

LA LANGUE DE LA MÉDECINE

Rapport technique (2006-2007)¹

INTRODUCTION

Nous souhaitons développer une *approche linguistique* des textes médicaux. Les langues de spécialité ne se réduisent pas à des nomenclatures terminologiques ou à des réseaux conceptuels : elles constituent des « langues », au sens propre du terme, et ces langues se réalisent dans des discours et dans des textes.

Notre point de départ est une base de données que nous avons constituée, il y a quelques années, dans le cadre d'un projet européen (projet Esprit *Webling*, 1998-2001). Elaborée à partir des principaux dictionnaires de médecine disponibles en français, cette base inclut plus de 40 000 noms, 4 000 adjectifs et plus de 600 verbes.

L'idée est de réactiver ces données, de les enrichir à partir des ressources en ligne, et surtout d'y appliquer notre méthodologie. Cette application se fait, principalement, selon deux axes complémentaires :

- la constitution de dictionnaires électroniques intégrant les « classes d'objets », afin d'articuler le plus étroitement possible le lexique, la syntaxe et la sémantique ;
- la constitution de corpus à partir du Web.

1. DICTIONNAIRES ÉLECTRONIQUES : MÉTHODOLOGIE

L'adaptation du modèle des « classes d'objets », mis au point initialement pour décrire la langue générale, à la langue de la médecine, implique deux types de dictionnaires : les dictionnaires de prédicats (verbes, adjectifs, noms prédictifs) et les dictionnaires d'arguments.

1.1. Dictionnaire des verbes

Par « verbes médicaux », il faut entendre aussi bien des verbes strictement spécialisés, c.-à-d. limités au domaine (par ex. *hospitaliser, cautériser, dialyser*), que des verbes généraux qui prennent en médecine un sens particulier. Dans une phrase comme *Mon médecin, qui me suit depuis longtemps, m'a arrêté pour un mois et m'a donné des antibiotiques*, il est clair que les verbes *suivre, arrêter* et *donner* n'ont pas le même sens que dans la langue générale.

Il convient donc d'inclure dans la description, à côté des verbes purement médicaux, des « emplois » médicaux de verbes généraux. Dans un cas comme dans l'autre, la description implique l'identification des schémas d'arguments, en précisant, quand il y a lieu, leur typage sémantique, en termes de traits et de classes d'objets.

¹ Ce rapport résume les travaux effectués, au cours de l'année 2006-2007, par un groupe de travail que j'ai animé, avec la participation de Sophie de PONTONX (pour le codage des termes), Taoufik MESSOUSSI (pour l'enrichissement lexical) et Fabrice ISSAC (pour le projet de corpus).

Exemples de verbes strictement spécialisés :

N ₀	Verbes	N ₁	Dérivés	Spécialités
hum:médecin	cautériser	inc:plaie	<i>cautérisation</i>	thérap.
hum:dentiste	cautériser	inc:dent	<i>cautérisation</i>	odont.
hum:médecin	hospitaliser	hum:patient	<i>hospitalisation</i>	
hum:médecin	néphrectomiser	hum:patient		néphrol., chir.
inc:plaie	suppurer		<i>suppuration</i>	pathol.
inc:plaie	surinfecter (se)		<i>surinfection</i>	pathol.
inc:plaie	surinfecté (être)		<i>surinfection</i>	pathol.
hum:chirurgien	trépaner	hum:patient	<i>trépanation</i>	chir.
hum:chirurgien	trépaner	inc:boîte crânienne, os	<i>trépanation</i>	chir.

Exemples de spécialisation de verbes généraux :

