

RECHERCHE TERMINOGRAPHIQUE EN TRADUCTION JURIDIQUE RÉALISÉE AU MOYEN D'OUTILS INFORMATIQUES

Danielle VERRIER

Traductrice agréée ♦ Terminologue agréée, Canada

1. INTRODUCTION

Je n'insisterai jamais assez sur l'avènement en tant que tel de la terminotique, issue de la rencontre de " la terminologie appliquée et de l'informatique " (Auger *et al.*, 1991 : 121).

1.1 Informatique, terminologie et traduction juridique

Un véritable processus d'automatisation s'est amorcé il y a quelques années. Sont alors apparus les outils relatifs aux tâches terminographiques comme les correcteurs syntaxiques, orthographiques et grammaticaux; les concordanciers, les dépouilleurs de syntagmes et autres outils de dépouillement assisté par ordinateur (DAO).

En outre, l'environnement DOS fut remplacé par l'environnement Windows, ce qui a facilité l'utilisation des outils de terminotique et l'apparition des supports de stockage de grande capacité. Internet a également fait son apparition. " Ce n'est qu'au début des années 1980, avec l'implantation et la diffusion de la micro-informatique (...) qu'on assiste à la naissance de termes comme : *terminologie informatique, terminologie assistée par ordinateur et terminotique* " (Auger *et al.*, 1991 : 121).

Mon exposé s'inscrit dans cette problématique. En effet, j'ai utilisé le plus possible d'outils informatiques dans ma démarche de consignation de la terminologie propre au domaine de la traduction juridique.

1.2 L'objet de l'exposé

Je vais donc décrire la méthode de confection de fiches terminologiques bilingues dans le domaine juridique et j'ai eu recours, pour ce faire, à des méthodes de travail informatisées. Qui plus est, j'en ai profité pour faire l'essai de Termplus, logiciel de dépouillement surtout utilisé pour des textes de spécialité.

1.3 Les objectifs de l'exposé

La pratique de la traduction juridique exige non seulement la consignation de termes, mais aussi l'automatisation maximale des méthodes de travail en terminographie¹.

L'exposé vise à décrire un travail dont les objectifs, en cette ère de productivité et de rentabilité, étaient les suivants :

- 1) vérifier jusqu'à quel point le processus des étapes de la chaîne de travail en terminographie est accéléré grâce à l'automatisation, notamment lorsque l'on a recours au dépouillement assisté par ordinateur;

- 2) faire correspondre les méthodes de travail traditionnelles (celles de l'humain) à celles du processus informatisé (utilisation de la machine) et indiquer la place de plus en plus importante occupée par la terminotique² (L'Homme, 1996 : 45);
- 3) expérimenter l'outil d'extraction terminologique Termplus dans un corpus bilingue constitué de textes spécialisés, soit des textes juridiques, et dépister les descripteurs complexes dans cette langue de spécialité.

1.4 Le contenu de l'exposé

Dans un premier temps, je dresserai un parallèle entre la méthodologie traditionnelle et la méthodologie informatisée dans le domaine de la terminologie, la chaîne de travail en terminographie s'étant grandement métamorphosée au cours des dernières années.

Dans un deuxième temps, je décrirai les étapes d'une recherche terminologique systématique informatisée dans le domaine juridique et leur degré d'automatisation.

En dernier lieu, après avoir en quelque sorte reconstitué la " terminogenèse " de la pratique terminographique, je tirerai une conclusion sur l'ensemble de l'expérience, notamment sur l'utilisation du dépouilleur terminologique (DT).

2. LA CHAÎNE DE TRAVAIL EN TERMINOGRAPHIE

2.1 Méthodologie traditionnelle

Je vous présenterai une vue d'ensemble des méthodes de travail traditionnellement utilisées en terminologie. Ensuite, j'établirai un parallèle entre la méthodologie traditionnelle naguère utilisée et la méthodologie informatisée.

Je ferai un survol des différentes étapes de la méthodologie traditionnelle, soit : 1) la recherche de documentation; 2) le dépouillement terminologique; 3) le repérage provisoire de termes et le découpage des unités; 4) le choix définitif de termes comprenant l'établissement de la nomenclature, la collecte de données sur chaque terme, le découpage et la sélection de contextes; 5) le traitement manuel des contextes; 6) la rédaction de la fiche terminologique.

2.1.1 Recherche de documentation

Cette importante étape précède le travail de terminologie. En fait, il s'agit de choisir les documents qui serviront au dépouillement et d'établir ainsi le corpus de base. Il est donc nécessaire de trouver des textes représentatifs du domaine à étudier et de ses sous-domaines.

2.1.2 Dépouillement terminologique

Le terminologue lit donc systématiquement le corpus et indique manuellement les termes à consigner. Il les " repère ". Il parcourt de ce fait une première fois chacun des textes et souligne les unités terminologiques à extraire.

2.1.2.1 Repérage provisoire de termes et découpage des unités

Le terminologue part en éclaireur pour repérer les termes intéressants qu'il " découpera " ensuite, selon l'expression de Rondeau (1984 : 78).

