

Utilisation d'une banque de textes en terminographie

L'article présente les résultats d'un projet d'expérimentation mené à l'Office de la langue française et visant à vérifier les possibilités et les modalités d'intégration d'un corpus informatisé à la chaîne de travail terminographique. Le projet se situe dans la continuité d'un projet de constitution de banques de textes déjà mené par l'Office dans le contexte de sa participation aux travaux de veille néologique du Rint. Après avoir décrit les logiciels et les méthodes de dépouillement utilisés, on formule des recommandations quant à la constitution et à l'exploitation de banques de textes en terminographie.

Termes-clés:
corpus informatisé;
gestion documentaire; terminotique;
banque de textes; logiciel.

1 Introduction

Le volume de textes disponibles sur Internet, donc en format numérique, est de plus en plus important. Cette situation présente un avantage indéniable du point de vue du terminographe; en effet, le texte écrit constitue la matière première de son travail et l'apparition de nombreux logiciels permettant le traitement informatisé de la langue lui offre la possibilité, pour peu qu'il intègre ces logiciels à ses méthodes, de traiter une quantité imposante de textes.

En 1998, dans le contexte de sa participation aux travaux de veille néologique du Rint, l'Office de la langue française a fait constituer par la Société de mathématiques appliquées (SMA) de Québec deux banques de textes, en imagerie médicale et en robotique industrielle, à partir de documents puisés à même le réseau Internet. L'objectif de l'opération était de procéder à l'analyse de ces banques de textes afin d'alimenter les systèmes de veille terminologique de la *Banque de terminologie du Québec* et de *Balnéo*⁽¹⁾. Les résultats de cette analyse ont amené l'Office à poursuivre l'expérimentation dans le domaine de la robotique industrielle.

Ce sont les résultats de cette nouvelle expérimentation que nous présentons ici: l'expérience visait à tirer d'un nouveau corpus, formé comme le premier de textes spécialisés recueillis sur Internet et constitués en banque textuelle, le plus grand

nombre possible de fiches contextuelles⁽²⁾, tout en analysant les méthodes utilisées pour le faire. On désirait ainsi évaluer plus en détail les possibilités et les modalités d'intégration d'un corpus informatisé à la chaîne de travail terminographique.

L'article précise d'abord les objectifs du projet et les logiciels utilisés, avant de s'attarder à la constitution du corpus et à la description des méthodes de dépouillement appliquées. Il présente ensuite les résultats sous forme de comparaison entre les différentes méthodes de dépouillement pour enfin formuler certaines propositions et recommandations.

2 Expérimentation

2.1 Objectifs

Notre objectif général était de vérifier comment l'utilisation d'une banque de textes informatisée pouvait s'intégrer dans la chaîne de travail d'un terminologue participant à l'alimentation d'un dictionnaire terminologique. Il s'agissait d'abord d'évaluer la pertinence de l'utilisation d'une banque textuelle comme corpus de dépouillement, par opposition, par exemple, à l'exploitation ponctuelle de chacun des textes au fur et à mesure de sa découverte. Il fallait aussi déterminer la meilleure façon d'analyser le corpus ainsi constitué, en tentant de voir comment les méthodes traditionnelles de dépouillement pouvaient être améliorées ou assistées par des outils

(1) Le système *Balnéo*, créé par le Rint, a pour objectifs la collecte, l'échange et la diffusion rapides de néologismes sur Internet: (<http://www.rint.org/cadreIndex7.htm>).

(2) Nous entendons par fiche contextuelle une fiche portant les données suivantes: terme français, contexte(s), source(s). Mentionnons également que les fiches rédigées devaient traiter de notions absentes de la *Banque de terminologie du Québec (BTQ)*.

logiciels. On désirait aussi vérifier la compatibilité et la convivialité des logiciels utilisés.

