

# **Nouvelles technologies et recherche terminologique**

## **Techniques d'extraction des données terminologiques et leur impact sur le travail du terminographe**

**Marie-Claude L'Homme**

**Université de Montréal**

### **Résumé**

*L'apparition de divers outils de traitement des données lexicales a modifié sensiblement, sans les bouleverser totalement, diverses tâches que le terminologue-terminographe devait réaliser différemment. Nous décrirons certaines applications – dont certaines sont conçues expressément pour la terminographie et d'autres peuvent être adaptées à ses besoins – et nous montrerons de quelle manière elles modifient, allègent ou enrichissent une terminographie plus classique. Les outils qui retiendront notre attention sont ceux qui extraient de textes qu'on leur soumet divers types de données terminologiques. Nous montrerons ce que permettent certaines applications commerciales, mais nous ferons également état des résultats de recherches universitaires dans le domaine. Les exemples illustreront des données anglaises et françaises, mais nous mettrons l'accent sur la recherche unilingue.*

### **Abstract**

*Computerized tools designed to process lexical unit have changed considerably – although not drastically – several tasks terminologists-terminographers were used to perform in a different manner. Some applications designed to extract various types of terminological data will be described. Some of these have been designed for terminology purposes, while others can be adapted to terminology research requirements. We will also show how these tools change, facilitate, or enrich conventional terminology tasks. We will show what can be done with commercial applications, but we will also describe results delivered by academic research in this area. Examples will show French and English data. However, our focus will be on monolingual terminographic research.*

**Mots clés :** terminologie, terminographie, corpus, extraction de termes, termes complexes, liens conceptuels, cooccurrents

**Keywords :** terminology, terminography, corpora, term extraction, complex terms, conceptual relationships, co-occurents

## 1. Introduction

La terminographie, à l'instar d'autres domaines langagiers s'appuyant fortement sur les textes comme la lexicographie ou la traduction, subit des transformations en raison de l'introduction de nouvelles technologies informatiques. Certaines d'entre elles rendent des services inestimables alors qu'on s'interroge encore sur l'opportunité des autres.

Nous examinons, dans les pages qui suivent, un groupe d'applications informatiques dont l'objet est d'extraire, de textes spécialisés qu'on leur soumet, des données terminologiques de natures diverses. Il s'agit de tâches particulièrement difficiles à automatiser et il est permis de s'interroger sur leur apport dans la chaîne de travail terminographique.

La section 2 rappelle quelques précisions sur la terminographie classique et « informatisée » telle que nous l'envisageons. La section 3 aborde différentes techniques mises au point ces dernières années pour extraire des données terminologiques de textes en format électronique. La première application qui retiendra l'attention (3.2) est l'extraction de termes complexes puisqu'elle est implantée dans un nombre croissant de logiciels. Nous nous pencherons également (3.3) sur d'autres applications moins répandues comme l'extraction de contextes ciblés et l'appariement de termes reliés sur le plan paradigmatique (liens conceptuels, par exemple) ou syntagmatique (groupes formés de termes et de leurs cooccurrents, par exemple). La section 4 étudie l'impact qu'ont ces techniques sur le travail du terminographe.

## 2. La recherche terminographique classique et réalisée à l'aide d'outils informatiques

Il existe diverses manières d'envisager les tâches terminographiques en fonction de l'environnement dans lequel on les accomplit. On peut les mettre en oeuvre pour décrire

des unités linguistiques dans des dictionnaires, pour représenter des connaissances, pour répondre à des besoins de traduction ou pour normaliser les usages au sein d'une entreprise ou d'une communauté linguistique. Les outils informatiques retenus pour réaliser l'une ou l'autre tâche varieront forcément.

## 2.1 Point de vue adopté

Les tâches qui retiendront l'attention dans ce qui suit sont celles déployées dans le cadre d'une recherche terminographique dont l'objet est la confection de dictionnaires spécialisés ou l'enrichissement de banques de terminologie à partir de corpus. Les techniques informatiques examinées sont celles qui fouillent un texte ou un groupe de textes spécialisés dans le but d'en extraire des informations. Ces dernières sont par la suite interprétées, analysées et synthétisées par le terminographe en vue de décrire des unités terminologiques. La figure 1 illustre le point de vue adopté.