Il faut donc, lors de cette opération, effectuer l'extraction des termes provisoires en fonction de leur pertinence par rapport au sujet visé; il s'agit de l'" élagage " du texte.

2.1.2.2 Choix définitif de termes

Le terminologue s'assure, notamment en recherche thématique, de tenir compte des termes revêtant un intérêt particulier eu égard au domaine visé. **Les termes sont choisis en raison de leur essentialité.**

2.1.2.3 Établissement de la nomenclature

Le terminologue fait un examen minutieux des textes et établit finalement une nomenclature, soit un " répertoire de termes présentant des relations notionnelles fortement structurées et correspondant à des règles systématiques de dénomination " (Boutin-Quesnel *et al.*, 1985 : 31).

2.1.2.4 Collecte de données sur chaque terme

Selon Auger (1978 : 32), il faut recueillir toutes les données nécessaires au traitement de la nomenclature, soit : la catégorie lexicale et grammaticale, le domaine, une définition, des contextes (terminologiques, langagiers ou métalinguistiques) et des illustrations, le cas échéant.

2.1.2.5. Découpage et sélection de contextes

Boutin-Quesnel *et al.* (1985 : 26) expliquent très bien le découpage terminologique qui consiste à identifier le statut terminologique d'une unité extraite d'un énoncé. La notion véhiculée par un terme, simple³ ou complexe, se trouve ainsi délimitée.

Une unité terminologique sera mieux comprise grâce à un "environnement" contextuel.

La plupart du temps, le terminologue relève plusieurs contextes dans le but d'en faire ensuite le tri. Il choisit alors le ou les contextes les plus pertinents, soit

" (...) l'environnement linguistique d'un terme dans un énoncé. Il s'agit le plus souvent d'un segment de phrase, d'une ou de plusieurs phrases juxtaposées où le terme étudié est occurrent (...) Il est essentiel que le contexte relevé présente une attestation exacte de l'unité étudiée " (Auger, 1973 : 34).

2.1.3 Traitement manuel des contextes

Les contextes doivent " renseigner " sur la notion traitée et " éclairer " le lecteur de la fiche. Le terminologue relit alors les diverses données recueillies.

2.1.3.1 Analyse des résultats et synthèse

Dubuc⁴ explique clairement le recoupement interlangue des notions. Une fois la nomenclature bien établie en langue de départ (LD), les unités et les contextes dûment découpés, le travail en langue d'arrivée (LA) n'a pas à se faire d'une façon étroitement cloisonnée. En effet, au fur et à

mesure des dépouillements en langue d'arrivée, le terminologue peut commencer l'appariement interlangue des termes, quitte à noter les termes sans équivalent en LD, pour les traiter ultérieurement en poussant ses recherches en LD.

Chaque nomenclature fait l'objet d'une dernière analyse terminologique avant l'établissement d'une nomenclature bilingue qui constitue en fait un point d'aboutissement.

2.1.4 Rédaction de la fiche terminologique

La consignation définitive des divers renseignements sur une fiche terminologique est faite dans des champs préétablis et les composantes de la fiche varient selon le terminologue ou l'entreprise.

2.2 Parallèle entre la méthodologie traditionnelle et la méthodologie informatisée

Le travail du terminologue a beaucoup évolué. Le cheminement méthodologique est à peu près le même, sauf que le travail est de plus en plus effectué au moyen de techniques nouvelles.

2.3. Méthodologie informatisée

La chaîne de traitement de la recherche terminographique a subi des métamorphoses sur le plan de l'automatisation et les tâches du terminologue ont, par conséquent, été grandement informatisées. J'expliciterais donc la méthodologie informatisée utilisée.

2.3.1 Collecte et choix des textes (mise en forme du corpus) à dépouiller en fonction du domaine de spécialité

Il est essentiel de recueillir des documents électroniques en fonction du domaine étudié et de mettre beaucoup de soin à les choisir. La qualité des textes à dépouiller est primordiale car elle aura une incidence sur la valeur de la recherche terminologique.

2.3.1.1 Téléchargement ou saisie optique

De plus en plus, les textes sont téléchargés à partir d'Internet ou déjà stockés sur support électronique. Sinon, il faudra numériser les documents.

2.3.1.2 Conversion des textes

Il faudra souvent convertir les textes du corpus dans un format compatible avec les outils utilisés par le terminologue.

2.3.2 Analyse du texte et dépouillement de termes grâce à un logiciel de dépouillement assisté par ordinateur (DAO)

Le passage de l'ère traditionnelle à l'ère informatisée a vu naître une nouvelle technologie. L'ordinateur peut dorénavant analyser des corpus du début jusqu'à la fin. Les fichiers-textes font alors l'objet d'un dépouillement assisté par ordinateur (DAO). Le contenu textuel est alors lu et traité par un dépouilleur et il en résulte une liste de mots ou de termes. Le terminologue n'a donc pas à lire intégralement les textes comme avant.