2.2 Logiciels utilisés

L'expérimentation a mené à l'utilisation d'une série de logiciels correspondant chacun à une étape du travail terminologique. Pour la recherche documentaire sur Internet, nous avons utilisé *Copernic*⁽³⁾, un moteur de recherche qui permet de conserver les archives des recherches déjà faites et qui conjugue les possibilités de plusieurs logiciels de recherche en les lançant simultanément pour ensuite faire la synthèse des résultats. Pour la gestion de la banque, nous avons utilisé l'*Atelier BDT*, logiciel mis au point par la société SMA de Québec, qui permet la constitution et la gestion de banques de textes et qui est doté d'un moteur de recherche (dans le corpus) assez performant. Pour effectuer la transition entre le format utilisé sur Internet et l'*Atelier BDT*, il est nécessaire de se servir d'un logiciel de traitement de texte; nous avons utilisé *Word*. En ce qui a trait à l'analyse textuelle, nous nous sommes servie du logiciel de dépouillement *Adept-Nomino*. Enfin, pour la rédaction des fiches, nous avons effectué le transfert des fiches rédigées avec *Adept-Nomino* vers *Sami*⁽⁴⁾, le logiciel d'alimentation et de mise à jour de la *Banque de terminologie du Québec (BTQ)*. Dans les cas où nous avons dépouillé les textes manuellement, les fiches ont été rédigées directement dans *Sami*.

(3) On peut télécharger une version gratuite de *Copernic* à l'adresse suivante:
<http://www.copernic.com/fr/>

(4) Pour plus d'informations sur *Sami*, voir Pelletier (1996).

2.3 Constitution du corpus

Nous avons constitué, à l'aide du moteur de recherche *Copernic*, une banque d'environ 73 000 mots, dans le domaine de la robotique. Nous avons privilégié les textes rédigés en français émanant de centres de recherche universitaires ou gouvernementaux (rapports d'activités, thèses, projets de recherches, articles, etc.). En effet, ce type de texte nous paraissait offrir le plus de potentiel pour la recherche de termes nouveaux, par la crédibilité de leurs auteurs – des chercheurs de haut niveau – parce qu'ils ont été rédigés en français (et non traduits) et parce qu'ils traitent de domaines de pointe.

Nous reviendrons plus loin sur l'importance à accorder à l'étape de constitution du corpus de dépouillement. Signalons simplement que nous n'avions pas pour ambition de former un corpus se suffisant à lui-même, c'est-à-dire qui contiendrait suffisamment d'informations provenant de sources assez diversifiées pour servir à lui seul de documentation pour la rédaction des fiches terminologiques. Notre but était plutôt de constituer une banque de textes servant à trouver des termes absents de la *BTQ*, et dont la nature permettrait de recueillir des contextes très « parlants », pouvant guider le terminologue abordant ces notions pour compléter les fiches (c'est-à-dire collecter les données conceptuelles nécessaires à la rédaction d'une définition et éventuellement de notes) et lui permettre de s'engager avec confiance dans la bonne direction pour la recherche d'informations supplémentaires.

Une fois les textes trouvés, encore fallait-il les constituer en corpus, ce que nous avons fait grâce à l'*Atelier BDT*. Nous reviendrons plus loin sur les manipulations liées à cette opération.

2.4 Dépouillement

Traditionnellement, lorsqu'il s'agit d'établir une nomenclature pour un domaine spécialisé, le terminologue s'arme de son crayon et de sa patience et entreprend la lecture de textes récents publiés dans le domaine qui l'intéresse. Au fil de cette lecture, il se familiarise avec le domaine en question et relève ce qu'on appelle des candidats-termes, c'est-à-dire des unités simples ou syntagmatiques qui semblent se distinguer par leur caractère d'univocité et de spécificité. Le terminologue relève donc ces unités avec leur contexte et, éventuellement, les informations terminologiques qu'il déniche dans le texte et qui lui permettront plus tard de rédiger définition et notes. Cette étape, le dépouillement, gourmande de temps, devra être suivie d'une recherche exhaustive permettant de vérifier les emplois du terme, sa fréquence, ses équivalents (dans le cas d'une banque multilingue), etc. Un premier problème soulevé à l'étape du dépouillement est donc un problème de temps.

Une autre difficulté relative au dépouillement est liée à l'absence de nomenclature de départ: en effet, en veille, l'objectif est d'abord d'établir une telle nomenclature, puisqu'on est alors à la recherche de termes que l'on n'a pas encore traités, c'est-à-dire relevés, attestés, consignés, définis, etc. On ne sait donc pas exactement ce qu'on cherche, et la démarche fait appel à une bonne dose d'instinct et d'expérience; les résultats obtenus dépendent donc beaucoup de l'expérience du terminologue et de sa connaissance du domaine traité.

