

Recherche terminographique semi-automatisée en veille terminologique : expérimentation dans le domaine médical

Le présent article fait état d'une recherche terminographique intégrant des outils informatiques (outil de dépouillement, concordancier, système de gestion de bases de données et module de création de fiches terminologiques). La recherche, qui portait principalement sur un sous-domaine de la médecine, à savoir la pharmacologie cardiovasculaire, se limitait à des concepts qui n'avaient pas fait l'objet d'un traitement préalable dans la *Banque de terminologie du Québec (BTQ)*. Nous relatons les étapes les plus importantes de la recherche et décrivons les principales difficultés rencontrées. Nous proposons également des solutions fondées sur les lacunes observées.

Termes-clés :
terminographie, terminotique,
veille terminologique,
dépouillement assisté par ordinateur.

1 Introduction

Le terminologue peut tirer profit d'un nombre croissant d'outils informatiques qu'il utilise pour accélérer ou systématiser des tâches répétitives ou fastidieuses. Nous faisons référence ici aux logiciels de dépouillement assisté par ordinateur, aux concordanciers, aux systèmes de gestion de bases de données et aux logiciels d'édition de fiches terminologiques. Ces outils sont également utilisés par d'autres professionnels, comme les traducteurs ou les lexicographes, mais le terminologue en adapte certaines fonctions à ses besoins particuliers. En effet, celui-ci applique une méthode de recherche qui rend compte d'un certain nombre de concepts rattachés à un domaine de spécialité qui a été circonscrit au préalable.

Le présent article porte sur l'utilisation de certains de ces outils dans un contexte particulier, à savoir la veille terminologique. La *veille terminologique* est une activité qui consiste à repérer des termes « nouveaux » et à colliger des renseignements sur ces termes (contextes, notes et définitions) dans le but de les décrire. Nous adoptons une définition fonctionnelle de la « nouveauté d'un terme » inspirée de Perron (1997) qui s'appuie sur l'utilisation d'un corpus d'exclusion : un terme nouveau, ou *néonyme*, est un terme qui n'a pas fait l'objet d'une description terminologique dans un répertoire donné (le corpus

d'exclusion utilisé ici est la *Banque de terminologie du Québec (BTQ)*).

La recherche a été réalisée dans le cadre d'une entente avec l'Office de la langue française, organisme gouvernemental québécois. Nous devons nous pencher sur environ 400 concepts liés au domaine de la médecine, plus précisément à la radiologie et à la pharmacologie cardiovasculaire⁽¹⁾. Les fiches terminologiques produites étaient bilingues (anglais-français), mais nous nous concentrerons sur les aspects touchant les termes français liés au domaine de la pharmacologie cardiovasculaire. Notre équipe⁽²⁾ devait procéder en outre à l'expérimentation de différents outils informatiques et déterminer s'ils présentaient un intérêt dans un contexte professionnel. Les outils testés sont *Adepto-Nomino* (Perron 1996) et *Sami-BTQ* (Pelletier 1996)⁽³⁾.

Nous présentons d'abord un panorama des principales étapes de la recherche, pour ensuite nous pencher sur les difficultés liées à chacune des

(1) Ces sous-domaines de la médecine ont été choisis en fonction de l'accessibilité de textes en format électronique. Toutefois, il aurait été préférable de travailler sur un seul sous-domaine et d'enrichir le corpus de textes. Une recherche portant sur deux sous-domaines simultanément présente des difficultés notables puisque le terminologue doit se familiariser avec des notions parfois éloignées. Nous n'aborderons pas ces difficultés dans le cadre du présent article, mais elles ont été présentes durant toutes les étapes de la recherche.

(2) Nous tenons à remercier les personnes qui, outre les trois auteurs, ont participé au projet de recherche : Anne Chudobiak, Claire-Hélène Lavigne, Sophie Milliard, Jean Ntakirutimana et Margreet de Rooij.

(3) Signalons par ailleurs que nous avons intégré d'autres logiciels plus largement répandus afin de prendre en charge certaines tâches périphériques (les logiciels en question sont cités au tableau 2).

étapes franchies. En conclusion, nous suggérons quelques avenues afin d'améliorer la recherche telle que nous l'avons réalisée.

2 Principales étapes de recherche

Les méthodes de recherche terminographiques font l'objet d'un consensus depuis déjà quelques décennies. Toutefois, l'intégration de certains outils informatiques vient bouleverser en partie ces méthodes. Nous présentons au tableau 1 les différences observées entre une recherche terminologique systématique « traditionnelle » et une recherche qui a recours à des outils informatiques (cette description est largement inspirée de Auger *et al.* (1991)).

Nous nous sommes alignées sur la démarche présentée dans la troisième colonne du tableau 1, tout en tenant compte des particularités liées à la veille terminologique (il en sera question un peu plus loin). À chaque étape de la recherche, nous avons intégré un logiciel donné ou fait appel à une fonction présente dans un logiciel : le tableau 2 dresse la liste des différents outils informatiques utilisés.

La recherche, comme on peut le constater, s'appuie, en dépit de l'intégration d'outils informatiques, sur des méthodes de recherche systématique (Auger et Rousseau 1978). L'aspect prédominant de ce type de recherche est la prise en compte de l'ensemble des termes d'un domaine : le terminologue définit, ultimement, les termes les uns par rapport aux autres. Cette stratégie permet de dégager les liens existant entre les concepts du domaine (hyponymie et hyperonymie, co-hyponymie, méronymie, etc.) et ce pour parvenir à dégager les

Tableau 1 :
Comparatif des méthodes de travail traditionnelles et des méthodes de travail intégrant des outils informatiques.