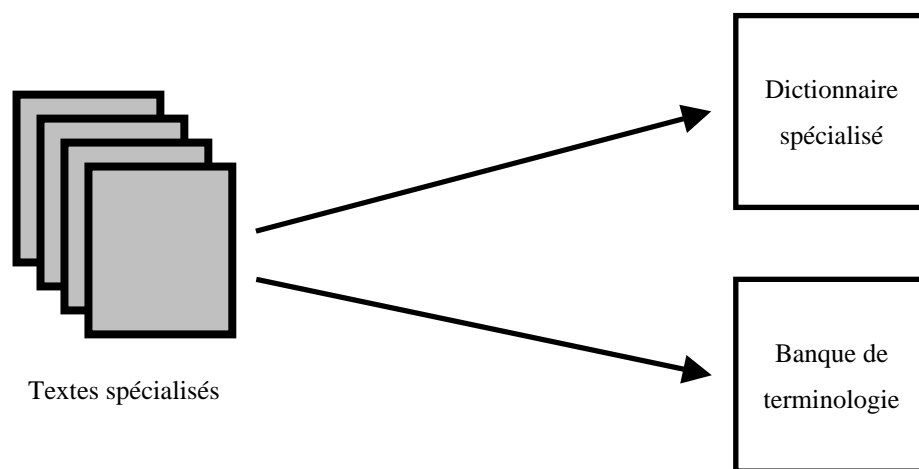


Figure 1 : Préparation de descriptions terminographiques à partir de textes spécialisés

## 2.2 Nouvelles générations de dictionnaires spécialisés

Nous nous intéresserons à des dictionnaires spécialisés et à des banques de terminologie qui appartiennent à une nouvelle génération. Les dictionnaires classiques décrivent des termes de nature nominale dans des fichiers ou des articles monoconceptuels et retiennent les rubriques suivantes : des vedettes dans une ou plusieurs langues, une définition, des contextes et des marques d'usage.

Depuis quelques années, les supports de données terminologiques tentent d'intégrer des éléments descriptifs différents et plus riches que ce que l'on retrouve dans les dictionnaires terminologiques élaborés selon les règles de l'art. Ils tiennent compte de la polysémie, d'unités appartenant à des catégories différentes de celle du nom et peuvent retenir des rubriques comme les cooccurrents ou les termes apparentés morphologiquement à la vedette. Si les descriptions sont faites sur support électronique, le terminographe peut relier les entrées en fonction d'affinités sémantiques ou syntagmatiques. Les trois figures qui suivent contiennent des illustrations de ces répertoires nouveaux.

La figure 2 renferme la reproduction d'une entrée apparaissant dans le Meynard (2000 : 22-23). Il s'agit d'un dictionnaire bilingue consacré au vocabulaire d'Internet dans lequel la terminographe dresse des listes de « combinaisons lexicales spécialisées », à savoir des groupes de mots contenant une unité terminologique et un autre mot se combinant avec elle.

### **CLIENT**

#### **Définition**

Un ordinateur ayant accès aux ressources partagées du réseau fournies par un autre ordinateur appelé serveur.

**Terme noyau anglais** : client

### **Nom cooccurrent + terme noyau**

· Exécution d'un client	Run of a client
· Fenêtre d'un client	Window of a client
· Infrastructure d'un client	Infrastructure of a client
· Installation d'un client	Set up of a client
· Requête d'un client	Request of a client
· Session d'un client	Session of a client

### **Verbe cooccurrent + terme noyau**

· Exécuter un client	To run a client
· Installer un client	To set up a client
· Lancer un client	To launch a client
· Quitter un client	To exit a client

### **Terme noyau + Verbe cooccurrent**

· Client affiche	Client displays
· Client charge	Client loads
· Client dialogue avec	Client communicates with
· Client envoie	Client sends
· Client ferme	Client closes
· Client fonctionne	Client runs
· Client interagit avec	Client interacts with
· Client interprète	Client interprets
· Client récupère	Client retrieves
· Client traite	Client processes
· Client transfère	Client transfers

### **Terme noyau + Adjectif cooccurrent**

· Client distant	Remote client
· Client local	Local client

Figure 2 : Entrée *Client* tirée du Meynard (2000 : 22-23)

La figure 3 montre un fragment de l'entrée *Économie* tirée du Binon *et al.* (2000 : 210-214) intitulé *Dictionnaire d'apprentissage du français des affaires* (DAFA). Ce dictionnaire spécialisé est celui qui comporte les entrées les plus riches sur le plan linguistique. Il présente les mots morphologiquement apparentés au mot vedette, effectue des renvois à des entrées décrivant des concepts connexes, décrit les sens multiples que

peut revêtir le mot vedette dans les corpus dépouillés, présente les expressions et les collocations dans lesquelles on retrouve la vedette, ces dernières étant classées en fonction de la catégorie grammaticale du cooccurrent et de son sens. Finalement, des contextes servent à illustrer l'emploi du mot vedette ou d'autres mots qui lui sont associés..