On a tenté, lors de l'expérimentation décrite ici, de vérifier s'il était possible de lever ces difficultés en automatisant l'étape du dépouillement des textes. Pour cela, nous avons choisi de comparer trois

approches; la comparaison s'est faite sur le plan du temps nécessaire pour produire un certain nombre de fiches. Un tableau résume cette comparaison dans la section *Analyse*.

La première approche, que nous avons appelée *dépouillement classique*, consistait à imprimer les textes et à souligner à la main les candidats-termes pour relancer ensuite ces termes dans la banque de textes afin de trouver les meilleurs contextes. Remarquons que, malgré son nom, cette démarche se distingue déjà d'une méthode plus traditionnelle puisqu'elle utilise, même si ce n'est que partiellement, les possibilités d'un logiciel comme l'*Atelier BDT* pour la recherche des contextes; pour suivre vraiment ce que nous avons décrit plus haut comme la méthode de travail habituelle du terminologue, il aurait fallu imprimer tous les textes et recueillir, en les recherchant visuellement, les meilleurs contextes pour chaque terme retenu. Cependant, notre façon de procéder a quand même donné des résultats assez probants lors de la comparaison avec les deux autres méthodes.

La deuxième approche est intermédiaire: dans ce que nous avons appelé *dépouillement semi-automatique*, il s'agissait d'établir une nomenclature d'éléments de base susceptibles de donner des termes complexes (par exemple *capteur* qui donnera entre autres *capteur à ultrason*) ou de racines susceptibles de fournir des termes simples et complexes (comme l'élément tronqué *holon** qui ramènera *holonomie*, *holonome*, *robot holonome*, *non-holonome*, etc.) et de lancer chacun des termes de cette nomenclature dans la banque de textes pour recueillir occurrences, cooccurrences et contextes. Il s'agit de la méthode qui correspond le mieux à l'optique dans laquelle l'*Atelier BDT* a été constitué: ce logiciel est en effet équipé d'un moteur de recherche

intéressant et performant qui permet de trouver aisément tous les contextes et cooccurrences d'un terme donné. Il suffit de lancer un terme pour trouver ses composés, des exemples de son utilisation, des contextes. Cependant, le logiciel ne fournissant pas à ce jour d'index, la principale contrainte liée à son utilisation vient du fait qu'il faut savoir quoi demander au moteur de recherche pour obtenir des résultats. Il est donc nécessaire de disposer d'une nomenclature de départ; or, comme nous l'avons déjà souligné, en veille, l'objectif est justement d'établir une telle nomenclature. Une des façons de contourner le problème est de lancer dans la banque de textes des termes de base (par exemple, pour le domaine de la robotique, des termes comme *robot*, *commande*, *manipulateur*, etc.) qu'on aura recueillis lors de lectures préliminaires. Cependant une telle méthode ne permet pas de trouver des termes composés à partir de bases auxquelles on n'aura pas songé, sinon par hasard. Si on veut agir de façon plus méthodique, il faudra donc aller plutôt vers la troisième méthode, le *dépouillement automatique* qui consiste simplement à faire analyser les textes de la banque par un logiciel de dépouillement textuel, ici *Adepte-Nomino*.

2.5 Résultats

Les résultats sont groupés selon la méthode utilisée, ce qui permettra plus loin la comparaison entre ces méthodes.

2.5.1 Dépouillement classique

Nous n'avons effectué que très peu de dépouillement selon cette méthode: elle a été utilisée surtout dans le but d'obtenir une référence, un point de comparaison avec les autres méthodes, étant donné que c'est celle qui «colle» le mieux à ce

que font traditionnellement les terminologues en situation de veille terminologique.

À partir de deux textes choisis au hasard dans la banque de textes, puis imprimés, nous avons souligné les candidats-termes, avant de constituer une fiche contextuelle pour chacun d'entre eux en les relançant dans la banque de textes pour trouver le ou les meilleurs contextes, après avoir vérifié leur absence de la BTQ. En une heure, 11 fiches contextuelles ont ainsi été rédigées.