	<i>Méthode traditionnelle</i>	<i>Méthode intégrant des outils informatiques</i>
Recherche documentaire : Mise en forme du corpus	Collecte de textes de nature variée portant sur un domaine de spécialité.	1. Recherche et collecte de textes en format électronique portant sur un domaine de spécialité. 2. Conversion de textes papier en format électronique (numérisation).
Repérage des unités terminologiques	Repérage de termes au moment de la lecture de textes et transcription des contextes.	1. Identification automatique de termes (logiciel d'identification automatique des unités potentiellement terminologiques). 2. Première sélection humaine des unités produites par l'identification automatique qui feront l'objet de recherches ultérieures.
Relevé de contextes	Relevé de contextes lors de la lecture des textes. Raffinement de la liste de termes devant faire l'objet de descriptions.	1. Repérage automatique de contextes pour les termes retenus et sélection humaine des meilleurs contextes. 2. Raffinement de la liste de termes devant faire l'objet de descriptions.
Analyse des contextes	Étude des contextes relevés pour la production de fiches terminologiques. Information complémentaire puisée dans des ouvrages de référence.	1. Étude des contextes relevés pour la production de fiches terminologiques. 2. Information complémentaire puisée dans des ressources électroniques (par exemple, Internet) et, si nécessaire, dans des ouvrages de référence papier.
Rédaction de fiches terminologiques	Reproduction, dans une fiche terminologique, d'une partie des informations recueillies lors du dépouillement. Ajout d'information complémentaire. Reproduction de la fiche dans une banque de données terminologiques.	1. Copie, dans une fiche terminologique, d'une partie des informations recueillies lors du dépouillement. 2. Ajout d'information complémentaire. 3. Importation de la fiche dans une banque de données terminologiques.

Tableau 2 : Outils utilisés lors de la recherche

<i>Tâche</i>	<i>Procédé ou outil informatique pour la réaliser</i>
Mise en forme du corpus	Numérisation (numériseur pleine-page), reconnaissance optique de caractères (<i>Omnipage Direct</i>) et vérification orthographique pour la correction des textes numérisés (vérificateur de <i>Word</i>). Sélection de textes à partir de sites Internet
Identification automatique de termes	Analyse linguistique (<i>Nomino</i> ⁽⁴⁾) Listes de termes gérées dans un fichier <i>Excel</i>
Relevé et analyse des contextes	Sélection de contextes proposés par le concordancier (<i>Adepte</i>) Copie des contextes sélectionnés dans <i>Adepte</i> (système de gestion de bases de données intégré dans l'environnement <i>Adepte</i>) Lecture et analyse des contextes relevés dans <i>Adepte</i> pour chacun des termes (établissement de renvois pour les termes apparentés)
Recherche d'information complémentaire	Consultation d'ouvrages de référence sur support électronique (banques de terminologie, dictionnaires sur CD-Rom) Consultation de sites Internet En dernier recours, consultation d'ouvrages de référence sur support papier
Rédaction de fiches	Copie et collage dans <i>Sami-BTQ</i>

caractéristiques qui définiront les termes dénotant ces concepts.

Nous avons dû cependant tenir compte de certaines difficultés liées à la recherche terminologique menée en contexte de veille telle qu'elle a été définie à la section 1 (rappelons que nous avons écarté des termes déjà décrits dans le corpus d'exclusion, à savoir la *BTQ*).

Dans ce contexte, il est parfois difficile d'obtenir une vue d'ensemble

(4) *Adepte* et *Nomino* sont fondus dans le même environnement de travail. Toutefois, *Nomino* fait le dépouillement et les autres tâches sont réalisées au moyen de fonctions présentes dans *Adepte*. Cette distinction ne sera pas toujours faite dans ce qui suit.

du système conceptuel d'un domaine comparable à celle donnée par la recherche systématique véritable. Cette lacune est causée par le fait qu'une partie des termes du domaine peut déjà faire l'objet d'une description terminologique, alors qu'une seconde partie reste à décrire. D'autres lacunes sont entraînées par le fait que les traitements automatiques restent imparfaits et reposent exclusivement sur un corpus qui peut lui-même présenter des imperfections (par exemple, nous avons parfois exclu des termes pour lesquels le corpus ne livrait pas suffisamment de renseignements). Pour ces raisons, il arrive que le terminologue doive composer avec des listes de termes partielles.

Pour illustrer le sens de la dernière observation, nous avons reproduit au tableau 3 une partie de la liste de termes construits à partir de la tête *abcès* (ce sont les termes trouvés dans les textes du corpus utilisé dans le cadre de la recherche).

Tableau 3 :
Liste de termes dont la tête est *abcès*

<i>Termes trouvés dans le corpus</i>	<i>Présence dans la BTQ</i>
abcès appendiculaire	
abcès bilatéral	
abcès cérébral	BTQ
abcès de Brodie	BTQ
abcès du cerveau	BTQ
abcès du foie	BTQ
abcès du poumon	BTQ
abcès du psoas	BTQ
abcès encapsulé	
abcès épidural	
abcès intra-osseux	BTQ
abcès médullaire	
abcès nécrotique	
abcès paravertébral	
abcès para-vertébral	
abcès pelvien	
abcès péri-péritreux	

On remarque que, sur un total de 17 termes, sept sont déjà traités dans la *Banque de terminologie du Québec*. Donc, en principe, ces unités ne doivent pas faire l'objet d'un traitement dans un contexte de veille terminologique. Cependant, les «trous» signalés dans les listes peuvent soulever des problèmes au moment de l'analyse des renseignements recueillis.

Par exemple, les trous posent des problèmes du point de vue de l'appréhension d'un domaine de spécialité. Nous allons commenter ces difficultés au moyen d'un certain nombre de concepts liés au concept d'«action» (effet recherché par l'administration de médicaments) qui sont reproduits à la figure 1.