Mentionnons, en outre, que le DAO consiste à extraire, au moyen d'un logiciel, le contenu terminologique de textes déjà sur support électronique. Bref, le **dépouillement est effectué par un logiciel au lieu d'un humain** et permet une automatisation partielle de la tâche de dépouillement terminologique. Le logiciel extrait des syntagmes composés de deux, trois et même quatre mots ou plus qui apparaissent dans un texte qui lui a été soumis et les inscrit dans une liste ordonnée par fréquence ou ordre alphabétique.

Grâce au DAO, une première liste de mots est produite et le terminologue sera appelé à en faire l'épuration.

2.3.2.1 Types de dépouilleur

Je n'ai malheureusement pas le temps ici d'élaborer sur les types de dépouilleur mais sachons qu'ils sont regroupés en trois catégories (L'Homme, 1999 : 229) :

- 1) les dépouilleurs en mots simples, soit les *dépouilleurs lexicaux (DL)*;
- 2) les dépouilleurs en mots complexes (en unités terminologiques complexes), soit les *dépouilleurs terminologiques (DT)*;
- 3) les dépouilleurs en séquences de mots significatives : ces outils dépistent des suites de mots qui peuvent se révéler significatives dans le texte où elles sont utilisées : des cooccurrents (ex. *traiter des données*), des expressions figées ou fréquentes (ex. *société d'investissement à capital variable, titres comportant droit de vote*).

Je me pencherai maintenant sur les modes de fonctionnement du dépouilleur terminologique, outil de dépouillement relativement nouveau.

2.3.2.2 Modes de fonctionnement des dépouilleurs terminologiques

L'Homme (1999 : 244) explique deux techniques d'extraction utilisées pour l'identification de termes complexes : le **calcul des segments répétés** qui s'appuie sur une reconnaissance de caractères et tient compte de la fréquence des suites de mots, et l'**identification de patrons typiques** qui " recherche des unités terminologiques complexes en s'appuyant sur le fait que celles-ci se composent de suites de catégories grammaticales régulières " (1999 : 247).

2.3.2.3 Limites des dépouilleurs

Il faut tenir compte que les dépouilleurs modernes présentent par ailleurs des limites : les listes obtenues sont rarement parfaites.

Il n'en demeure pas moins que le dépouilleur fait une analyse d'un texte en quelques minutes (en fonction de la taille du texte) et en extrait une première liste de mots qui seront éventuellement épurés par le terminologue.

Il fallait donner à l'ordinateur, selon Perron (1989 : 6) une "compétence" linguistique. Il est nécessaire de préciser que le DAO visait avant tout à assister les terminologues qui ont par la suite à valider la pertinence des termes retenus. Le dépouilleur identifie automatiquement des termes contrairement au terminologue qui le faisait naguère manuellement.

Comme l'indique L'Homme (1999 : 252), deux critères très importants sont à retenir : la liste conserve parfois des suites indésirables, qui ne présentent aucun intérêt pour l'utilisateur. On les qualifie du générique *bruit*. Il se peut également que le dépouilleur ne conserve pas des mots qui seraient d'intérêt pour l'utilisateur. On qualifie cette omission du générique *silence*.

2.3.3 Établissement d'une première nomenclature

Le terminologue repère dans la liste produite par le dépouilleur terminologique (DT), les termes qui feront l'objet d'une recherche. En fait, il identifie les termes présentant entre eux des relations notionnelles et les indique sur la liste.

2.3.4 Découpage des termes et délimitation des contextes à l'aide d'un concordancier

Lors du découpage des termes, soit des unités terminologiques, il est très important de bien analyser la relation qui existe entre le déterminant et le déterminé; cette relation doit absolument être jugée essentielle pour que le terme soit retenu. Aussi faut-il rigoureusement retenir le contexte qui précise et cerne le plus possible la notion que couvre l'unité terminologique.

a) Rôle du concordancier

Le concordancier tient lieu de gestionnaire de contextes. Il localise des chaînes de caractères au moyen d'un affichage spécial (trait de surligneur, etc.) et l'utilisateur peut analyser les contextes et procéder au découpage et à leur délimitation.

Pour chacun des termes à relever, le concordancier permet de repérer le terme et son contexte. Il fouille une base de textes constituée d'un ensemble de fichiers généralement issus d'un logiciel de traitement de textes.

L'utilisateur prend connaissance des contextes et découpe, avec texte à l'écran, les parties à retenir. " L'opération du découpage des contextes vise essentiellement à appuyer chaque unité de la nomenclature par un contenu notionnel significatif " (Dubuc, 1992 : 52). Les contextes sélectionnés et délimités en fonction de leur teneur sémantique font alors l'objet d'une analyse et sont, par la suite, importés dans une base de données.

Les modes de fonctionnement diffèrent mais, en soi, un concordancier " établit une liste de *concordances*, c'est-à-dire une liste contenant la chaîne recherchée assortie de contextes " (L'Homme, 1999 : 177).