## ÉCONOMIE

? épargne

1 l'économie 4 les déséconomies 4 la micro-économie 4 la macro-économie 4 l'économique 4 l'économisme 4 l'économétrie 4 un économat 4 un économiseur	4 un économiste, une économiste 4 un économe, une économe	2 (anti)économique 4 macro-économique 4 micro-économique 4 économe 4 économétrique 4 économique- 4 socio-économique 4 économiquement	3 économiser
--	--	---	--------------

### 1 l'ÉCONOMIE – (n.f.)

1.1 Ensemble des activités des personnes ...  
*La fragilité structurelle de notre économie ...*

1.2 (emploi au singulier) Science qui étudie l'économie (sens 1.1).  
*L'économie n'est pas une science exacte ...*

1.3 etc.

#### expressions

(sens 1.1)

**(Quand) économie rime avec écologie.** *Économie rime avec écologie : le procédé Orvert ...*

(sens 1.3)

• (Une personne) **faire l'économie de qqch.** : éviter d'utiliser qqch ...

#### + adjectif

TYPE D'ÉCONOMIE (sens 1.1 et 1.2)

**L'économie classique ; marxiste ; keynésienne ; néo-classique** : grande écoles ...

#### CARACTÉRISATION DE L'ÉCONOMIE (sens 1.1)

**Une économie ouverte** : économie d'un pays qui ...

#### + nom

(sens 1.1)

• **La (bonne x mauvaise) santé de l'économie...**

(sens 1.2)

**Un diplôme en économie.** (Syn. : **un diplôme en sciences économiques**, (fam.) **un diplôme en sciences éco...**)

+ verbe : qui fait quoi ? (sens 1.1)		
une mesure, le gouvernement ? l'~	? <b>stimuler</b> l'~  <b>se développer</b>	la stimulation de l'~  le développement de l'~
l'~	?? <b>tourner à plein régime à plein rendement</b> ➤ <b>se porte bien tourne bien, rond</b>	- -
etc.		

Figure 3 : Extrait de l'entrée *Économie* tirée du DAFA (2000 : 210-214)

Enfin, la figure 4 illustre une partie de la modélisation des liens conceptuels entre *règle* et d'autres concepts appartenant au domaine de l'intelligence artificielle (Otman 1996 : 119). Les modélisations de ce genre sont faites dans ce qu'il est convenu d'appeler

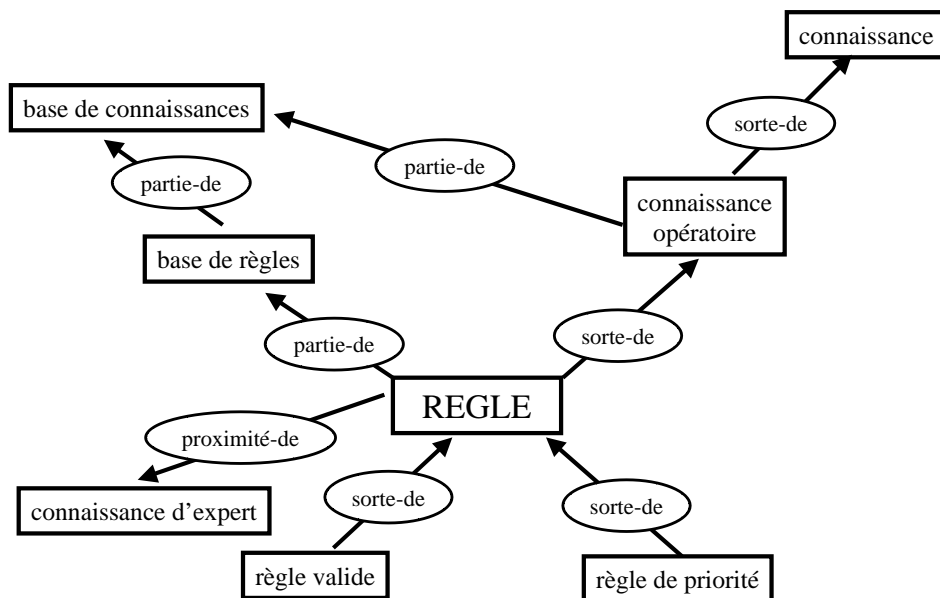


Figure 4 : Partie du réseau sémantique de *règle* (Otman 1996 : 119)