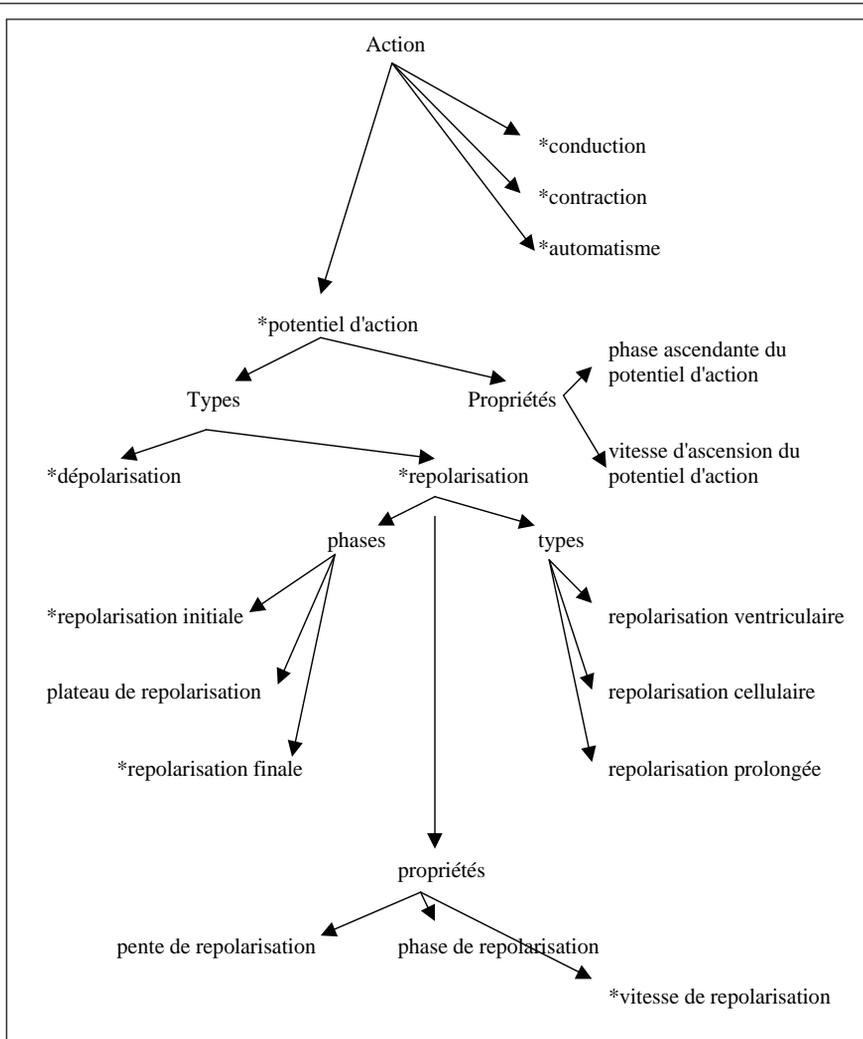


Figure 1: Concepts liés à « action »

Tableau 4: Termes construits à partir de *conduction* et de *conductibilité*

<i>Termes construits à partir de conduction</i>	<i>Termes construits à partir de conductibilité</i>
conduction tawarienne conduction cardiaque	conductibilité auriculaire conductibilité atrio-ventriculaire conductibilité cardiaque conductibilité ventriculaire conductibilité de ré-entrée

Tous les concepts présentés à la figure 1 n'ont pas fait l'objet d'un traitement pour différentes raisons (les concepts mis de côté sont signalés par l'astérisque). Nous présentons, ci-dessous, quelques causes pour lesquelles ils n'ont pas été retenus dans le cadre du projet.

Conduction: terme non relevé lors du dépouillement initial des textes français (qui accordait la priorité aux termes complexes). En revanche, certains hyponymes ont été retenus: *conduction cardiaque*, *conduction tawarienne*. De plus, certains termes faisant référence à la propriété de conduire les impulsions ont été traités: *conductibilité auriculaire*, *conductibilité atrio-ventriculaire*, *conductibilité cardiaque*, *conductibilité de ré-entrée* et *conductibilité ventriculaire*. On constate (tableau 4) le manque de parallélisme entre ces deux derniers groupes de concepts (qu'une véritable recherche systématique aurait corrigé).

Contraction: terme traité dans la *BTQ*. Ici encore, certains hyponymes ont été retenus: *contraction cardiaque*, *contraction myocardique*, *contraction spontanée*, *contraction tonique*, *contraction vasculaire*, *contraction ventriculaire*. De plus, certains termes étroitement liés à ces premières unités ont été traités dans le cadre du projet: *contractilité cardiaque*, *contractilité myocardique*, *contractilité ventriculaire*. De nouveaux parallélismes n'ont pas été saisis de façon appropriée (tableau 5).

Automatisme: terme non relevé lors du dépouillement des textes français. Certains hyponymes d'*automatisme* ont pourtant été traités: *automatisme auriculaire*, *automatisme cardiaque*, *automatisme ectopique*, *automatisme sinusal*. Un certain nombre de termes dénotant la propriété de générer des automatismes ont également été retenus: *automaticité ectopique*, *automaticité myocardique*, *automaticité*

Tableau 5 : Termes construits à partir de *contraction* et de *contractilité*

<i>Termes construits à partir de contraction</i>	<i>Termes construits à partir de contractilité</i>
contraction auriculaire contraction cardiaque contraction myocardique contraction tonique contraction vasculaire contraction ventriculaire	contractilité cardiaque contractilité myocardique contractilité ventriculaire

Tableau 6 : Termes construits à partir de *automatisme* et de *automaticité*

<i>Termes construits à partir de automatisme</i>	<i>Termes construits à partir de automaticité</i>
automatisme auriculaire automatisme cardiaque automatisme ectopique automatisme sinusal	automaticité ectopique automaticité sinusale automaticité ventriculaire

sinusale, automaticité ventriculaire.

Ici encore, des parallélismes n'ont pas été saisis de façon appropriée (tableau 6) :

Potentiel d'action: terme traité dans la *BTQ*;

Dépolarisation: terme traité dans la *BTQ*. Nous avons toutefois traité des hyponymes de ce terme :

dépolarisation diastolique, dépolarisation diastolique lente, dépolarisation spontanée, dépolarisation ventriculaire.

Repolarisation: terme traité dans la *BTQ*. Des hyponymes de ce terme ont été décrits: *repolarisation cellulaire, repolarisation en phases, repolarisation prolongée, repolarisation ventriculaire.*

Repolarisation initiale, repolarisation finale, vitesse de repolarisation: termes non relevés dans le corpus de départ.

Comme le montrent les exemples ci-dessus, seule une partie

des concepts fait l'objet d'un traitement terminographique. Cependant, tous les concepts d'un domaine doivent être appréhendés pour décrire les termes adéquatement. En contexte de veille, le terminologue doit lui-même combler les trous pour les concepts qui ne sont pas spécifiquement pris en charge dans le cadre d'un projet.