2.5.2 Dépouillement semi-automatique

Rappelons que pour utiliser cette méthode il faut disposer d'une nomenclature de départ; nous avons donc choisi de lancer les 15 composantes des 11 termes trouvés avec la première méthode⁽⁵⁾, ce qui a permis d'isoler 62 nouvelles notions et de rédiger les fiches contextuelles correspondantes en environ 8 heures. Rappelons qu'il s'agissait ici de lancer chacun des éléments de base dans la banque de textes, de relever chacun des candidats-termes formés à partir de ces éléments, puis de relever le ou les meilleurs contextes pour chacun de ces termes avant de rédiger la fiche contextuelle correspondante dans *Sami*.

(5) Nous entendons par *composante* chacun des éléments de base d'un terme: si on a trouvé par exemple *capteur à ultrason*, on relancera *capteur* et *ultrason*.

(6) Nous avons en premier lieu tenté l'analyse de l'ensemble des textes placés dans la banque de textes avec *Tact*, logiciel mis au point à l'Université de Toronto (et dont on peut télécharger une version Web à l'adresse suivante: <http://tactweb.humanities.mcmaster.ca/>). *Tact* fournit une liste des mots trouvés dans le texte, accompagnés de leur fréquence d'utilisation; il faut, à partir de cette liste, sélectionner les termes dont on pense qu'ils pourraient donner des résultats intéressants (soit qu'il s'agisse de termes simples, ou d'éléments de base de termes complexes) et les marquer un par un avant de demander à *Tact* de produire une liste de cooccurrences ou de contextes. Pour notre banque de textes, *Tact* a produit une liste d'environ 10 000 mots, qu'il aurait fallu lire et sélectionner, sans l'aide, à cette étape du traitement, d'un contexte pour décider de la pertinence de chacun d'eux. Devant l'ampleur de la tâche, nous avons décidé de tenter immédiatement un essai avec *Adept-Nomino*. Il faut également souligner que les manipulations préliminaires à l'analyse du texte sont peu conviviales avec *Tact*, qui fonctionne en environnement Dos, avec lequel la plupart des utilisateurs sont peu familiers.

(7) Il faut noter que la mise en forme des textes faite automatiquement par l'*Atelier BDT* nuit à l'utilisation d'*Adept-Nomino*, à cause des différents marqueurs qui y sont introduits; pour obtenir les meilleures performances d'*Adept-Nomino*, nous avons donc dû partir des textes dans leur format «original», c'est-à-dire sauvegardés en format *texte seulement* par *Word*.

(8) Ne sont décrites ici que les «sorties» d'*Adept-Nomino*: pour une description plus détaillée de son fonctionnement, voir Perron (1996: 32).

(9) Le temps nécessaire à ce transfert n'a pas été comptabilisé. Dans le cadre de ce projet, il a été effectué par les services informatiques de l'Office; c'est une opération très rapide (à titre d'exemple, il faut moins d'une minute pour verser 170 fiches).

2.5.3 Dépouillement automatique⁽⁶⁾

Nous avons lancé *Adept-Nomino*, à partir encore une fois de l'ensemble des textes placés dans la banque⁽⁷⁾. Après une analyse (qui dure entre 20 et 25 minutes pour une banque de textes telle que la nôtre), *Adept-Nomino* produit cinq listes⁽⁸⁾: les unités complexes nominales (ucn), les unités complexes nominales additionnelles (ucna), une liste des verbes, une liste des adjectifs et une liste des noms. C'est surtout à partir de la liste des ucn et de la liste des noms que nous avons identifié des candidats-termes.

Lorsqu'on interroge n'importe lequel des éléments des listes fournies par *Adept-Nomino*, le logiciel le replace dans son contexte (autant de fois qu'il y a d'occurrences). Il est donc possible de choisir immédiatement les contextes qui paraissent les meilleurs et de les transférer directement dans le champ approprié. En environ trois heures et demie, 89 candidats-termes ont été recueillis et leurs fiches contextuelles rédigées (après vérification dans la BTQ). Le résultat se trouvait sous la forme de fiches au format *Adept-Nomino*, que nous avons versées dans *Sami* par la suite⁽⁹⁾.

