplinaire de l'anthropologie et de la sociologie. Ces études abordent les consultations de manière contextualisée, en les replaçant en particulier dans le cadre de l'histoire médicale du patient. Seuls certains passages de l'interaction sont mis en exergue, servant d'exemples pour illustrer telle ou telle attitude. Les objectifs de ces études concernent des questions comme les relations de pouvoir entre médecin et patient, la négociation, l'expression de la subjectivité, etc. Dans ce même cadre de l'analyse conversationnelle, le sociologue Have (1989) a cherché, par exemple, à formaliser les aspects rituels de la consultation, notamment en identifiant les différentes phases typiques de son déroulement.

Dans le cadre du projet INTERMEDE, des études qualitatives ont été menées par un des groupes de sociologues, incluant principalement l'observation *in situ* des consultations ainsi que l'analyse des entretiens semi-directifs, et débouchant sur la définition de profils de médecins (postures soignantes), et de patients (autonomes ou dépendants).

1.3. Spécificités de notre approche

Dans le contexte que nous venons de dessiner (contexte général des études sur les consultations médicales, contexte propre au projet), quelle peut être la contribution spécifique d'une équipe de linguistes s'inscrivant dans le champ de l'analyse de corpus et du traitement automatisé des données langagières ? Nous avons eu à relever un double défi. Tout d'abord, nous devons trouver les moyens de répondre à une question émanant de la sphère médicale, que l'on peut reformuler ainsi : peut-on constater des disparités dans les interactions, qui puissent être mises en rapport avec les variables sociales qui renseignent le corpus ? Cette question ouverte nous laissait libres de la méthode et des modes d'approche du corpus, et notamment des marques linguistiques à considérer pour caractériser

l'interaction. Ensuite, nous voulions mettre en œuvre une démarche complémentaire par rapport à celles adoptées par les autres partenaires, c'est-à-dire à la fois différente et permettant d'engager une collaboration. Nous nous sommes ainsi efforcés de proposer une approche facilitatrice de dialogue, en travaillant sur les modalités d'accès aux données langagières (visualisation, exploration) et sur l'intégration des retours / hypothèses des autres partenaires.

Nous avons dégagé au cours du projet quatre caractéristiques de notre approche.

1) L'opposition entre méthodes quantitatives et méthodes qualitatives nous a semblé réductrice pour une étude linguistique de ce corpus. Sans vouloir prétendre réconcilier les deux démarches (comme Roter & Mc Neilis (2003) invitent par exemple à le faire), il nous a semblé possible à la fois de mener une approche quantifiée des descripteurs linguistiques que nous mettons au jour (cf. 3.1), et de proposer des méthodes facilitant l'exploration du corpus et ainsi le repérage et l'examen de passages particuliers (cf. 3.2).

2) Nous restons au plus près de la matérialité langagière des consultations. C'est ce qui doit faire la spécificité de l'approche linguistique des données langagières par rapport aux investigations menées par des chercheurs d'autres disciplines. Nos variables correspondent à des caractéristiques linguistiques formelles de l'échange : nous nous gardons d'ajouter une couche d'annotation qui chercherait à interpréter et codifier le contenu des discours, réduisant la complexité du matériau langagier.

3) Nous avons cherché à travailler sur des descripteurs linguistiques variés permettant de cerner au mieux la dimension interactionnelle du corpus. L'étude de la répartition lexicale donnant plutôt accès à des informations de nature thématique, nous avons opté pour un autre type d'exploration, de manière à construire une gamme de traits plus adéquats pour rendre compte du comportement verbal – interactionnel et énonciatif – des deux acteurs. Cette tâche d'élaboration des variables

spécifiques à l'étude a été guidée par la contrainte de disposer de traits faciles à repérer automatiquement pour les projeter de manière systématique sur l'ensemble du corpus. Cette phase a donné lieu à de nombreux tâtonnements – nous avons provisoirement renoncé à calculer certains traits peu productifs ou difficiles à repérer (comme la présence de termes spécialisés ou de marques de reformulation), ou regroupé des traits fortement corrélés. La constitution de traits linguistiques pertinents pour l'étude nous semble constituer un aspect central de ce qui fait la spécificité de notre démarche.