Ainsi, on peut penser, lors de la planification d'un projet de veille terminologique, qu'une description terminographique exigera moins de temps puisqu'elle ne porte que sur un nombre réduit de concepts. Toutefois, il n'en va pas tout à fait ainsi puisque les liens entre tous les concepts doivent être saisis. De plus, la systématisme, qui fait ordinairement la qualité d'une recherche terminographique, est perdue partiellement dans ce type de démarche.

3 Détail de la recherche

Nous allons maintenant nous pencher sur les étapes proprement dites de la recherche et sur certaines difficultés ponctuelles. Les sections qui suivent présentent les étapes décrites dans la troisième colonne du tableau 1. Précisons d'entrée de jeu que les différentes tâches sont décrites successivement dans le but de faciliter leur présentation. Dans les faits, cependant, les chevauchements et les aller et retour sont fréquents.

3.1 Mise en forme du corpus

Les termes à traiter devaient être extraits de textes en format électronique (qui étaient soumis, par la suite, à un logiciel de dépouillement, voir 3.2). Nous avons mis en forme un corpus de textes de pharmacologie cardiovasculaire qui comptait environ 300 000 mots⁽⁵⁾. Les textes sélectionnés avaient été numérisés lors d'un précédent projet de recherche, c'est-à-dire que les textes papier ont été convertis en format électronique.

Les textes du corpus devaient être rédigés en français, c'est-à-dire non traduits. Nous avons également pris soin de choisir des textes écrits par des auteurs différents pour constituer un corpus représentatif. À cette étape, nous n'avons pas mis l'accent sur la variété des niveaux de spécialisation des textes en tenant pour acquis qu'un échantillon suffisamment élaboré permettrait de constituer un corpus bien équilibré. Nous avons regroupé les textes accessibles en format électronique.

(5) Rappelons que nous avons travaillé sur un autre sous-domaine de la médecine, à savoir la radiologie; ces textes ont été téléchargés à partir de sites Internet.

Cependant, lors du traitement des termes, nous avons rencontré des difficultés importantes qui découlent des choix faits au départ. Il aurait été primordial, dès la première étape, de sélectionner des textes qui respectent des critères très précis.

En veille terminologique, les documents qui forment le corpus de dépouillement devraient contenir la plupart des renseignements essentiels à la rédaction d'une fiche puisque, en principe, ces renseignements ne se trouvent nulle part ailleurs (dans les ouvrages de référence, par exemple). De plus, le succès des traitements automatiques repose en très grande partie sur la qualité des textes formant le corpus de départ. Voici les principaux critères auxquels les documents devraient normalement répondre :

1. Une certaine uniformité des thèmes abordés dans un groupe de documents ;
2. Une documentation « équilibrée » : c'est-à-dire suffisamment pointue pour contenir des renseignements susceptibles d'éclairer les concepts terminologiques appartenant à un domaine relativement restreint et suffisamment accessible pour permettre au terminologue d'en tirer des définitions qui s'adressent à un vaste public ;
3. Des documents comportant un nombre d'occurrences suffisamment élevé des termes faisant l'objet d'un traitement ;
4. Un nombre élevé de textes de provenances diverses, afin d'accéder à des attestations des variantes terminologiques (synonymes, abréviations, etc.).

Ainsi, les documents intégrés dans un corpus de départ doivent faire l'objet d'une sélection rigoureuse. Cette observation reste valable peu importe le type de recherche terminologique entreprise – traditionnelle ou informatisée –, mais le problème est amplifié par le fait qu'en recherche informatisée on tente

de réunir des sources en format électronique qui ne répondent pas nécessairement à tous ces critères⁽⁶⁾. De plus, de nombreux textes présentant ces caractéristiques « idéales » (des manuels didactiques, des encyclopédies spécialisées) sont rarement en format électronique. Le terminologue peut numériser les textes idéaux, mais la numérisation devient rentable dans la mesure où le projet en est un de grande envergure.

À notre avis, la constitution d'un corpus électronique (dans une optique de recherche en contexte de veille) devrait se dérouler de la manière suivante :

1. L'identification des objectifs de la recherche (domaine, langue, nombre de concepts à décrire, mode de description des concepts) ;
2. Le repérage de documents de types variés portant sur le domaine choisi ;
3. La vérification de l'accessibilité des textes repérés en format électronique (copie de textes accessibles, conversion, si nécessaire, de textes intéressants en format électronique).

3.2 Identification automatique de termes

L'ensemble des textes sélectionnés ont été soumis au logiciel *Nomino* pour en extraire le contenu terminologique⁽⁷⁾. *Nomino* propose des listes de mots simples et de syntagmes nominaux à partir desquelles le terminologue sélectionne les unités qui feront l'objet d'un traitement terminographique.

Nous aborderons, dans les sections qui suivent, quelques difficultés soulevées par l'utilisation d'un logiciel de dépouillement. Nous décrirons d'abord les imperfections des listes proposées par le logiciel et leurs répercussions sur le travail terminographique (3.2.1). Nous nous pencherons par la suite sur deux méthodes d'utilisation des listes (3.2.2).

Dans le cadre du projet, le temps consacré à l'interprétation des listes proposées par *Nomino* a été prolongé en raison du fait qu'il fallait nous assurer que les termes retenus n'avaient pas fait l'objet d'une description préalable dans la *BTQ*. La consultation de la banque a dû être faite ponctuellement pour chacune des unités potentiellement intéressante puisqu'aucune procédure automatique n'allégeait cette tâche.

3.2.1 Listes proposées par le logiciel de dépouillement *Nomino*

Comme plusieurs autres logiciels de dépouillement, *Nomino* génère des listes de mots et de syntagmes nominaux qui sont imparfaites du point de vue de l'utilisateur : dans le cas qui nous préoccupe, les listes comportaient un grand nombre d'entrées qui n'ont pas fait l'objet d'un traitement terminographique. Le tableau 7 présente les principaux problèmes relevés dans les listes de termes complexes⁽⁸⁾.