Analyse des résultats

Le tableau 1 résume les données relatives au nombre de fiches et au temps nécessaire à leur production pour chaque méthode:

Notons d'abord que malgré son apparente incongruité (il est moins élevé que le résultat du dépouillement classique), le résultat correspondant au dépouillement semi-automatique est tout à fait normal; en effet, comme nous l'avons mentionné dans la section *Résultats*, il s'agit d'une méthode en quatre étapes:

1. Établir une nomenclature de base;
2. Lancer les éléments de cette nomenclature dans la banque de textes à l'aide du moteur de recherche de l'*Atelier BDT*;
3. Recueillir les candidats-termes complexes produits par cette recherche;
4. Pour chacun d'entre eux, rechercher le ou les meilleurs contextes.

On pourrait bien sûr recueillir les contextes (étape 4) en même temps qu'on identifie les termes complexes (étape 3), mais il semble que cette façon de faire soit plus compliquée et produise plus de «bruit»; on risque en effet de se retrouver avec des contextes inutiles qu'on devra éliminer par la suite. C'est pourquoi nous avons choisi d'établir d'abord la liste des termes complexes pour ensuite faire une recherche de contextes (cette recherche est d'ailleurs plus restreinte que la précédente, puisqu'on ne lance dans la banque de textes que le composé qui nous intéresse).

La méthode la plus performante semble être de loin, au vu des résultats reproduits dans le tableau, celle qui fait intervenir un logiciel de

Tableau 1

Méthode	Nombre de fiches	Temps nécessaire	Nombre de fiches à l'heure
Classique	11	1 heure	11
Semi-automatique	62	8 heures	7,75
Automatique	89	3 heures et demie	24,4

dépouillement comme *Adepte-Nomino*. Nous pensons qu'une telle performance s'explique par la rapidité et l'efficacité permises par l'ordinateur, combinées à la liberté laissée au terminologue; en effet, avec une telle méthode, c'est le logiciel qui propose les candidats-termes, et c'est le terminologue qui dispose de ces propositions comme il l'entend. On pourrait objecter que le terminologue, avec une telle méthode, ne lit plus les textes, et donc n'est plus aussi familier avec son domaine. À ceci nous répondrons que la performance d'un outil de dépouillement automatique de textes donne au contraire plus de temps au terminologue pour lire, et qui plus est pour lire des textes imprimés qui, de toute façon, n'auraient pas pu être analysés de façon automatique (à moins d'être d'abord numérisés, ce qui pose des problèmes différents). Nous dirions que le logiciel s'occupe des textes disponibles sous format numérique, et l'être humain des textes imprimés. De plus, c'est le terminologue qui a choisi les textes à faire analyser (et donc qui les a au moins parcourus) et c'est lui qui analyse chaque contexte proposé par *Adepte-Nomino* pour juger de sa pertinence, ce qui contribue de façon appréciable à sa familiarisation avec le domaine.

3 Propositions et recommandations

3.1 Constitution d'un corpus

La qualité et la diversité des textes choisis constituent l'assise d'un travail terminologique basé sur le corpus; elles déterminent en effet en grande partie l'utilité et la pertinence du résultat final. L'établissement de critères régissant le choix de textes destinés à faire partie d'un corpus de dépouillement nous paraît donc essentiel; la poursuite d'essais portant

sur la constitution de banques textuelles permettrait d'établir des orientations méthodologiques aptes à guider le terminologue lors de cette étape primordiale qu'est la mise sur pied de son corpus de dépouillement.