Passons maintenant à l'étape suivante laquelle consiste à traiter les contextes de façon automatisée de même qu'à en faire l'analyse sémantique.

2.3.5 Traitement automatisé des contextes

Une fois les contextes à l'écran, nous pouvons les structurer et utiliser les plus pertinents. Comme ils sont dans une base de données, il est relativement simple de les manipuler. Les contextes doivent donner une idée précise du rapport du terme à sa notion ou sinon l'expliquer.

2.3.5.1 Analyse sémantique et synthèse

Le passage des données terminologiques dans l'utilitaire permet d'en effectuer une analyse sérieuse et, s'il y a lieu, de compléter le dépouillement. " À ce stade, l'intervention du terminologue vise donc à assurer, pour chaque contexte significatif, la cohérence de lecture et la présence du maximum de traits sémantiques pertinents " (Dubuc, 1992 : 135). Il faut choisir ou rédiger des définitions. Il s'agit ici de faire des recoupements entre les termes extraits d'autres ouvrages de référence, le cas échéant, et d'ajouter des renseignements susceptibles d'appuyer la recherche.

2.3.6 Dernière analyse et consignation directe sur la fiche de travail à l'écran

Il existe des systèmes de gestion de base de données (SGBD) génériques, à savoir Access, FileMaker, dBASE, etc., de même que des SGBD spécialisés, tels que les logiciels de terminologie (*Multiterm*, *Termex*, etc.) qui, comme leur nom l'indique, sont " spécialement adaptés à la gestion des données de traduction " (L'HOMME, 1999 : 161).

Le moment est venu de rédiger les fiches sous forme définitive. Les termes et leurs contextes figurent sur les fiches, mais on y ajoute les autres renseignements que l'on avait jugé bon d'inscrire dans un protocole de rédaction de la fiche, soit date, auteur, information linguistique, etc.

2.4 Conclusion

Je dois insister sur le fait que malgré l'utilisation d'outils informatiques, les étapes de la démarche terminologique demeurent pratiquement inchangées. De plus, il est important de souligner la nécessité pour le terminologue de maîtriser le domaine dans lequel il travaille; sinon, son travail terminologique risque de manquer de rigueur et de précision. En dernière analyse, il faut se rendre à l'évidence que malgré tous les progrès techniques, seul le terminologue peut " juger de la pertinence des unités terminologiques, saisir les notions et assurer leur recoupement interlangue " (Dubuc, 1992 : 133).

3. RECHERCHE INFORMATISÉE DANS LE DOMAINE JURIDIQUE

3.1 Présentation des étapes

La recherche comprend cinq grandes étapes dont je donnerai un aperçu. Attardons-nous tout d'abord au premier élément de cette recherche informatisée.

3.2 Mise en forme du corpus

Il est impossible de travailler à partir de textes stockés sur des cédéroms sur le marché (par ex. code, lois et règlements) car ils sont protégés et encryptés; par conséquent, je ne peux en extraire de l'information ou encore effectuer des fonctions Copier et Coller (notamment pour l'importation dans Access), etc. J'ai jugé bon de m'adresser directement à la *Commission des valeurs mobilières du Québec* pour obtenir des textes de base sur support électronique et sous forme bilingue. Cette dernière m'a fourni *Loi et Règlement sur les valeurs mobilières du Québec*.

C'est donc ainsi que le corpus de base a été constitué. Il faut être particulièrement conscient du fait que le support électronique est nécessaire pour l'automatisation des tâches ultérieures.

Une fois stockés les textes dans l'ordinateur, il fallait les enregistrer en format de fichier *Texte seulement* car le produit alors obtenu était dépourvu de caractères gras, d'italiques et de ses attributs de présentation (tabulations, colonnes, etc.), ce qui permettait à Termplus une analyse de textes beaucoup plus performante.

3.3 Dépouillement

Le dépouillement a été réalisé au moyen du dépouilleur Termplus.

3.3.1 Historique

Termplus⁵ a été mis au point par une équipe de chercheurs de l'Université Laval. Il s'agit d'un logiciel visant la gestion informatisée de l'information textuelle et qui extrait les termes complexes (TC) **en ne tenant pas compte des termes simples**.

Ses concepteurs soutiennent également que Termplus est appelé à servir à des fins de lecture conceptuelle⁶, soit une nouvelle approche pour accéder au contenu d'un texte. Néanmoins, Termplus extrait d'ores et déjà les termes complexes contenus dans des textes de spécialité bilingues sur support électronique et les aligne.

3.3.2 Fonctionnement du dépouilleur

Ce logiciel fonctionne en version DOS, en version Windows 95, en version Windows 98 et ne possède pas de module de conversion de données. Il dépouille notamment des textes en anglais et en français. Termplus constitue une liste des termes les plus représentatifs du contenu d'un texte de spécialité et permet, entre autres, de les retrouver en tête de liste par ordre inverse de fréquences. Le travail du terminologue consiste à épurer ensuite cette liste.