(6) On a par ailleurs souvent tendance à exagérer l'intérêt que peuvent présenter les ressources trouvées sur Internet. Il est vrai que le *Web* offre une masse documentaire colossale mais l'expérience a montré qu'il peut se révéler difficile d'isoler, d'une part, puis de réunir, d'autre part, un nombre suffisamment élevé de documents de qualité et rédigés dans une autre langue que l'anglais (ce fait s'observe dans un certain nombre de domaines).

(7) Pour une présentation détaillée du fonctionnement du logiciel, voir Perron (1996).

(8) *Nomino* propose également des listes de mots simples. Bien que ces listes proposent également des unités terminologiques, la proportion de bruit est nettement plus élevée.

Tableau 7: Principales causes de bruit et exemples tirés des textes dépouillés⁽⁹⁾

Problèmes graphiques (fautes de typographie contenues dans le texte traités)	• <i>sur la conduction, effet indésirable</i>
Syntagme nominal non terminologique (syntagme construit à partir d'un nom mais qui n'est pas une unité terminologique)	<i>conduction au niveau des synapses des ganglions sympathiques, effet additionnel, effet bien connu; hypotension avec des variations; insuffisance préexistante</i>
Terme n'appartenant pas au domaine	<i>conduite automobile, année de l'enfant</i>
Mot mal catégorisé (un mot dans le syntagme relevé est mal analysé: verbe traité comme un nom, par exemple)	<i>contraction cardiaque cesse, hypertension artérielle comporte; insuffisance doit; résistance vasculaire pulmonaire montre</i>
Collocation plutôt qu'unité terminologique (groupe de mots ayant des affinités et qui apparaissent fréquemment dans les textes, mais qui ne sont pas des termes complexes)	<i>effet de la drogue, effet des médicaments, traitement de la migraine, traitement de l'angine de poitrine</i>
Mauvais découpage (le terme apparaissant dans le texte est soit plus long, soit plus court que le terme proposé dans la liste)	<i>hypertension à activité: hypertension à activité rénine basse; médicament à action: médicament à action alphabloquante</i>

De plus, même si les listes proposées par le logiciel de dépouillement étaient parfaites, c'est-à-dire qu'elles ne comportaient que des unités terminologiques valables compte tenu d'un projet donné (ce qui est loin d'être le cas), elles soulèvent des difficultés du point de vue de la démarche normalement adoptée par le terminologue. Ce dernier doit désormais procéder à partir d'une liste de formes linguistiques et en extraire le contenu conceptuel pertinent pour rédiger une définition et établir des liens entre concepts apparentés. Ce contenu conceptuel est moins immédiatement perceptible à partir d'une liste de

syntagmes nominaux proposés par un logiciel de dépouillement. (Nous verrons plus loin qu'une méthode permet toutefois de simuler la démarche conceptuelle.)

3.2.2 Utilisation des listes proposées par le logiciel de dépouillement Nomino

Le terminologue peut aborder les listes de syntagmes nominaux de deux manières. La première méthode consiste à procéder à un balayage systématique des entrées afin d'écartier toutes les unités ne devant pas faire l'objet d'une recherche plus poussée et de conserver les entrées qui

semblent intéressantes. Le terminologue doit d'abord s'assurer que l'unité proposée est effectivement un terme (certaines entrées sont éliminées rapidement; d'autres, au contraire, sont plus difficiles à traiter, car seule la lecture du contexte permet de prendre une décision éclairée). De plus, si la recherche est réalisée dans un contexte de veille terminologique, il doit s'assurer que le terme n'apparaît pas dans la BTQ. Nous avons testé cette technique sur une partie d'une liste de syntagmes nominaux générée par *Nomino*: nous avons éliminé environ 90 % des entrées de cette manière. Ce travail est très mécanique et peu satisfaisant, puisqu'on doit consacrer beaucoup de temps à isoler environ 10 % de termes dont l'intérêt n'est pas nécessairement assuré à cette étape.

La seconde méthode consiste à repérer, dans les listes proposées par le logiciel, des blocs de termes potentiellement intéressants sans tenir compte de l'ensemble des entrées dans les listes. Cette dernière méthode est nettement plus intéressante pour des motifs qui ressortiront clairement dans ce qui suit. Le terminologue travaille sur une série de termes apparentés sur le plan de la forme ou du sens, ou se trouvant dans un même environnement conceptuel⁽¹⁰⁾. Il utilise, comme point de départ, une série de termes construits à partir de la même tête.

(9) Les éléments entraînant du bruit varient en fonction de la stratégie d'identification adoptée par le logiciel et, parfois, de la langue traitée. De plus, la proportion de bruit dépend de l'application visée (pour de plus amples renseignements, voir Lauriston (1994); L'Homme *et al.* (1996) et Otman (1991)).

(10) Cette méthode a été suggérée dans Perron (1996).

Voici un exemple de la méthode utilisée dans les textes que nous avons dépouillés. Nous savons que les médicaments exercent un certain nombre d'«actions». Nous utilisons les termes construits à partir d'*action* comme point de départ d'un dépouillement. Nous avons reproduit dans le tableau 8 une partie de la liste des unités complexes proposées par *Adepte-Nomino* dont la tête est *action*:

Tableau 8 : Liste partielle des termes dont la tête est *action*

Termes complexes
action adrénérurgique
action adrénolytique
action agoniste
action alpha
action alpha-adrénérurgique
action alpha-adrénolytique
action alpha-agoniste
action alpha-agoniste adrénérurgique
action alphabloquante
action alphabloquante accessoire
action alphabloquante non spécifique
action analgésique
action analogue
action anesthésique
action anesthésique locale
action anorexigène
action antagoniste

Un premier repérage permet d'effacer les unités complexes relevées par *Adepte-Nomino*, mais qui ne sont pas des termes. Par exemple, *action analogue* a été écarté de la liste à la suite d'une première lecture. De même, *action alpha* a été exclu, car l'examen des contextes a permis de constater qu'*alpha* était toujours placé devant un adjectif (*alpha-adrénérurgique*, *alpha-adrénolytique*, etc.).