4) Concrètement, notre approche s'est effectuée en deux temps, suivant une méthodologie qui s'inscrit dans la lignée de celle de D. Biber (1988). Tout d'abord, nous avons pratiqué un marquage du corpus, sur la base de procédés automatiques ou semi-automatiques, pour annoter l'ensemble des caractéristiques langagières établies. Dans un second temps, l'analyse de ces données et de leurs variations dans le corpus nous a permis de faire émerger des configurations, autrement dit de proposer une typologie des consultations, puis de corrélérer celles-ci aux caractérisations sociales disponibles.

2. Données de l'étude et définition des variables

2.1. Le corpus

Le matériau sur lequel nous avons travaillé comprend 79 consultations enregistrées, correspondant chacune à un patient différent et concernant au total 11 médecins généralistes. Parmi ces consultations, 13 n'ont pu être exploitées pour une ou plusieurs raisons (truncation de l'enregistrement, perturbation majeure de l'entretien par un événement extérieur, consultation impliquant plus d'un patient, etc.). Si nous n'avons pas été impliqués dans la phase de collecte, nous avons pu néanmoins intervenir sur la phase de transcription de manière à disposer d'un corpus exploitable pour des analyses automatiques. Nous avons également procédé à un marquage XML du corpus inspiré de la TEI-P5.

La figure 1 montre un extrait d'une consultation. La transcription est orthographique et normalisée. Y sont indiqués : les tours de parole (<u>) avec des attributs indiquant le locuteur (who, M signifiant ici le médecin et P le patient) et le numéro de tour (n), les commentaires (<comment>) les pauses (<pause> avec deux versions, courte ou longue) et les chevauchements (<overlap>). A partir des données ainsi retranscrites et annotées, nous avons effectué une segmentation en mots, une lemmatisation et un étiquetage morpho-syntaxique à l'aide d'une version modifiée du programme TreeTagger.

```

<u who="M" n="96">allez-y respirez fort <comment>téléphone
sonne</comment> excusez-moi</u>
<comment>médecin répond au téléphone</comment>
<pause length="long"/>
<u who="M" n="97">respirez bien profondément <pause
length="long"/> et la glycémie ? vous l'avez faite ?</u>
<u who="P" n="98">oui</u>
<u who="M" n="99">alors ?</u>
<u who="P" n="100">1,40</u>
<u who="M" n="101">vous dépassez pas ?</u>
<u who="P" n="102">non et là j'avais 1,13 là</u>
<u who="M" n="103">même avec les <overlap type="begin"
n="4">bons repas, même avec les bons repas</overlap></u>
<u who="P" n="104"><overlap type="end" n="4">dépasser oui,
un petit peu ça m'arrive oui</overlap> mais bon en moyenne
hein</u>
<u who="M" n="105"> et après les bons repas vous le faites
aussi ? les lendemains de bons repas ?</u>
<comment>rires</comment>

```

Figure 1 : Extrait de consultation au format XML

2.2. Les variables de l'étude

Une fois les données annotées, nous avons calculé (à l'aide de programmes dédiés) un ensemble de mesures globales permettant de caractériser au mieux l'interaction médecin / patient. Ces variables sont décrites ci-dessous et identifiées par le code utilisé lors des analyses (section 3). Nous

présentons les variables linguistiques issues de ces données³, puis les variables externes obtenues par d'autres sources.

a) Variables linguistiques

Les trois premières variables cherchent à capter la taille de la consultation et le partage de la parole entre les protagonistes.

[N_Mots] : cette variable compte le nombre de mots de chaque consultation. Cette mesure prend en compte tout signe (mot, mais également ponctuation, interjection) prononcé par le médecin et le patient. Les 66 consultations transcrites comportent entre 795 et 7 717 mots, avec une moyenne de 2 636 (écart-type 1 350).

[Parole_M] et **[Parole_P]** : ces deux variables symétriques mesurent le taux de parole par locuteur (M pour le médecin, P pour le patient). Celui-ci est calculé en considérant le nombre de mots prononcés par chaque locuteur par rapport au nombre total de mots de la consultation. La parole du médecin couvre entre 13% et 90% de la consultation (moyenne de 57%). De façon générale, c'est donc le médecin qui parle le plus dans le corpus étudié.

Les trois variables suivantes s'intéressent au partage du vocabulaire entre le médecin et le patient.