L'approche du logiciel Termplus repose sur l'extraction et la manipulation des termes complexes dans un texte de spécialité. D'après les chercheurs de l'Université Laval, Termplus se limite au dépouillement de termes complexes (environ 80 % des termes qu'il faut extraire

dans un corpus) et ne retient aucun terme simple (approximativement 20 % des termes dans un document). Ce constat revêt une très grande importance à ce stade-ci car le dépouillement de textes, en dépit de l'automatisation, demeure pour ainsi dire inachevé.

Terminus met en application le calcul des segments répétés, technique d'extraction de termes complexes dont j'ai déjà parlé. Ce genre d'extraction tient compte des chaînes de caractères paraissant plus d'une fois dans les textes analysés.

Diverses formes de filtrage peuvent être complémentaires à la technique d'extraction de termes complexes que Terminus met en application.

3.3.3 Terminus et le corpus de textes juridiques

Je vous entretiendrai donc du traitement effectué par Terminus, des résultats du dépouillement, de l'épuration des listes obtenues et de la sélection des termes par le terminologue. Je soulèverai, en dernier ressort, certains problèmes constatés à la suite de l'utilisation de ce logiciel.

a) Traitement effectué par Terminus

Le dépouillement des termes retenus par le logiciel apparaît sous la colonne de droite *Index conceptuel*.

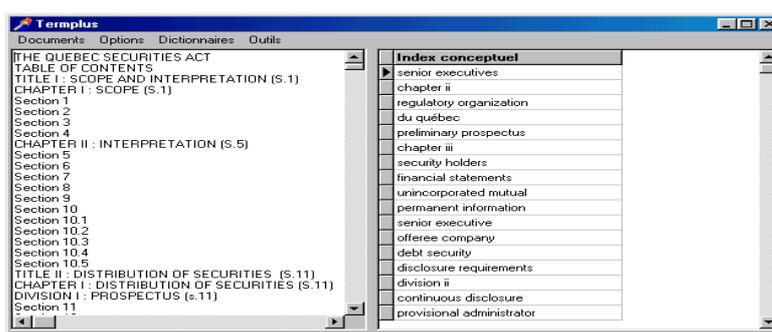


Figure 3.1 Écran Terminus – Index conceptuel

À partir de l'Index conceptuel obtenu, on obtient un Résumé comportant l'entrée à gauche et l'entrée surlignée en bleu dans le contexte apparaît dans la partie de droite.

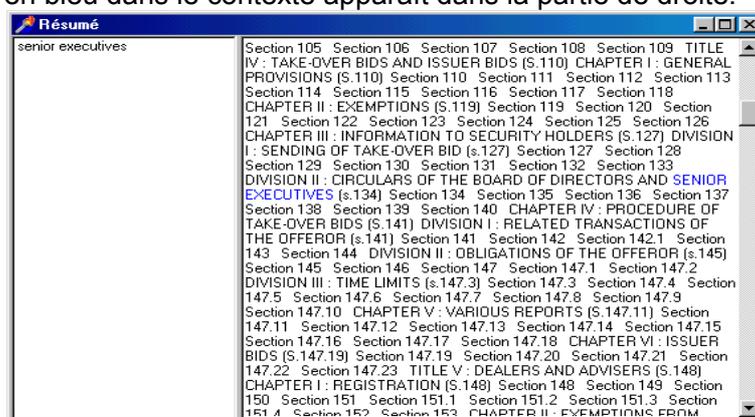


Figure 3.2 – Terminus – Génération de résumé.

Tableau I

Nombre de termes complexes retenus par Termplus					
CORPUS	Loi anglaise	Loi française	Règlement anglais	Règlement français	TOTAL
Nombre de mots	32 578	38 855	32 504	34 963	138 900
Termes complexes retenus	432	1967	390	1543	4332

Le **Tableau I** donne une idée globale des résultats obtenus au moyen de l'analyse faite par Termplus. Le corpus analysé était donc constitué de quatre textes (138 900 mots) et Termplus en a extrait 4332 mots complexes distribués comme il suit : 432 dans la loi anglaise, 1967 dans la loi française, 390 mots complexes dans le règlement anglais et enfin 1543 dans le règlement français. Les listes produites étaient donc assez volumineuses.

Le temps de traitement automatique prend quelques minutes seulement et s'accroît en fonction de la taille du texte.

b) Résultats du dépouillement

Il est très important de noter que la longueur maximale des segments est variable et paramétrable (2-6 éléments) et, rappelons-le, que la liste ne comporte aucun terme simple.

Je me retrouve toutefois en présence de bruits et de silences. Voici quelques exemples de suites indésirables tirées de mes listes françaises :

fin du dernier exercice,
présente loi, etc., et

de mes listes anglaises :

business unless,
client solely, etc.

Il en est de même pour les silences relevés. Je ne retrouve pas, par exemple, des termes comme

closed company ou
franchise agreement en anglais ou

émetteur associé ou
soucription à forfait en français,

termes qu'un terminologue aurait sans doute retenus.