Ensuite, un recouplement avec les termes de la *BTQ* permet d'éliminer les unités ayant déjà fait l'objet d'une description terminologique. Aucun des termes ci-dessus n'apparaissait dans la *BTQ*, mais *action coagulante*

(listé un peu plus loin) a été repéré et, par conséquent, écarté de la liste.

Les éléments qui restent peuvent en principe faire l'objet d'une recherche. Le terminologue peut, par la suite, utiliser les contextes dans lesquels ces termes apparaissent pour repérer d'autres termes liés sur le plan conceptuel. Les contextes affichés contiendront en effet d'autres termes qui seront utilisés comme point de départ pour une autre recherche. Par exemple, une recherche sur *action adrénérurgique* donne accès au contexte suivant :

La plupart des bêta bloquants sont antagonistes des actions adrénérurgiques et au niveau des récepteurs bêta 1 et au niveau des bêta 2.

Ce contexte révèle la présence d'autres termes, à savoir *bêta bloquant*, *récepteur bêta 1* et *récepteur bêta 2*. Le terminologue peut effectuer des recherches sur ces termes qui serviront eux-mêmes de pistes pour d'autres unités et ainsi de suite jusqu'à ce qu'il épuise le filon. Cette méthode a permis de repérer, dans les textes de pharmacologie, des termes associés aux familles conceptuelles suivantes :

- Noms de médicaments ou de substances biochimiques (ex.: *alcaloïde de l'ergot*; *barorécepteur aortique*; *céliprolol*);
- Propriétés ou mécanismes d'action des médicaments (ex.: *absorption digestive*; *action adrénolytique*)
- Maladies et symptômes (ex.: *angor primaire*; *bradycardie réflexe*; *déficit enzymatique*);
- Propriétés ou mécanismes d'action des parties du corps (ex. *automaticité ectopique*, *conductibilité atrio-ventriculaire*, *débit coronarien*);
- Termes anatomiques (ex.: *ganglion végétatif*; *innervation parasymphatique*).

Ainsi, la démarche qui consiste à procéder à partir de termes apparentés est moins contraignante et plus satisfaisante pour le terminologue. C'est également celle qui se rapproche

le plus des méthodes de travail traditionnelles.

3.3 Extraction et analyse des contextes

Adepte comporte un concordancier qui affiche, pour toutes les unités présentes dans les listes, les phrases dans lesquelles ces unités apparaissent. Les contextes peuvent être sélectionnés ou simplement consultés. Les contextes sélectionnés sont copiés, à la demande de l'utilisateur, dans une fiche recevant l'ensemble des contextes choisis pour un terme.

À l'étape de l'analyse des contextes, nous nous sommes aperçu que les contextes extraits des corpus ne comportaient pas l'ensemble des éléments essentiels à la rédaction de définitions. Étant donné que les corpus étaient constitués de textes très spécialisés, c'est-à-dire destinés à des spécialistes du domaine, il a fallu puiser des renseignements complémentaires dans d'autres sources plus accessibles : par exemple, des manuels pédagogiques de format papier destinés à des étudiants en médecine ou encore des sites Internet qui s'adressent à la population en général pour vulgariser certaines connaissances scientifiques. Il aurait fallu choisir, et ce dès le début, des textes de type pédagogique ou informatif grand public pour éviter un surplus de travail à cette étape.

Même si on ne peut pas toujours en tirer des définitions proprement dites, les contextes contiennent de nombreux éléments qui permettent de cerner les concepts décrits. Nous avons reproduit quelques exemples ci-dessous :

- Éléments définitoires

Absorption intestinale : les graisses, insolubles dans l'eau, sont transformées en micelles composées d'acides gras, de monoglycérides et de sels biliaires.

(Notion de transformation liée à l'absorption intestinale)

Les bêta-bloquants n'ont, aux doses thérapeutiques, qu'une action adrénolytique, qui déplace le tonus neurovégétatif en faveur du parasympathique, ralentissant le nœud sinusal et freinant la conduction auriculoventriculaire. (Effet de l'action adrénolytique exercée par certaines substances)

– Relations explicites avec des concepts connexes

Certains neuroleptiques tels que l'halopéridol ou la chlorpromazine ont une action alphabloquante non spécifique. D'autres neuroleptiques tels que le sulpiride, la thioropérazine, ou le pimozide possèdent une action alphabloquante assez marquée à prédominance alpha 2 présynaptique qui pourrait expliquer leur profil «désinhibiteur», probablement lié à la libération accrue de catécholamines (fig. 34).* (Médicaments qui exercent une action alphabloquante).

Dissocier l'action de ces produits en une action antiarythmique et une action antifibrillante est plus ou moins légitime, car les deux ne sont pas sans relation. (Relation entre deux types d'actions voisines).

Le mécanisme de l'action hypotensive des hydrazinophthalazines est encore très mal connu. Certains auteurs ont supposé qu'elles inhibent directement les centres vasomoteurs. (Effets de l'action hypotensive).

Les contextes sont également utiles pour dénicher d'autres formes de renseignements :

– Attester certaines relations synonymiques établies dans les autres sources consultées (par exemple, la série synonymique suivante : *nœud atrioventriculaire, nœud auriculoventriculaire, nœud d'Aschoff-Tawara, nœud atrioventriculaire de Tawara, nœud d'Aschoff et Tawara* peut être établie à la suite d'une consultation combinée du corpus et de la documentation complémentaire) ;

– Vérifier l'exactitude des renseignements trouvés dans les autres sources consultées (par exemple, de tels renseignements pouvaient se révéler plus ou moins pertinents, être trop généraux ou porter sur une notion médicale différente) ;

– Fournir de l'information plus technique qui pouvait faire l'objet de note (ex. : durée d'une action exercée par un médicament, effets indésirables provoqués par une substance, affections traitées par un médicament, etc.) ;

– Donner le genre d'un terme, notamment d'un terme dénotant un médicament ;

– Fournir l'explicitation d'un sigle (ex. : *MIBG* est expliqué dans un contexte entre parenthèses) :

Une scintigraphie à la MIBG (méthyl iodo benzo guanidine) est réalisée en cas de suspicion de neuroblastome.