[Voc_Com] : cette variable mesure la part de vocabulaire commun aux deux interlocuteurs d'une consultation. Le vocabulaire commun est constitué des lemmes des noms, verbes et adjectifs qui apparaissent à la fois dans les énoncés du médecin et dans ceux du patient sans toutefois appartenir à ce qu'on appelle le « vocabulaire banal », c'est-à-dire les mots qui apparaissent dans plus de 50% des consultations. En moyenne, 9% des lemmes d'une consultation relèvent de ce vocabulaire commun.

³ Certaines caractéristiques, bien que disponibles dans les transcriptions, n'ont pas été exploitées.

[Voc_M] et **[Voc_P]** : ces deux variables mesurent la part de vocabulaire propre au médecin ou au patient. Lorsqu'un lemme non banal apparaît exclusivement dans les énoncés de l'un des interlocuteurs, il est considéré comme *spécifique* à cet interlocuteur. En moyenne, la part de vocabulaire spécifique au médecin est de 26% contre 42% en ce qui concerne le vocabulaire spécifique au patient.

Les deux couples de variables suivants visent chacun des types d'énoncés particuliers.

[Question_M] et **[Question_P]** : proportion de questions dans une consultation (rapportée au nombre d'énoncés), par locuteur. Elles sont repérées de façon superficielle sur la base de la présence de points d'interrogation dans la retranscription. Ce nombre est globalement plus élevé pour les médecins (23%) que pour les patients (5%).

[Repet_M] et **[Repet_P]** : proportion d'énoncés de répétition par locuteur. Dans un énoncé de répétition, au moins 20% des lemmes de type adjectif, nom ou verbe apparaissent également dans l'énoncé précédent⁴. Au total, environ 13% des énoncés du corpus sont de ce type pour les deux locuteurs.

Pour terminer, nous avons projeté sur le texte une série de lexiques spécifiques dans le but d'évaluer la subjectivité des énoncés, en nous inspirant de la typologie de Kerbrat-Orecchioni (1980). Après plusieurs tentatives, nous avons opté pour une distinction entre deux classes de marqueurs.

[Subj_M] et **[Subj_P]** : pour chaque locuteur, la proportion de termes affectifs (*heureux, insupportable, trouille...*), évaluatifs (*chaud, costaud, mini...*) ou axiologiques (*bête, moche, positiver...*).

⁴ Par exemple, cet échange articule deux énoncés de répétition :

Médecin : c'est lourd

Patient : ben oui c'est lourd

Médecin : bien sûr c'est lourd, bien sûr c'est lourd

[**Modal_M**] et [**Modal_P**] : pour chaque locuteur, la proportion de termes modalisateurs (*peut-être, certainement, sembler...*).

b) Variables sociales

En plus de ces variables linguistiques, nous disposons, grâce aux questionnaires remplis par les patients et les médecins, d'un ensemble d'informations, résumées ci-dessous.

[**Age_M**], [**Age_P**], [**Sexe_M**] et [**Sexe_P**] : l'âge et le sexe du médecin et du patient. L'âge des patients varie de 17 à 85 ans (médiane 54), celui des médecins de 36 à 57 (médiane 54). 32 patients sur 66 sont des femmes, 18 des consultations impliquent des médecins femmes.

[**CSP**] : une version binarisée de la catégorie socio-professionnelle du patient. Les catégories initialement renseignées (utilisant la nomenclature de l'INSEE) montraient une trop grande dispersion et un déséquilibre important. Nos collègues sociologues ont proposé une distinction binaire sur la base des entretiens menés auprès des patients⁵. Sont ainsi classés négativement (notés CSP- par la suite) sur cette variable les patients disposant de faibles revenus, ayant un faible sentiment d'emprise sur le monde. Il a été observé de plus que ces patients sont peu soucieux de leur santé et ont des attentes minimales envers leur médecin.

La dernière variable rend compte de la place que chacune des thématiques identifiées par nos partenaires occupe dans les consultations.

[**Type_consultation**] : cette variable à trois modalités - *écoute clinique* (17%) ; *hygiène de vie* (31%) ; *traitement médical* (51%) - a été établie à partir de l'analyse des thématiques abordées au cours de la consultation, et de la répartition séquentielle de celles-ci. Cette variable a l'avantage

⁵ Seule une partie des patients ont pu être ainsi qualifiés, les autres consultations ont reçu une valeur indéfinie.

de proposer une classification assez transparente des grands types de consultation médicale.