À noter que la liste obtenue n'a pas fait l'objet de filtrage " post-dépouillement ", c.-à-d. que les résultats obtenus auraient pu être améliorés si l'on avait utilisé un filtre, par. ex. pour éliminer

tous les segments contenant des chiffres, ou si l'on avait ajusté le module linguistique du logiciel, par ex. pour ne garder que le singulier d'un terme repéré. Toutefois, certains termes auraient été à tort ignorés dans ce dernier cas.

Il ressort que le logiciel peut dépouiller un texte en produisant une liste partiellement exacte bien qu'incomplète.

c) Épuration des listes et sélection des termes

L'humain doit épurer les listes de termes complexes produites par le logiciel, ce qui est moins long que de parcourir tout un texte. Vu l'option *Ordre inverse des fréquences* retenue (en fonction de leur nombre d'occurrences), les termes les plus fréquents figurent au début du texte. Parcourant les listes et surlignant dans un premier temps les termes intéressants, je les épure et y indique clairement les termes à retenir à des fins de consignation.

La **pertinence** guide le choix des termes. En effet, chaque terme retenu pour ma base de données doit être dans la mesure du possible rattaché à l'"arbre de domaine", soit celui des valeurs mobilières. S'y greffent nécessairement certains termes reliés aux domaines de la comptabilité et de la gestion financière. Mon expérience et la documentation obtenue (lexiques fournis par des praticiens du domaine) ont aussi déterminé la "pré-validation" ou encore l'épuration des termes obtenus après l'extraction terminologique.

J'établis ensuite la nomenclature de l'anglais pour laquelle il me faudra trouver des équivalences françaises.

3.3.4 Problèmes relevés

Il serait opportun de souligner certains problèmes à l'origine de bruits et de silences dont

- l'absence de distinction terme/non-terme

Le dépouilleur extrait souvent beaucoup de termes qui ne sont pas des unités terminologiques. Des expressions telles que *option previously* et *dispense d'inscription à titre* ne sont d'aucune utilité.

- un mauvais découpage

Ainsi, Termplus retient des éléments tels que *titres d'emprunt émis* alors qu'il a déjà *titres d'emprunt* dans la liste extraite du règlement et des termes comme *permanent information* dans la loi anglaise alors que ce terme complexe devrait être le suivant : *permanent information record*.

- la présence de chaînes numériques ou alphanumériques et de mots vides

Des chaînes de caractères telles que des chaînes numériques (\$1 000, january 1983) ou alphanumériques (c-4,v-1,c-37, etc.) sont retenues⁷. Nous constatons également la présence de mots vides comme le, de, etc., dans certains termes complexes retenus (par ex. la confédération des syndicats nationaux).

- la discrimination majuscules/minuscules

Termpus exerce une discrimination vis-à-vis certains caractères et, par conséquent, ne prend pas en considération les majuscules des termes qu'il retient. À titre d'exemple, il retient en minuscules *united states*.

- la non-équivalence entre les entrées anglaises et françaises

Certains termes, en fait des équivalents de termes anglais qui ont été repérés, ne sont pas forcément présents dans la liste française. Il nous faut donc les chercher.

3.4 Collecte des contextes

3.4.1 Rôle de concordancier exercé par Termpus et copie dans Access

Termpus joue également le rôle de concordancier. Il surligne les endroits où figurent les termes. En effet, il était possible d'extraire directement les contextes dans les résumés et de les copier sur des fiches dans la base de données Access.

3.5 Analyse des contextes

Les contextes sont affichés dans une fenêtre. La Figure 3.3 illustre le cas d'*organisme d'autoréglementation*. Termpus regroupe les contextes et surligne les termes complexes retenus lors de l'analyse.