N ₀	Verbes	P ₁	N ₁	P ₂	N ₂	Dérivés
hum:col	admettre		hum:patient	à, dans	loc:hôpital	<i>admission</i>
hum:médecin	administrer		inc:médicament	à	hum:patient	<i>administration</i>
évé:maladie	affecter		hum:patient			<i>affection</i>
inc:médicament	agir	sur	inc:organisme			<i>action</i>
état	aggraver (s')					<i>aggravation</i>
hum	arrêter		in:traitement			<i>arrêt</i>
hum:médecin	arrêter		hum:patient	pour	tps:durée	<i>arrêt</i>
hum:médecin	arrêter		évé:hémorragie			<i>arrêt</i>
hum:fœtus	bouger					
hum:médecin	consulter					<i>consultation</i>
hum:patient	consulter		hum:médecin			<i>consultation</i>
hum:médecin	couronner		inc:dent			
inc:dent	déchaussé (être)					<i>déchaussement</i>
inc:dent	déchausser (se)					<i>déchaussement</i>
hum:médecin	déclencher		évé: accouchement			<i>déclenchement</i>
inc:médicament	dégager		inc:nez, gorge, bronches			<i>dégagement</i>
hum:médecin	donner		inc:médicament	à	hum:patient	
hum:médecin	donner		tps:période d'arrêt	à	hum:patient	
hum:médecin	endormir		hum:patient			
hum:médecin	examiner		hum:patient			<i>examen</i>
hum:médecin	prélever		inc:organe			<i>prélèvement</i>
hum:médecin	prélever		inc:sang, excrétiens			<i>prélèvement</i>
hum:patient	prendre		inc:médicament			<i>prise</i>
hum:médecin	prescrire		inc:médicament	à	hum:patient	<i>prescription</i>
état:douleur, maladie	réveiller (se)					<i>réveil</i>
hum:patient	réveiller (se)					<i>réveil</i>
hum:médecin	suivre		hum:patient			<i>suivi</i>
hum:patient	suivre		in:traitement			

(On reconnaîtra ici les verbes *arrêter*, *donner* et *suivre* dont nous parlions précédemment.)

1.2. Dictionnaire des adjectifs

Les mêmes principes s'appliquent aux adjectifs prédicatifs. Ils doivent être, eux aussi, intégrés dans leurs schémas d'arguments. Soit par exemple l'adjectif *bon* :

Ce sirop est bon pour la toux.

Ce sirop est bon pour la gorge.

Malgré leur ressemblance, ces deux énoncés doivent être distingués, comme le prouve la variation suivante :

Ce sirop est bon contre la toux.

* *Ce sirop est bon contre la gorge.*

Dans le premier cas seulement, *contre* est synonyme de *pour* ! Ce qu'on peut représenter ainsi :

N ₀	Adjectifs	P ₁	N ₁
inc:médicament	bon	pour + contre	ina:maladie
inc:médicament	bon	pour	inc:npc (partie du corps)

1.3. Dictionnaire des noms prédicatifs

Les prédicats nominaux exigent le même type de description :

N ₀	Vsup	Npréd	P ₁	N ₁	P ₂	N ₂
hum	avoir	rougeole				
hum	faire	prise de sang	à	hum		
hum	pratiquer	ablation	de	npc	sur	hum

1.4. Classes de prédicats

On peut aller au-delà de la distinction morphologique (nom/verbe/adjectif), pour rapprocher les prédicats de même racine :

amputer/ G:v/ N0:chir/ N1:hum/ N2:de Npc/

amputation/ G:nf/ N0:chir/ N1:de Npc/ N2:sur hum/ Vsup:pratiquer

anorexie/ G:nf/ T:état/ C:état pathol/ N0:hum/ Vsup:souffrir de

anorexique/ G:adj/ C:état/ N0:hum/ Vsup:être

En outre, il est possible de regrouper les prédicats en classes cohérentes, en fonction de leur mode de fonctionnement. On identifiera ainsi :

- des prédicats d'action : consultations, examens, soins, prescriptions, opérations, anesthésies...
- des prédicats d'événements : maladies infectieuses, inflammations, attaques...
- des prédicats d'état : états pathologiques, états sains, positions, etc.

1.5. Dictionnaires d'arguments

Aux dictionnaires de prédicats correspondent, symétriquement, les dictionnaires d'arguments, où tous les termes se trouvent catégorisés en traits et en classes d'objets, en fonction des prédicats qui les sélectionnent. Citons, entre autres classes spécifiques, des ensembles tels que :

- les parties du corps : os, muscles, vaisseaux, nerfs, glandes, liquides organiques... ;
- les micro-organismes : bactéries (*pneumocoque*), virus (*adénovirus*), champignons microscopiques (*actinomyète*)... ;
- les remèdes : médicaments (*antihistaminique*), plantes médicinales (*aspérule odorante*)... ;

- le matériel médical : instruments d'observation (*stéthoscope*), instruments de mesure (*asthmomètre*), instruments chirurgicaux (*bistouri*)... ;
- les locatifs : lieux de soins (*salle d'opération*), etc.

À chaque classe correspondent des prédicats appropriés :

- un os peut se *fracturer*, un vaisseau *se boucher*, un muscle *se froisser* (mais on ne saurait parler de la *fracture* d'un muscle ou du *froissement* d'un os...).