3. Caractérisation des consultations

Nous proposons ici deux types d'analyses des données recueillies et annotées. La première est une méthode quantitative classique permettant d'étudier les corrélations entre les caractéristiques linguistiques des consultations et les variables externes, et de proposer une première typologie. La seconde, plus exploratoire, propose d'envisager la complexité de la consultation, en étudiant les paramètres de l'interaction au fil de son déroulement, de manière plus qualitative.

3.1 Analyse multidimensionnelle

L'annotation automatique des variables linguistiques et externes fournit un ensemble de données décrivant 66 individus au moyen de 14 variables linguistiques et 6 variables externes. Les variables linguistiques constituent autant de traits dont on peut observer les relations par le biais d'une analyse en composantes principales (ACP). Dans un premier temps, l'objectif est de faire émerger les dimensions langagières qui décrivent au mieux la variation observée dans le corpus. Ces dimensions sont obtenues à partir du regroupement statistique des variables linguistiques. Dans un second temps, les variables externes sont projetées pour observer leur corrélation avec les dimensions mises au jour, et donc leur interaction avec les variables linguistiques. Les deux premières composantes principales captant plus de 47% de la variance totale, nous nous limitons à celles-ci pour notre interprétation. La figure 2 montre la carte factorielle positionnant les variables, qu'elles soient linguistiques (flèches noires) ou externes (en italiques bleues).

La première dimension, horizontale (33,08% de la variance), est la dimension du locuteur. Elle oppose les interactions médicales en fonction de la répartition de la parole de chacun des deux locuteurs : on peut isoler les interactions dans lesquelles la parole du médecin domine (Parole_M, à gauche dans la figure) de celles où le nombre de mots prononcés par le patient est plus élevé (Parole_P, à droite). Du

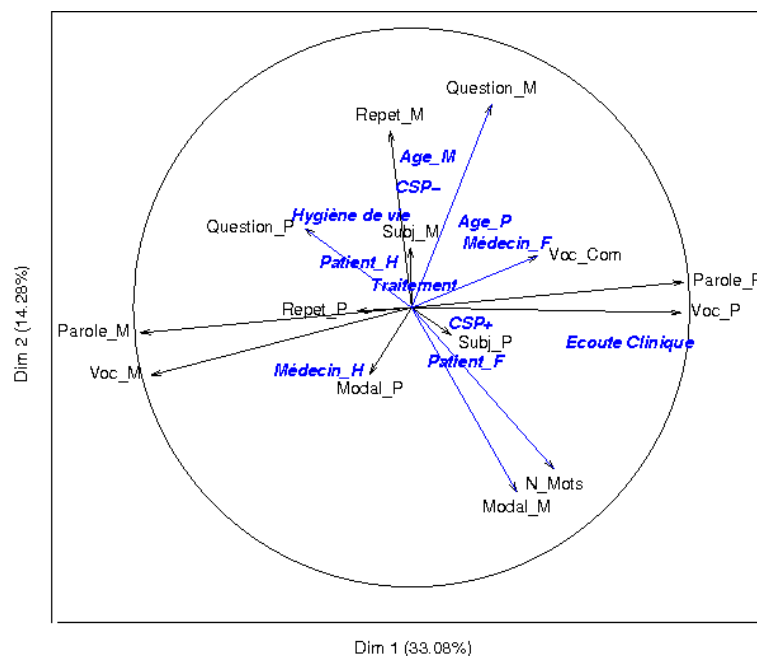


Figure 2 : Carte factorielle de l'ACP

point de vue des variables externes, les cas où la parole du patient est majoritaire correspondent nettement à des consultations de type « Ecoute clinique », et sont généralement menées par des médecins femmes. Ces deux situations opposées se caractérisent symétriquement par une très forte part de vocabulaire spécifique au locuteur majoritaire (Voc_P, Voc_M). On ne retrouve cependant pas de symétrie sur les autres aspects langagiers : les consultations où le patient

s'exprime plus sont plus longues (N_Mots), et voient le vocabulaire des deux interlocuteurs mieux partagé (Voc_Com). De l'autre côté, celles où le médecin est le locuteur majoritaire sont également celles où le patient pose plus de questions (Questions_P), et produit une part plus importante d'énoncés de répétition (Repet_P). Cette première dimension suggère donc l'existence de deux pôles, de deux types d'attitude interactionnelle, sans pour autant permettre de dégager un lien avec des paramètres sociaux liés au patient.