3.2 Utilité d'une banque de textes

Dans l'optique où l'on considère qu'une banque de textes doit être «rentable», c'est-à-dire que le temps qu'on a investi dans sa constitution doit avoir été utilement dépensé, il est permis de se poser des questions sur l'utilité d'une telle banque. Nous soulignerons simplement ici deux faits en réponse à ces doutes: en situation de veille terminologique, où on ne dispose pas de nomenclature de départ, l'expérimentation décrite ici fait ressortir le rapport entre le temps investi et les résultats. Pour une demi-journée passée à récolter des textes sur Internet, on peut tirer, en deux jours environ, plus de 400 fiches contextuelles⁽¹⁰⁾ de la banque ainsi constituée. L'expérimentation, bien qu'elle n'ait pas directement porté sur une situation de mise à jour du fonds terminologique, a également fait ressortir, de façon sous-jacente, le fait suivant: on a dû éliminer une certaine quantité de termes qui étaient déjà dans la BTQ; étant donné la qualité des textes stockés (consistant, rappelons-le, en grande partie en rapports de recherche universitaires), il paraît donc probable que pour ces termes la banque de textes aurait pu servir de ressource documentaire fort utile (jointe aux ressources habituelles) en situation de mise à jour. Il faudrait mener d'autres expérimentations, avec des corpus plus importants, pour vérifier cette hypothèse.

3.3 Manipulations informatiques

Il faut mentionner ici que les opérations de stockage des textes dans l'*Atelier BDT* comprennent des manipulations informatiques (création et initialisation de la banque, création d'un nouveau répertoire, puis marquage, segmentation et importation des textes, mise à jour de la page d'accueil) susceptibles de rebuter de nombreux terminologues peu familiers avec les dessous de la manipulation de fichiers, ce qui a amené certains terminologues à s'interroger sur les étapes de la constitution d'une banque de textes (Michel, 1998:10): «Qui doit faire les recherches menant à sa constitution? Les bibliothécaires sont-ils les mieux qualifiés pour ce genre de démarche? Comment en arriver à jumeler l'expertise du terminologue familiarisé avec les sites les plus intéressants et celle des documentalistes?»

En réponse à ces questions, nous proposons ici⁽¹¹⁾ un mode de formation d'une banque de textes – avec l'*Atelier BDT* – qui décharge le terminologue des tâches qui sont par trop étrangères à son travail habituel et qui, lui paraissant lourdes et malaisées, pourraient le décourager de

(10) On considère le travail accompli par un terminologue habitué à la procédure et utilisant un outil de dépouillement automatique; comme pour toute nouvelle technique, les premières expériences seront peut-être moins «rentables»...

(11) On trouvera en annexe une proposition détaillée de partage des tâches qui ne demande plus au terminologue que de sauvegarder sous le bon format le texte qu'il a trouvé en ajoutant la date de consultation et l'adresse du site.

mettre sur pied cette banque, tout en tenant pour acquis qu'il est le mieux placé pour repérer les meilleurs textes dans son domaine de recherche.

3.4 Exploitation des textes

Une autre préoccupation des terminologues portait sur la pertinence de consacrer du temps à intégrer un texte intéressant à une banque textuelle au lieu d'en extraire au moment même toutes les informations utiles. Cette question nous amène à rappeler ici le concept de *banque évolutive*: un terminologue, tant qu'il travaillerait dans le même domaine, continuerait d'enrichir sa banque et la mettrait éventuellement en commun avec ceux de ses collègues qui travaillent dans un domaine voisin du sien. Cette façon de voir est inspirée par l'observation suivante: l'expérience démontre qu'une notion ayant fait l'objet d'une recherche sur Internet «ramène» souvent plusieurs autres notions avec elle; le texte ainsi trouvé risque donc de s'avérer utile ultérieurement et il est parfois difficile de retrouver un site Internet (même s'il est marqué d'un signet – on court toujours le risque qu'il soit devenu inaccessible ou qu'il ait été déplacé) alors qu'il est possible de le conserver sur son propre disque dur, et qui plus est, avec les possibilités qu'offre le moteur de recherche d'un logiciel de gestion documentaire. De plus, le terminologue, au fur et à mesure qu'il fait connaissance avec son domaine, sait de mieux en mieux quoi chercher et où le chercher; sa banque de textes gagne donc en qualité au fil du temps, et peut devenir une source inestimable d'informations dans un domaine donné. Elle peut servir ultérieurement à des opérations de mise à jour, à familiariser un nouveau terminologue avec un domaine donné, à faire des études d'évolution de la terminologie dans une sphère