Nous appliquons ainsi à la médecine la méthode mise au point pour la langue générale, en associant étroitement, dans le cadre des dictionnaires, lexicque, syntaxe et sémantique.

2. DICTIONNAIRES ÉLECTRONIQUES : RÉALISATIONS

Nous sommes partis d'une base de 44690 noms simples et composés, prédicats ou arguments :

- prédicats : *dermatose, rupture du tympan, crise de panique, coma diabétique hyperosmolaire anesthésie, lipiodiagnostic, transfusion, ablation...*
- arguments : *cavité amniotique, humérus, pneumocoque, stéthoscope, goutte-à-goutte, antihistaminique, aspérule odorante, bloc opératoire...*

Certains de ces noms avaient reçu un premier codage en termes de domaine, de trait ou de classe.

2.1. Domaines

Nous avons adopté, provisoirement, la nomenclature suivante, conforme à la tradition lexicographique (cf. GDEL, TLF, etc.) :

- | | |
|--------------------------------------|--|
| - anatomie (anat.) | - microbiologie (microbiol.) |
| - andrologie (androl.) | - néphrologie (néphrol.) |
| - anesthésiologie (anesth.) | - neurobiologie (neurobiol.) |
| - angiologie (angiol.) | - neurologie (neurol.) |
| - biochim. (biochim.) | - nutrition (nutrit.) |
| - biologie (biol.) | - obstétrique (obstétr.) |
| - cardiologie (cardiol.) | - odonto-stomatologie (odonto-stomat.) |
| - chirurgie (chir.) | - oncologie (oncol.) |
| - cytologie (cytol.) | - ophtalmologie (ophtalm.) |
| - dermatologie (dermat.) | - oto-rhino-laryngologie (ORL) |
| - embryologie (embryol.) | - orthopédie (orthopéd.) |
| - endocrinologie (endocrin.) | - parasitologie (parasitol.) |
| - épidémiologie (épidém.) | - pathologie (path.) |
| - examens médicaux (exam.) | - pédiatrie (pédiat.) |
| - gastro-entérologie (gastro-entér.) | - pharmacologie (pharm.) |
| - génétique (génét.) | - physiologie (physiol.) |
| - gériatrie (gériat.) | - pneumologie (pneumol.) |
| - gynécologie (gynécol.) | - psychologie (psych.) |
| - hématologie (hémat.) | - psychiatrie (psychiat.) |
| - hépatologie (hépatol.) | - radiologie (radiol.) |
| - histologie (histol.) | - rhumatologie (rhumat.) |
| - immunologie (immunol.) | - tératologie (tératol.) |
| - matériel médical (mat.méd.) | - thérapeutique (thérap.) |
| - médecines douces (méd. douces) | - toxicologie (toxicol.) |
| - médecine générale (méd.gén.) | - urologie (urol.) |

Cette liste est toutefois disparate et doit être restructurée en fonction de paramètres plus homogènes :

- spécialités liées à des parties du corps : *ophtalmologie, cardiologie, etc.*
- modes d'action : *chirurgie, anesthésie, etc.*
- âge du patient : *gériatrie, pédiatrie, etc.*
- etc.

Nous nous sommes limités pour l'instant à dégager une liste de 8 grands « secteurs d'activité médicale », ainsi constituée :

- disciplines scientifiques en amont de la médecine : *physique, chimie, biologie, psychologie*
- description et fonctionnement du corps humain : *anatomie, physiologie*
- description et traitement de la maladie : *pathologie, thérapeutique*

2.2. Classes

Il a fallu dans un premier temps nettoyer la base et attribuer des classes aux entrées afin de parvenir à un codage cohérent et permettre ainsi leur traitement.

Etant donné la quantité de noms à classer, nous avons décidé, en cours de traitement, de commencer par répartir ces noms en hyperclasses plutôt qu'en classes trop précises. Cela nous permettait d'avoir une vision plus rapide et globale de la nomenclature.

Ce travail nous a conduits à mener une réflexion sur la façon de décrire les classes et de les hiérarchiser entre elles.

Nous avons dégagés 4 niveaux d'analyse :

- Les *traits* : liste fermée de 8 codes :
 - *humain (hum)*
 - *animal (ani)*
 - *végétal (vég)*
 - *inanimé (in)*
 - *inanimé abstrait (ina)*
 - *inanimé concret(inc)*
 - *locatif (loc)*
 - *temps (tps)*

Remarque : Les noms difficilement analysables en termes d'opposition concret/abstrait reçoivent le code neutre *in (inanimé)*. Ceci vaut pour des classes comme les *documents* ou les *tumeurs* par exemple qui peuvent justifier selon le contexte l'une ou l'autre interprétation.