La deuxième dimension (14,28%) est de nature principalement énonciative et semble cette fois suggérer une corrélation entre certaines attitudes énonciatives et certaines catégories de patients. Elle oppose les médecins qui recourent fortement à un vocabulaire subjectif (Subj_M, en haut) à ceux qui modalisent leur discours (Modal_M, en bas). Ceux dont les tours de parole laissent une grande place à la subjectivité sont aussi ceux qui posent beaucoup de questions (Questions_M) à leurs patients et qui répètent leurs énoncés (Repet_M). Dans le premier cas, il s'agit de médecins plus âgés, s'adressant à des patients également plus âgés et issus de couches sociales plus populaires. Cette dimension semble donc opposer deux styles d'interaction de la part du médecin, selon que celui-ci mise plutôt sur l'usage d'un vocabulaire affectif et évaluatif, ou qu'il s'exprime de manière plus nuancée. Son identification constitue dès lors une piste de réponse à la question des variations corrélées socialement dans les consultations médicales.

3.2 Profil de l'interaction et déroulement de la consultation

Cette première typologie des caractéristiques des interactions demande à être complétée par des observations plus précises. En effet, l'analyse multidimensionnelle ne fournit qu'une description partielle du corpus, dans la mesure où elle est fondée sur des caractéristiques synthétiques qui résument en quelques mesures un matériau langagier sans permettre d'en saisir toute la richesse et la complexité. Ce constat est celui de toutes les approches quantitatives de l'analyse des textes, au

premier rang desquelles se trouve la lexicométrie, qui, sous sa forme traditionnelle, résume par des fréquences globales des distributions lexicales parfois très disparates d'un point à un autre du texte. La notion émergente de topologie textuelle (Mellet et Barthélémy 2007, Mayaffre 2007) cherche à répondre à ce problème en se dotant de moyens d'analyse et de représentation à l'échelle du texte. C'est sous cet angle que nous avons choisi de compléter notre étude des variations de l'interaction dans notre corpus. Plus précisément, notre objectif est de fournir une représentation de chacune des consultations médicales qui permette non seulement d'observer simultanément les principaux traits linguistiques de l'interaction patient-médecin mais aussi de visualiser l'évolution de ces paramètres au fil d'une consultation.

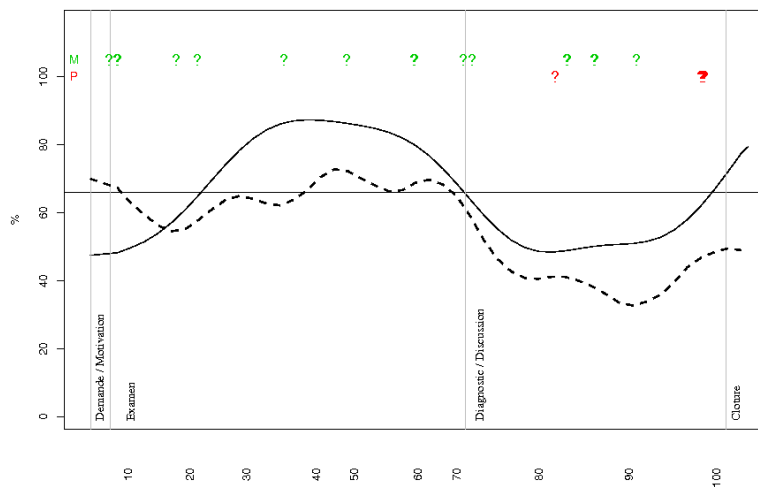


Figure 3 : Profil de consultation

La figure 3 montre ainsi (courbe continue) l'évolution du taux de parole du médecin (Parole_M), c'est-à-dire la proportion de mots prononcés par lui dans une fenêtre glissante de 50 mots, avec un lissage linéaire. La moyenne sur l'ensemble de la consultation est représentée par le trait horizontal. Les marques sur l'axe des abscisses sont les

numéros de tours de parole. Cette courbe, ici très simple, indique une phase de parole accrue du médecin autour du premier tiers de la consultation, ainsi que vers la fin. La courbe en pointillé mesure la fréquence relative des changements de tour de parole : une valeur élevée indique un échange plus rapide, et des énoncés généralement plus courts, une valeur basse correspond à des passages plus monologiques. On peut voir que cette courbe suit globalement la première, et indique donc que les passages où le médecin parle le plus (première moitié) sont ceux qui correspondent à une interaction plus rapide (avec des énoncés courts de la part du patient, essentiellement des réponses aux questions). A l'inverse, le médecin semble moins interrompre son patient en fin de consultation, lorsque celui-ci devient le locuteur majoritaire.