d'activité, etc. On pourrait objecter, en se référant au phénomène de saturation déjà souligné par plusieurs auteurs (voir par exemple Otman 1991b: 61), qu'on observe qu'à partir d'une certaine taille (au demeurant difficile à déterminer), un corpus de textes ne fournit plus beaucoup de terminologie nouvelle. Ce phénomène nous semble s'appliquer de façon plus marquée dans certaines situations: dans les cas de corpus en langue générale par exemple, ou dans un domaine qui connaît une évolution plutôt lente. Dans les domaines spécialisés identifiés comme des priorités par l'Office, ce phénomène ne paraît cependant pas représenter un obstacle à la constitution d'une banque évolutive, pouvant au contraire servir à indiquer que le domaine est à peu près couvert lorsque l'ajout de nouveaux textes n'apporte plus ou presque plus de terminologie nouvelle. Dans tous les cas, de nouvelles expérimentations pourraient fournir des indicateurs sur la taille optimale d'un corpus en langue spécialisée.

3.5 Convivialité et compatibilité des logiciels

Comme nous l'avons brièvement mentionné plus haut, en ce moment l'Atelier BDT et *Adepté-Nomino* ne sont pas pleinement compatibles; les textes une fois marqués par l'*Atelier* deviennent inutilisables (ou posent du moins plusieurs difficultés d'utilisation) par *Adepté-Nomino*. C'est ce problème relatif à la compatibilité entre les logiciels qui nous amène à formuler, après discussion avec les concepteurs de l'*Atelier BDT*, deux recommandations dont voici la première (qui constitue une étape vers la mise en œuvre de la situation idéale): munir l'*Atelier BDT* d'un index interactif qui permettrait d'interroger n'importe quel mot du

texte pour le replacer dans son contexte et éventuellement découvrir ses composés. Ceci éliminerait (ou du moins réduirait) le besoin d'une nomenclature de départ. Mais la solution qui nous semble la meilleure est celle-ci: mettre au point une interface de communication entre l'*Atelier BDT* et *Adepté-Nomino*, qui permettrait d'analyser directement en tout ou en partie la banque de textes constituée grâce à l'*Atelier BDT*. De plus, comme nous y avons fait allusion plus tôt, l'interface entre *Adepté-Nomino* et *Sami* existe déjà; une nouvelle interface entre l'*Atelier BDT* et *Adepté-Nomino* résulterait donc en un outil intégré constitué d'un logiciel de gestion documentaire, d'un logiciel de dépouillement et d'un logiciel d'alimentation d'une banque de terminologie, outil qui s'inscrirait aisément dans la chaîne de travail terminographique.

3.6 Proposition de chaîne de travail informatisée en situation de veille

Même s'il évolue assez rapidement, le travail du terminologue-terminographe comporte toujours les mêmes grandes étapes (Auger, 1994: 50): «1) préparation du travail terminologique [...] (exploration du domaine, recherche de documentation, sélection et enregistrement des sources); [...] 2) travail terminographique proprement dit [...] (mise en forme et traitement d'un corpus de textes, dépouillement, établissement de la nomenclature, extraction de données terminologiques et paraterminologiques, [...] etc.); 3) stockage et traitement des données recueillies [...]; 4) édition et diffusion du produit final (dictionnaire terminologique) [...]»

Ce que nous proposons ici, c'est de mieux appuyer certaines de ces

Tableau 2

Étape	Logiciel
Recherche de textes pertinents sur Internet	<i>Copernic</i>
Constitution d'une banque de textes	<i>Microsoft Word</i> et <i>Atelier BDT</i>
Extraction des termes	<i>Adepte-Nomino</i>
Extraction des contextes, notes, définitions, etc.	<i>Adepte-Nomino</i>
Rédaction des fiches	<i>Sami</i>

étapes à l'aide de logiciels permettant d'automatiser en partie la tâche du terminologue.

Le projet d'expérimentation mené à l'Office de la langue française concernait surtout les premières étapes du travail terminologique, soit la préparation du travail par la recherche documentaire et la sélection de sources, de même que le travail d'extraction des données terminologiques. Le tableau 2 présente sous forme schématique la chaîne de travail proposée, avec pour chaque étape un exemple de logiciel disponible⁽¹²⁾.