- Les *hyperclasses* : liste restreinte à 17 pour l'instant :
 - *action*
 - *état*
 - *événement*
 - *document*
 - *élément-biologique*
 - *fonction*
 - *matériel*
 - *micro-organisme*
 - *partie du corps*
 - *phénomène*
 - *processus*
 - *production pathologique*
 - *propriété*
 - *quantification*
 - *spécialité*
 - *substance*
 - *théorie*

- Les *classes* : liste ouverte (environ 500).
Exemples : *muscle, os, opération, maladie transmissible, inflammation, analyse...*
- Les *sous-classes* : liste ouverte. Ce niveau d'analyse n'est activé que pour un nombre restreint d'entrées, selon les besoins de la description syntactico-sémantique.

S'agissant du degré de spécialisation, on distinguera deux situations.

Certaines classes ont un spectre qui dépasse le champ de la médecine. C'est le cas de la classe des *instruments de mesure médicaux* par exemple. Si, du point de vue lexical, les termes qui la composent sont propres à la médecine, ils partagent du point de vue syntaxique le même comportement que tous les autres instruments de mesure. Ainsi,

Un asthmomètre, un cardiotachymètre, un esthésiomètre, un tensiomètre... peuvent être précis, justes, exacts... ou au contraire inexacts, déréglés..., tout comme un anémomètre ou un baromètre.

D'autres classes sont quant à elles spécifiques à la médecine, qu'il s'agisse de classes de prédicats comme les *maladies* ou les *opérations chirurgicales* ou d'arguments comme les *parties du corps*. Ces classes feront l'objet des prochaines descriptions.

3. CONSTITUTION D'UN CORPUS MÉDICAL

Notre objectif est de rassembler sous forme de corpus un ensemble représentatif de textes médicaux. Ce corpus regroupera un ensemble de textes récupérés sur le Web.

3.1. Méthodologie

La constitution du corpus se fera en plusieurs étapes :

1. Récupération d'URL candidates : nous utilisons la technique de l'amorce, à partir des ressources lexicales dont nous disposons, pour identifier les pages susceptibles d'être liées au domaine de la médecine.
2. Tri et sélection des URL : ce travail permet de cibler plus précisément les pages « intéressantes » par l'utilisation d'heuristiques (fréquence de rappel par exemple).
3. Récupération et nettoyage des pages : nous utilisons un aspirateur de pages classique pour récupérer les URL issues de la sélection. L'étape de nettoyage est relativement simple, la structure des documents n'étant pas notre objectif.

3.2. Typologie des pages

La constitution du corpus se fera dans un premier temps « en aveugle » en ce qui concerne le type de documents rencontrés. Il est en effet possible, dans le domaine de la médecine, d'identifier différents types de sources :

- Vulgarisation (*Doctissimo*)
- Supports pédagogiques (universités)
- Dictionnaires (mono ou multilingues)
- Pages personnelles
- Sites d'information pour les professionnels
- Sites institutionnels (Sécurité sociale)
- etc.

Les termes, mais aussi la langue, rencontrés dans ces types de documents présentent un certain nombre de différences. Il est donc nécessaire, puisque nous souhaitons formaliser la langue de la médecine sous forme de classes d'objets, de pouvoir indiquer ces différences.

3.2. Exploitation linguistique du corpus

Ce corpus permettra ainsi :

- d'élaborer des concordances pour étudier le contexte des termes médicaux ; ainsi, à partir d'une première liste de noms d'inflammations, il est possible d'étudier systématiquement les contextes phrastiques dans lesquels ils se trouvent insérés ;
- de repartir de la syntaxe ainsi établie pour interroger les corpus : tous les termes apparaissant dans les contextes caractéristiques d'une classe sont virtuellement susceptibles d'appartenir à cette classe ; après vérification de la pertinence des candidats termes ainsi repérés, il est possible de les ajouter au dictionnaire.

Cette procédure comporte plusieurs avantages : elle offre une confrontation directe et à double sens entre dictionnaires et corpus ; elle permet un enrichissement progressif et raisonné de la nomenclature ; elle assure une amélioration de la description syntaxique des classes, grâce à un meilleur contrôle de son adéquation aux textes réels.