En complément de ces courbes, les occurrences des questions posées par les interlocuteurs sont indiquées en haut du graphique (en vert, ligne supérieure, pour le médecin, en rouge, ligne inférieure, pour le patient). Afin d'approfondir l'interprétation de ces représentations, nous avons parallèlement segmenté les consultations (sur la base d'une lecture/écoute au fil du texte) pour dégager les moments typiques de leur déroulement (découpage en phases fondé sur le modèle de Have 1989), à savoir : *introduction – expression de la demande et motivation de la visite – examen – diagnostic et discussion – clôture*. Le résultat de ce découpage pour la consultation est présenté dans la figure 3 par les traits verticaux (la phase d'introduction est absente car réduite dans ce cas à un simple « bonjour »). On peut observer une parfaite adéquation entre ce découpage et la courbe décrite par le partage de la parole : la phase où le médecin est le plus productif est bien celle de l'examen, la parole étant plus équitablement répartie lors de la phase de discussion.

Dans d'autres consultations, cette adéquation est loin d'être parfaite, et la courbe peut présenter des profils bien plus accidentés. Dans ces cas-là, la représentation globale est très utile pour repérer des lieux particuliers : zones denses de

questions, variations brusques ou plateaux indiquant des ruptures significatives du dialogue voire des phases de monologue. Nous avons illustré cette diversité dans Vergely *et al.* (2009).

Ces graphiques sont eux-mêmes intégrés à des fiches synthétiques générées automatiquement pour chaque consultation, tout en conservant un accès direct au contenu textuel. Le type d'annotation proposé rend donc possible des modes de visualisation originaux de chaque consultation : il permet à la fois d'en représenter le déroulement et de mettre en évidence des types de consultation qui se caractérisent par des combinaisons particulières des différentes variables disponibles (par exemple, telle consultation sera caractérisée par une parole dominante de la part du patient, par l'abondance des questions et une part importante de vocabulaire commun). Ce type de représentation permet de compléter et de dépasser la simplification quantitative qu'entraînent des calculs globaux à l'échelle d'une consultation. Si nous en sommes encore aux premières étapes méthodologiques, nous envisageons ensuite la possibilité d'exploiter cette représentation pour élaborer un profilage plus fin des consultations. Cette seconde phase présente un intérêt majeur pour les objectifs pratiques du projet, en permettant notamment d'identifier des consultations-types et des passages remarquables utilisables comme matériau pédagogique dans la formation des médecins généralistes.

Conclusion

Les résultats obtenus jusqu'à présent ont permis d'amorcer quelques éléments de réponse à la question initiale à laquelle le projet INTERMEDE cherchait à répondre. A ce stade, nous ne concluons pas sur la question des inégalités de santé : pour aller plus loin dans cette direction nous aurions besoin d'informations externes plus riches (mesure de la compréhension du patient, voire de l'efficacité de son traitement). Toutefois, nous avons pu constater que des

caractéristiques de l'interaction entre patient et médecin sont corrélées avec certaines caractéristiques sociales des acteurs. Les consultations les plus longues, où la parole du patient domine et où la part du vocabulaire partagé est importante, sont caractéristiques des médecins femmes. Le médecin semble utiliser un discours plus subjectif lorsqu'il s'adresse à des patients âgés et d'un niveau social inférieur, réservant un discours plus modalisé aux patients jeunes et socialement plus favorisés. Ces premières pistes d'analyse doivent être poursuivies, en particulier dans le sens d'une prise en compte d'autres marques de la subjectivité, et d'une étude plus fine de certains traits, comme les répétitions ou les questions, qui peuvent correspondre à des situations interactionnelles variées.