4 Conclusion

Il est apparu lors de cette expérimentation que l'utilisation des outils informatiques étudiés pouvait rendre plus agréable et plus efficace le travail du terminologue. À partir de logiciels existants, comme l'*Atelier BDT* et *Adepte-Nomino*, il est possible de mettre au point des outils sur mesure parfaitement adaptés au travail à accomplir, surtout en situation de veille, mais aussi pour la mise à jour, où ils s'ajouteraient aux ressources déjà disponibles. Il y a bien sûr tout un travail d'adaptation à accomplir, autant des outils aux personnes que l'inverse. Cette adaptation demande du temps et risque même de ralentir la production pendant une période donnée, où

chacun doit apprivoiser les nouveaux outils mis à sa disposition. Mais dans un monde où l'information est de plus en plus considérée comme une richesse, il est essentiel de se préoccuper de trouver des moyens de la conserver sous la meilleure forme et de la traiter le plus efficacement possible. Les terminologues sont des spécialistes du traitement de l'information sous bien des aspects; il leur revient donc de se tenir au courant des possibilités offertes par Internet et par l'informatique en général afin d'exploiter au mieux ces possibilités; ils demeureront ensuite les meilleurs juges des outils les mieux adaptés à leur travail.

*Anouk Jaccarini,
Office de la langue française,
Québec.*

(12) Il faut souligner qu'il existe de nombreux autres logiciels ayant les mêmes fonctions ou des fonctions semblables; on s'en tient ici à ceux qui ont servi au projet.

Bibliographie

Auger (Pierre), 1994: «Les outils de la terminotique: typologie des logiciels d'aide à la terminologie et/ou d'automatisation de la chaîne de travail en terminographie» dans *Terminologies nouvelles*, n° 11, p. 46-52.

Côte (Normand): «Formation aux banques virtuelles de textes – commentaires de l'équipe du tronc commun industriel», *document interne*, Office de la langue française, mars 1999.

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

Otman (Gabriel) 1991a: «Aspects de l'informatisation des activités terminologiques et traductionnelles» dans Clas (A.) et Hayssam Safar, dir., *L'environnement traductionnel: la station de travail du traducteur de l'an 2001 – Actes du colloque de Mons, 1991*, Sillery, Presses de l'Université du Québec, p. 213 à 221.

Otman (Gabriel) 1991b: «Des ambitions et des performances d'un système de dépouillement terminologique assisté par ordinateur» dans *La banque des mots*, numéro spécial 4/1991, p. 59-96.

Pelletier (Sylvie) 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*, n°15, p. 66-76.

Perron (Jean), 1996: «*Adepte-Nomino*, un outil de veille terminologique» dans *Terminologies nouvelles*, n°15, p. 32-47.

Annexe

Proposition de partage du travail pour la constitution d'une banque de textes avec l'*Atelier BDT*⁽¹³⁾.

FRÉQUENCE	TEMPS	RESPONSABLE DE LA DOCUMENTATION	TERMINOLOGUE
lors de la création d'une nouvelle banque de textes	5 minutes	<ul style="list-style-type: none"> – sous l'<i>Atelier BDT</i>, crée une nouvelle banque – informe le terminologue du répertoire où il doit placer ses textes 	
variable (hebdomadaire, mensuelle)	une demi-journée pour environ 35 000 mots		– copie les textes qu'il trouve pertinents sous format <i>WP5.0</i> (en ajoutant l'adresse du site et la date de consultation) dans le répertoire indiqué par le responsable de la documentation
sur notice du terminologue, afin que les textes deviennent disponibles le plus rapidement possible	environ 2 minutes par texte	<ul style="list-style-type: none"> – sous l'<i>Atelier BDT</i>, marque (adresse Internet, titre et titre abrégé) et segmente les textes – importe les textes dans la banque – met à jour le fichier <i>bdt.htm</i> (page d'accueil de l'<i>Atelier BDT</i>) – vérifie la banque (sous <i>consultation</i>) 	
			– la banque devient disponible et prête à être utilisée par le terminologue

(13) Note: dans ce texte, nous nommons *responsable de la documentation* la personne (technicien ou documentaliste) qui serait chargée, entre autres, d'assister le terminologue dans la constitution d'une banque de textes.