*base de connaissances terminologiques* (Meyer et al. 1992). Le formalisme utilisé par Otman est celui des *réseaux sémantiques* (Quillian 1968).

## 2.3 Les étapes de la recherche terminographique

La recherche terminographique peut être découpée en un certain nombre de tâches qui apparaissent dans la partie gauche de la figure 5. Ce découpage, essentiellement didactique, se retrouve dans cette forme ou dans une autre dans la plupart des manuels de terminologie (Dubuc 1986, Rondeau 1984, pour n'en nommer que quelques-uns).

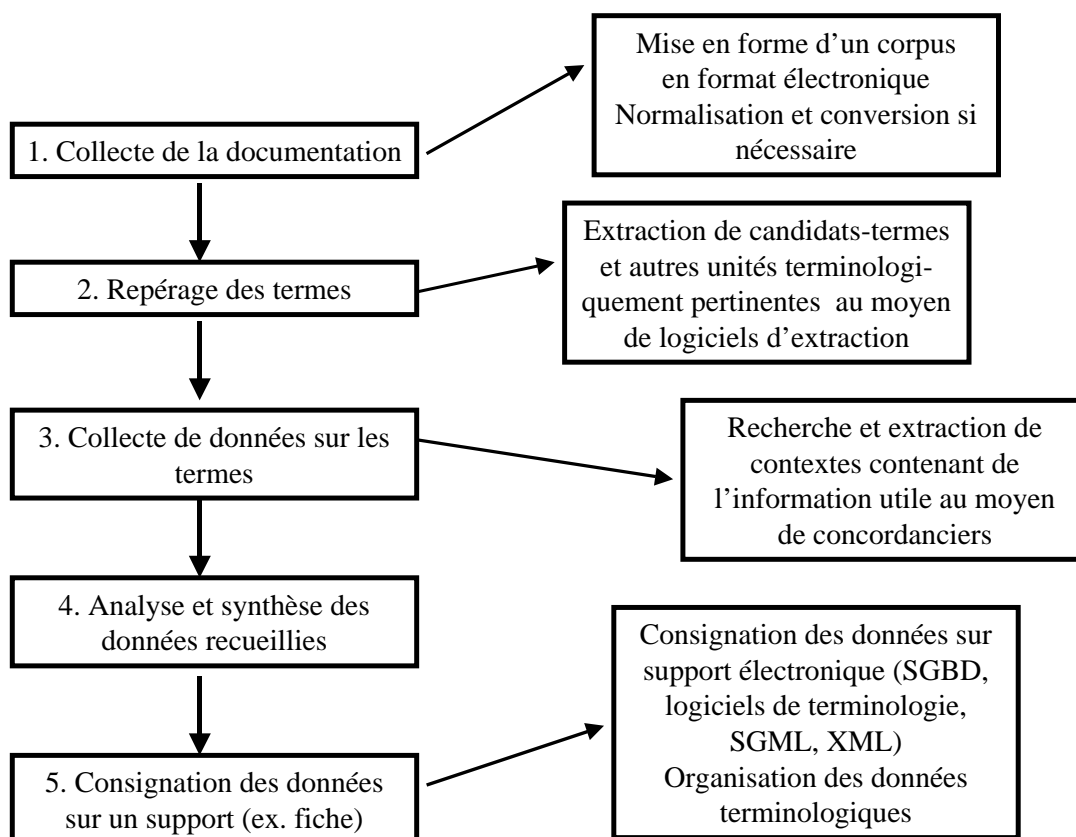


Figure 5 : Séquence de tâches en recherche terminographique classique et automatisée

La première tâche consiste à assembler une documentation, c'est-à-dire réunir des textes représentatifs du domaine dont on souhaite décrire le vocabulaire (médecine, droit, sécurité industrielle, informatique, etc.). Le terminographe procède ensuite à un repérage des unités qui feront éventuellement partie du dictionnaire spécialisé qu'il prépare ou la banque de terminologie qu'il enrichit. Ce premier repérage sera constamment raffiné au



cours des étapes suivantes. Ensuite, il convient de recueillir des données sur les termes repérés. Une grande partie de ces données proviennent des textes formant la documentation. D'autres informations sont recueillies auprès de spécialistes du domaine. Le terminographe procède ensuite à l'analyse et à la synthèse des données pour ensuite inscrire les informations définitives sur un support communément appelé la *fiche terminographique*.