Du point de vue méthodologique, nous avons montré comment une approche de linguistique outillée pouvait fournir à l'ensemble des partenaires des modalités nouvelles de visualisation et d'exploration du corpus. D'abord, en travaillant sur la normalisation et l'annotation du corpus transcrit, facilitant ainsi l'ensemble des traitements ultérieurs. Ensuite, en proposant de nouveaux modes d'accès au corpus : chaque consultation est indexée à l'aide de l'ensemble des variables linguistiques que nous avons mises au jour ; son déroulement est visualisé sous une forme graphique qui permet d'en saisir l'organisation générale et d'y repérer des moments d'interaction particuliers. Enfin, notre rôle d'interface est facilité par la capacité de produire une quantification des faits que nous avons choisi d'examiner. Cela nous a permis en particulier d'intégrer dans les traitements des variables produites par les autres partenaires (CSP reformulées, types de consultation).

Le contexte de ce projet nous a permis de prendre la mesure du rôle et des spécificités de la linguistique dans un cadre interdisciplinaire : s'il est clair que les approches outillées de type lexicométrique sont désormais utilisées prioritairement par les sciences sociales pour exploiter le matériau langagier, nous avons montré l'intérêt d'explorer d'autres dimensions linguistiques et de ne pas se limiter à une technique unique bien

éprouvée qui n'examine que la dimension lexicale. Nous envisageons dans un deuxième temps de prolonger ce travail par la prise en compte d'autres aspects de l'interaction, comme l'explicitation et la reformulation de la part du médecin, ou encore le recours à un vocabulaire spécialisé. Après un travail spécifique de description et de modélisation des marques linguistiques correspondantes, ces caractéristiques pourront être quantifiées et ainsi entrer dans le cadre méthodologique que nous avons déployé.

Références bibliographiques

- Ainsworth-Vaughn, N. (2001). The discourse of medical encounters, in D. Schiffrin, D. Tannen & H.E. Hamilton (eds) *The Handbook of Discourse Analysis*, Wiley-Blackwell, 453-469.
- Biber, D. (1988). *Variation across speech and writing*, Cambridge University Press.
- Génolini, J.-P. Roca, R., Rolland, C. & Membrado, M. (à paraître), L'éducation du patient en médecine générale : une activité périphérique dans la relation de soin, *Sciences Sociales et Santé*.
- Have, P. ten (1989). The consultation as a genre, in B. Torode (ed.), *Text and talk as social practice*. 115-133. Dordrecht: Foris Publications.
- Heritage, J. & Maynard, D. W. (eds.) (2006). *Communication in Medical Care. Interaction Between Primary Care Physicians and Patients*. Cambridge University Press.
- Kerbrat-Orecchioni, C. (1980). *L'énonciation*, Armand Colin.
- Marchand, P., Lang, T. & Rolland C. (2009) Exploration textométrique des relations praticien / patient (projet Intermède), 2ème Congrès de l'ADELF, *Epidémiologie et Inégalités sociales de santé*, Toulouse, 28-29 mai 2009 (communication affichée).

- Mayaffre, D. (2007) L'analyse de données textuelles aujourd'hui : du corpus comme une urne, au corpus comme un plan. *Lexicometrica*, 7.
- Mellet, S. & Barthélémy, J.-P. (2007). La topologie textuelle : légitimation d'une notion émergente, *Lexicometrica*, 7.
- Roter, D.L. & McNeilis, K.S. (2003). The nature of the therapeutic relationship and the assessment of its discourse in routine medical visits, in Thompson, T.L., Dorsey, A.M., Miller, K.I. and Parrot, R. (eds), *Handbook of Health Communication*, Erlbaum.
- Schieber A-C., Kelly-Irving, M., Rolland, C., Afrite, A., Cases, C., Dourgnon, P., Lombrail, P., Pascal, J., Lang, T., (à paraître). Do doctors and patients share perceptions of cardiovascular risk factors management during a consultation? The INTERMEDE study, *British Journal of General Practice*.
- Skelton, J.R. & Hobbs, F.D. (1999). Descriptive study of cooperative language in primary care consultations by male and female doctors. *British Medical Journal*, 318 : 576-579.
- Vergely, P., Condamines, A., Fabre, C., Josselin-Leray, A., Rebeyrolle, J. & Tanguy, L. (2009). Analyse linguistique des interactions patient/médecin, *Actes éducatifs et soins : entre éthique et gouvernance*, atelier « l'acte de soin approché par les aspects langagiers », Nice, 4-5 Juin 2009.
- Waitzkin, H. (1990). On studying the discourse of medical encounters: a critique of quantitative and qualitative methods and a proposal for reasonable compromise. *Medical Care*, 28(6), 473-488.