3.4 Recherche d'informations complémentaires

Comme les contextes extraits lors des premières étapes de la recherche se sont révélés insuffisants, il a fallu nous tourner vers d'autres ressources documentaires. Les renseignements recueillis étaient versés au fur à mesure dans une fiche terminologique (gérée par *Adepté*).

À cette étape, nous avons utilisé deux types de catégories de ressources documentaires, à savoir les ouvrages de référence, monographies, revues et encyclopédies (sur support informatique ou papier) et les sites Internet.

Les documents du premier groupe (dictionnaires et ouvrages spécialisés) ont aidé à produire un certain nombre de définitions. Toutefois, la recherche dans certains ouvrages requiert beaucoup de temps. Mentionnons également que les ouvrages de référence (dictionnaires, banques de terminologie) ont rarement permis de retrouver les

termes recherchés tels quels (ce que nous pouvions anticiper, puisqu'il s'agissait de néonymes). Toutefois, ils présentaient le plus souvent des renseignements d'ordre conceptuel sur des termes apparentés (hyponymes, hyponymes ou co-hyponymes).

L'Internet s'est révélé particulièrement utile parce qu'il permet d'atteindre l'information spécialisée rapidement. Il héberge un grand nombre de sites qu'on pourrait qualifier de «généralistes» comme des compilations de revues médicales, des listes de résumés scientifiques, etc. (voir Michel 1998 pour une présentation d'un certain nombre de sites médicaux).

Lorsque les recherches dans les premiers sites se révélaient infructueuses, nous tentions de localiser d'autres sites moins largement connus au moyen des moteurs de recherche (ex. : *Altavista, Copernic, Yahoo*, etc.). Les mécanismes de recherche, bien que parfois rudimentaires (la recherche se fait par simple correspondance entre chaînes de caractères), permettent d'isoler des documents web à partir d'un terme simple ou complexe. Signalons que les requêtes sur les termes complexes ont livré les résultats les plus intéressants. Par ailleurs, nos efforts ont été facilités par le fait que la recherche portait sur des termes médicaux construits à partir de racines grecques et latines. L'ambiguïté souvent générée par les recherches simples sur des chaînes de caractères a été réduite considérablement pour cette raison.

Lorsque les recherches de termes complexes se révélaient vaines, nous reprenions une nouvelle recherche à partir d'une composante d'un terme, notamment à partir de sa tête. Nous repérions fréquemment au moyen de cette méthode des contextes ou des informations sur l'hyperonyme. Fréquemment, ces contextes présentaient des informations sur le modificateur – même s'il n'était pas

mentionné explicitement – qui nous aidaient ainsi à saisir le concept dans sa totalité. Toute information pouvant aider potentiellement la description d'un concept était retenue et stockée.

Comme dans bien d'autres domaines, les sites Internet anglais sont nettement plus nombreux et riches en renseignements. Nous avons repéré ces sites de langue anglaise à partir de l'équivalent anglais d'un terme donné lorsque nous l'avions déniché dans un répertoire (pour les équivalents non relevés tels quels, nous procédions à partir des composantes d'un terme complexe). D'autres termes se sont révélés plus difficiles à traiter de cette manière, soit parce que nous n'avions pas relevé d'indices sur leur traduction, soit parce que l'équivalent anglais repéré avait d'autres acceptions non médicales. Dans les faits, ce sont les sites anglais qui nous ont permis de rédiger la plupart des définitions.

La recherche d'information complémentaire a été beaucoup plus ardue qu'on ne l'aurait pensé au départ. Cette étape a souffert des imperfections liées à la constitution du corpus.

3.5 Rédaction des fiches

Les fiches ont été saisies dans un logiciel d'édition de fiches terminologiques, *Sami-BTQ*⁽¹¹⁾, à partir des renseignements extraits des autres sources documentaires reproduites dans *Adepto-Nomino*. *Sami-BTQ* permet de conserver des versions provisoires et de les modifier tant que les versions définitives ne sont pas prêtes. De plus, ses mécanismes de recherche permettent d'atteindre des termes apparentés et la *BTQ* peut être consultée par l'utilisateur à tout moment.

(11) Pour une présentation détaillée des fonctions du logiciel, voir Pelletier 1996.

Nous avons, dans certains cas, copié les informations utiles à partir d'*Adepto* et nous les avons collées dans *Sami* (les entrées françaises et anglaises). Quant aux définitions et aux notes, elles étaient rédigées directement dans *Sami*.

Sami peut importer directement des fiches produites dans *Adepto*, ce qui évite d'avoir à reproduire les rubriques ou à les coller dans les champs prévus à cette fin. Nous n'avons pas eu recours à cette stratégie puisque nos entrées *Adepto* comportaient de nombreux renseignements dont la saisie s'est

faite sans véritable gestion de l'information dans les champs. En outre, nous trouvions plus utile de conserver tous les renseignements retenus que d'éditer les fiches de dépouillement afin de les rendre compatibles avec la structure des fiches *Sami-BTQ*. Certains renseignements apparaissant dans la fiche de dépouillement qui décrivait un premier terme ont servi à traiter d'autres termes liés sur le plan conceptuel. La figure 3 présente quelques fiches rédigées au terme de la recherche.

Figure 5: Fiches de terminologie

```

1. Conduction cardiaque
FICHER TRAVAIL
005 q010274
025 V7818
030 PUMS
035 UM
040 9812
060 médecine|
105 cardiac conduction
110 E; a
205 conduction cardiaque
210 P5; n f; F; b
215 Propagation d'une impulsion nerveuse qui prend naissance dans le nœud
auriculo-ventriculaire et qui se propage dans les artères entraînant la contraction
de ces dernières. |c|
295 SOURCE ANNÉE VOLUME NUMÉRO MOIS PAGE MENTION
a 605268 1998 12
b 603134 1998 12
c 603135 1998 12

2. Dépolarisation diastolique
FICHER TRAVAIL
005 q011024
025 V7818
030 PUMS
035 UM
040 9809
060 médecine|
105 diastolic depolarization
110 E; a
130 1- phase-four depolarization |E; a|
205 dépolarisation diastolique
210 P5; n f; F; a
215 Diminution de la polarisation de la diastole, c'est-à-dire du mouvement de
dilatation du cœur et des artères, qui alterne avec le mouvement de contraction
(systole). |b|
295 SOURCE ANNÉE VOLUME NUMÉRO MOIS PAGE MENTION
a 603134 1998 10
b 603135 1998 10

```

4 Suggestions d'améliorations à la méthode de travail

En guise de conclusion, nous rappellerons les principales étapes abordées dans le présent article en mettant l'accent sur les lacunes que nous avons observées. Nous suggérons également des éléments qui permettraient d'améliorer la méthode de travail.