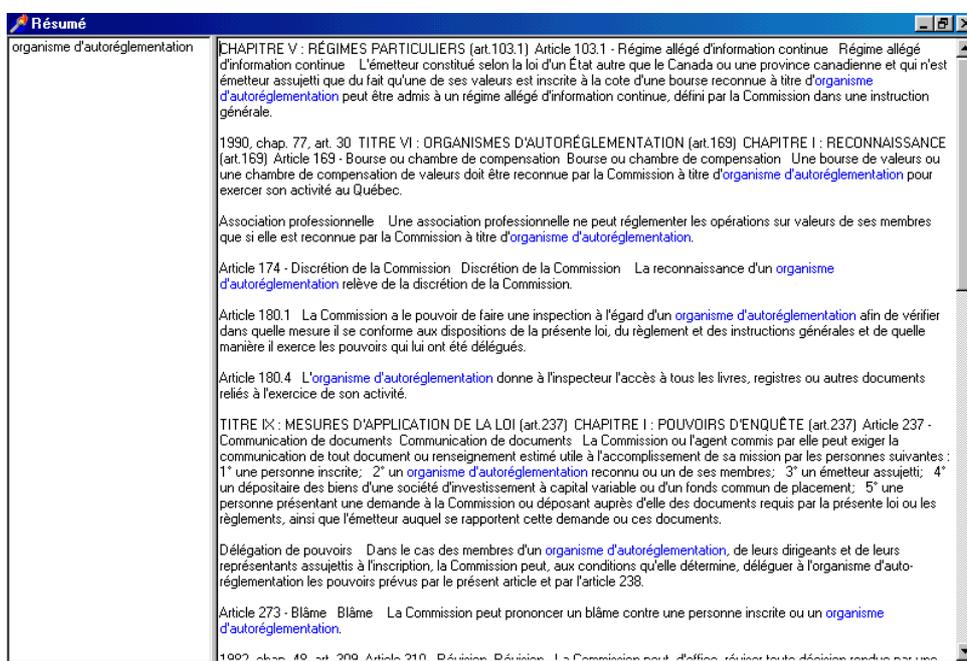


Figure 3.3 – Analyse des contextes

J'analyse donc les contextes affichés à l'écran et j'en dégage les contenus notionnels qui identifient le mieux le terme. Les contextes sont alors conservés dans la mesure où ils décrivent bien la notion.

Le regroupement des contextes effectué par Termplus accélère l'analyse terminologique. Il ne faut cependant pas en déduire qu'il n'y a pas de problèmes pour autant, soit dit en passant mineurs par rapport aux avantages.

Je procède alors au découpage et à la délimitation des contextes. J'ai habituellement retenu des contextes en LA équivalents à ceux de la LD. Il serait bon de souligner que Termplus est un concordancier "monolingue", c.-à-d. que j'ai à l'écran tantôt les contextes anglais, tantôt les contextes français.

3.5.1 Remarques sur Access

Les contextes retenus sont par la suite versés à l'aide des fonctions *Copier – Coller* dans un SGBD. J'ai utilisé Access. C'est un progiciel très convivial mis au point par Microsoft. Sa particularité réside dans le fait que la base est relationnelle, c.-à-d. que le champ d'une table peut être repris dans une autre table et que les tables peuvent se partager des informations.

J'ai jugé bon d'établir trois tables dans Access :

- Valeurs (table française)
- Valeurs1 (table anglaise)
- CVMQ (Contextes français, contextes anglais, fiches bilingues).

La base Valeurs est la base française et contient les " fiches " où figurent, outre l'entrée française, les rubriques suivantes : la numérotation automatique, le Contexte 1 sous forme de contexte ou de définition, et la Source contexte 1 où figure le code de la source. La fiche n° 1 donne ceci : l'entrée **notice d'offre** accompagnée du contexte :

*C. L'information financière prospective figurant dans un prospectus, une **notice d'offre** prévue par la Loi ou le règlement, ou dans un document dont la Commission autorise l'utilisation au lieu d'un prospectus est établie selon les instructions générales de la Commission et accompagnée du rapport du vérificateur.*

et de la source COREGL 50 dont je vous donnerai l'explication plus loin.

La base Valeurs 1 quant à elle, est la base anglaise et contient les mêmes rubriques.

3.6 Constitution de la fiche terminologique dans Access

À partir des phases embryonnaires d'établissement de fiches, je constitue maintenant une base où les champs sont repris dans leur intégralité et j'établis la fiche où peuvent figurer des champs complémentaires qui ont fait l'objet d'un protocole.

3.6.1 Rédaction des fiches Access

Une fiche complète est mise au point. Je pourrai ensuite achever mes recherches en ayant des champs bien délimités.

Dans le but d'alléger la fiche, j'ai établi des codes de source succincts pour les contextes anglais et les contextes français, par ex.

- COLOI pour signifier la Commission des valeurs mobilières du Québec. Loi sur les valeurs mobilières du Québec. Montréal, 12 juin 1997.

D'autres codes ont été utilisés dans la rédaction de mes fiches, comme par ex.

CCQ : Code civil du Québec.

La fiche ainsi faite est très claire et facile à consulter. Contrairement à la base de données sur papier, le contenu de la base de données informatisées peut être affiché dans tous les ordres possibles et imprimé à tout moment.

3.7 Conclusion

Je peux remanier automatiquement tout le contenu de la base de données Access et classer les informations dans un ordre préétabli, tout en sachant que je pourrai en ajouter ou éventuellement en retirer en un tour de main. Il s'agit d'un progrès immense.

4. CONCLUSION

Je me suis attachée, dans le présent exposé, à reconstituer la " terminogenèse " de la pratique terminographique. La méthodologie traditionnelle a cédé le pas à la méthodologie informatisée. En effet, plusieurs maillons de la chaîne terminographique ont été automatisés au cours des dernières années et une " terminotechnosynergie " s'est créée entre les étapes automatisées. Par ailleurs, la mise au point d'une base de données m'a permis de constater que l'ordinateur affiche un taux certain de fiabilité sur le plan du traitement. Le but ultime d'un travail de dépouillement terminologique n'en demeure pas moins inchangé : faire correspondre les terminologies anglaise et française et les valider.