Premièrement, au cours des différentes étapes de la recherche, nous nous sommes rendu compte que les textes sélectionnés n'étaient pas parfaitement adaptés à la description de 400 termes complexes du domaine médical. Nous avons donc passé beaucoup de temps à rechercher de l'information complémentaire pour mobiliser les renseignements nécessaires à la préparation des fiches terminologiques.

La recherche aurait gagné en efficacité si les objectifs avaient été bien définis au préalable et si le corpus avait été constitué de façon à les atteindre. Un corpus équilibré peut se révéler difficile à constituer dès le départ, mais ces efforts permettraient d'accroître la productivité, notamment en réduisant le temps consacré à la recherche complémentaire.

Deuxièmement, le dépouillement des textes a reposé sur un traitement automatique (par le logiciel *Nomino*) et nous avons accordé la priorité aux termes complexes. Toutefois, comme nous l'avons vu, l'acquisition de connaissances nécessaires à la description d'un terme complexe passe d'abord et fréquemment, par l'acquisition de connaissances sur un autre terme. Or, comme tous les termes du domaine ne faisaient pas nécessairement l'objet d'un traitement, notre recherche perdait en systématisme. Le travail véritablement terminologique est par excellence

systématique: un domaine de la connaissance humaine ne peut être appréhendé adéquatement que de façon systématique.

Nous croyons dès lors que la décision de réaliser ou non une description systématique doit être prise lors de la planification du travail et cela est directement lié au profil de l'utilisateur qui aura accès à cette terminologie. Ce dernier doit déterminer l'orientation que doit adopter une recherche terminographique. Le terminologue canalise et décrit les concepts d'un domaine et, de ce fait, procède à sa délimitation.

Enfin, le projet a permis de relativiser la croyance pourtant bien répandue voulant que l'utilisation d'outils informatiques accroisse considérablement la productivité. Il est vrai que certains traitements automatiques allègent nettement des tâches fastidieuses. Toutefois, ils modulent de façon sensible des méthodes de travail implantées depuis longtemps. La chaîne de travail terminographique, par exemple, devient plus séquentielle et le terminologue doit s'adapter à ces nouvelles stratégies. Le gain de productivité se fait sentir pour des projets de grande envergure et non pour de petits projets.

*Marie-Claude L'Homme,
Claudine Bodson,
Renata Stela Valente,
Département de linguistique
et de traduction,
Université de Montréal.*

Bibliographie

Auger (P.), Drouin (P.) et L'Homme (M.C.), 1991: «Un projet d'automatisation des procédures en terminographie», *Méta* 36 (1), pp. 121-127.

Auger (P.) et Rousseau (L.-J.), 1978: *Méthodologie de la recherche terminologique*. Québec: L'Éditeur officiel du Québec.

Bergenholtz (H.) et Tarp (S.), 1995: *Manual of Specialised Lexicography*, Amsterdam / Philadelphia: John Benjamins.

Boulanger, (J.C.), 1979: «Problématique d'une méthodologie d'identification des néologismes en terminologie», dans *Néologie et lexicologie. Hommage à Louis Guilbert*, coll. Langue et langage, Paris: Librairie Larousse, pp. 36-46.

Duplain (J.) et coll., 1997: *Guide d'utilisation de la fiche de terminologie*, édition mise à jour, Québec: Gouvernement du Québec.

Lauriston (A.), 1994: «Automatic Recognition of Complex Terms: The TERMINO Solution», dans *Terminology* 1(1), pp. 147-170.

L'Homme (M.C.), 1999: *Évaluation de méthodes d'informatisation de la chaîne de travail terminographique (en vue de la description de néologismes liés au domaine de la médecine). Rapport d'évaluation présenté à l'Office de la langue française* (avec la collaboration de C. Bodson, A. Chudobiak, C.H. Lavigne, S. Milliard, J. Ntakirutimana et R. S. Valente).

L'Homme, (M.C.), 1998: *Expérimentation du logiciel Adepto-Nomino en contexte de veille terminologique. Rapport d'évaluation présenté à l'Office de la langue française* (avec la collaboration de M. de Rooij).

L'Homme, (M.C.), 1996: «Definition of an Evaluation Grid for Term-extraction software», dans *Terminology* 3(2), pp. 291-312.

Meyer (I.) et Mackintosh (K.), 1996: «The Corpus from a Terminographer's Viewpoint», dans *International Journal of Corpus Linguistics* 1(2), pp. 257-285.

Michel (F.), 1998: «Banques virtuelles de textes scientifiques, une aide possible à la veille terminologique?», dans *Terminologies nouvelles* 18, pp. 5-10.

Otman (G.), 1991: «Des ambitions et des performances d'un système de dépouillement terminologique assisté par

ordinateur», dans *La banque des mots* 4 (numéro spécial), pp. 59-96.

Pelletier (S.), 1996: «Le système *Sami-BTQ*: un système intégré de production et de mise à jour des données de la Banque de terminologie du Québec», dans *Terminologies nouvelles* 15, pp. 66-76.

Pelletier (S.), 1995: *Sami. Système d'alimentation et de mise à jour de la Banque de terminologie du Québec*. Québec: Office de la langue française.

Perron (J.), 1996: «*Adepto-Nomino*: un outil de veille terminologique», dans *Terminologies nouvelles* 15, pp. 32-47.

Sager (J.C.), 1990: *A Practical Course in Terminology Processing*. Amsterdam / Philadelphia: John Benjamins.

Sinclair (J.), 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.