Par la suite, j'ai fait ressortir l'intérêt général que représente le dépouillement assisté par ordinateur (DAO) dans le domaine de la traduction de textes de spécialité, particulièrement de textes volumineux, en privilégiant l'option d'ordre inverse des fréquences. Le dépouilleur terminologique présente également des avantages pour d'autres applications linguistiques, notamment pour la lecture contextuelle, une nouvelle approche pour accéder au contenu d'un texte.

Des améliorations sont encore à apporter à Termplus car l'épuration des résultats requiert souvent beaucoup de temps et il faut parfois se reporter au texte de départ pour trancher. Malgré les bruits observés dans les listes et les limites du logiciel, Termplus m'a grandement étonnée. Il permet avant tout d'avoir une idée plus précise du contenu textuel et terminologique.

L'exposé de ma méthode permet de faire ressortir le bien-fondé de l'intervention du terminologue dans le processus de validation des données et dans l'enchaînement méthodologique. Compte tenu de ces considérations, je pense que le travail de dépouillement sera de plus en plus effectué par des dépouilleurs terminologiques " supervisés " par des terminologues spécialisés, en raison des facteurs de productivité et de rentabilité sans cesse

accrus de même que de l'évolution des mentalités face à l'utilisation de la machine. J'estime plutôt que le travail terminographique sera de plus en plus automatisé, mais qu'on ne pourra jamais remplacer totalement l'humain malgré la réingénierie du travail. L'informatique permet d'être plus efficace indépendamment de la méthode utilisée, la qualité des recherches et du travail reposant principalement sur une **méthodologie rigoureuse**.

Par ce travail, j'espère ouvrir la voie à d'autres projets de jumelage en matière de dépouillement assisté par ordinateur (DAO) et de terminologue spécialiste. La terminotique est omniprésente. D'après moi, le jour n'est pas si lointain où le commun des mortels pourra accéder à une fonction " Dépouiller " intégrée dans Word.

¹ " Consignation, traitement et présentation des données qui résultent d'une recherche terminologique " (Boutin-Quesnel *et al.*, 1985 : 17) .

² Ensemble des situations et activités dans lesquelles l'informatique intervient pour constituer, organiser, gérer ou exploiter des données (Gouadec (1993) cité dans L'Homme, 1996 : 45).

³ Selon Boutin-Quesnel *et al.* (1985 : 20), il s'agit d'une " unité signifiante constituée d'un mot (**terme simple**) ou de plusieurs mots (**terme complexe**) et qui désigne une notion de façon univoque à l'intérieur d'un domaine. Note. - Aussi appelé **unité terminologique** " . Le terme diffère d'un mot en ce sens qu'il peut être composé d'un seul mot ou de plusieurs mots, par ex. *fonds commun de placement* .

⁴ Note à Danielle Verrier, le 20 janvier 2000.

⁵ Notions est son ancêtre.

⁶ " Une approche basée sur l'extraction et la manipulation des termes dans un texte de spécialité permettant une lecture non linéaire. " (Exposé donné par Arman Tajarobi à l'Université de Montréal le 20 octobre 1999.)

⁷ Toutefois, ceci pourrait être contrôlé à l'aide d'un filtre.

5. BIBLIOGRAPHIE

AUGER, P., B. de Bessé, B. Salvail *et al.* (1973) : *Guide de travail en terminologie*, Cahiers de l'Office de la langue française n° 20, Québec : Ministère de l'Éducation.

AUGER, P., P. Drouin et M.-C. L'Homme (1991) : " Automatisation des procédures de travail en terminographie ", *Meta* 36 (1), pp. 121-127.

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

BOUTIN-QUESNEL, Rachel, N. Bélanger *et al.* (1985) : *Vocabulaire systématique de la terminologie*, Cahiers de l'Office de la langue française, Québec : Gouvernement du Québec.

Commission des valeurs mobilières du Québec (1998) : *Loi et Règlement sur les valeurs mobilières du Québec*, décembre 1998.

DUBUC, R. (1992) : *Manuel pratique de terminologie*, 3^e édition, Brossard (Québec) : Linguatex.

-
- GOUADEC, D. (éd.) (1993) : *Terminologie et terminotique. Outils, modèles et méthodes. Actes de la première Université d'automne en terminologie.* Rennes 2. 21-26 sept. 1992, Paris : La Maison du Dictionnaire.
- LADOUCEUR, J. et A. Tajarobi (1999) : Lecture conceptuelle – Une nouvelle approche pour accéder au contenu d'un texte. Exposé donné à l'Université de Montréal le 20 octobre 1999.
- L'HOMME, M.-C. (1999) : *Initiation à la traductique (Édition provisoire).* Brossard (Québec) : Liguattech.
- L'HOMME, M.-C. (1996) : Traduction et informatique. **TRA 2000 : Outils informatiques de langagiers à la traduction.** Recueil de notes, Montréal : Université de Montréal, pp.36-52.
- PERRON, J. (1989) : “ Termino : un système de dépouillement terminologique ”. *Terminogramme* 54, pp.3-9.
- RONDEAU, G. (1984) : *Introduction à la terminologie*, 2^e éd., Chicoutimi : Gaëtan Morin.