Le terminographe peut automatiser une partie de ces tâches en y intégrant différents types de logiciels. Les possibilités sont illustrées dans la partie droite de la figure 5 (d'après Auger *et al.* 1991).

La recherche documentaire consiste à réunir des textes en format électronique, dans ce qui est convenu d'appeler un *corpus*. Cette tâche peut se résumer au stockage de textes déjà accessibles dans cette forme ou nécessiter des opérations de conversion laborieuses (si les textes sont en format papier ou s'ils sont dans un format qui n'est pas compatible avec les logiciels utilisés par la suite).

Le repérage des termes est automatisé en partie à l'aide de logiciels appelés *extracteurs de termes* qui se concentrent le plus souvent sur le dépistage de termes complexes. Toutefois, de nouveaux outils recherchent d'autres unités « terminologiquement pertinentes » (nous y reviendrons plus loin).

La collecte de données sur les termes est grandement facilitée par les concordanciers qui extraient rapidement des contextes associés à un terme donné à la demande d'un utilisateur. Ici, encore, de nouveaux outils permettent d'isoler des contextes spécifiques ou d'exploiter l'information qui se trouve de part et d'autre des termes préalablement recherchés.

Enfin, la consignation des données se fait depuis fort longtemps sur support électronique, c'est-à-dire dans les banques de terminologie classiques dont les premières ont été mises sur pied dans les années 1960 et 1970. Toutefois, aujourd'hui, les outils

existants autorisent des modes de consignation nettement plus polyvalents et flexibles que ce qui était permis auparavant. Un champ est associé à plusieurs autres dans une même base ou entre bases de données différentes (le mode relationnel); les champs d'une base de données ou les données elles-mêmes peuvent être hiérarchisés (le mode hiérarchique).

### 3. Extraction de données terminologiques

Nous allons maintenant décrire certaines techniques d'extraction de données terminologiques et réfléchir à leur impact sur le travail du terminographe. La plupart des outils examinés n'ont pas été mis au point d'abord pour la recherche terminographique (par exemple, de nombreux travaux portant sur l'extraction de termes sont faits dans le contexte de la recherche d'information), mais les descriptions se limitent aux aspects pertinents pour la terminographie.

Précisons également que les commentaires qui suivent s'appuient sur des expériences d'utilisation et d'évaluation d'outils d'extraction de termes menées à l'Université de Montréal (L'Homme *et al.* 1996; Love 2000, Verrier 2000) ainsi que sur une abondante littérature sur le sujet (notamment, Cabré *et al.* 2000, Habert et Jacquemin 1993, Jacquemin 2001, Lauriston 1994 et Otman 1991; voir également les travaux portant sur des systèmes particuliers et cités à l'Annexe A).

Enfin, nous mettons de côté, faute d'espace, les outils et travaux qui portent sur l'alignement de textes ainsi que l'exploitation des textes une fois appariés. Ces travaux présentent un intérêt indéniable pour la terminographie bilingue et multilingue et feraient l'objet, à eux-seuls, d'un autre article.

### 3.1 Ce qu'on peut tirer automatiquement ou semi-automatiquement des textes en format électronique

Nous regrouperons les renseignements qu'on peut tirer des textes en format électronique en deux catégories. La première que nous appelons *version de base* renferme l'extraction de concordances et l'extraction de termes (surtout de termes complexes). Ces deux premières applications sont illustrées aux figures 6 et 7.

```
ravels to and from the CPU and the other parts of the computer. Comp  
e microprocessor; this CPU is also used in the PC/XT, the Portable  
second, a measure of CPU speed). The 8086 Microprocessor The 808 I  
ng is the ability of a CPU to perform several tasks at a time -- s  
hold." But because the CPU switches among tasks so quickly, it appea  
sor used i 80486. This CPU offers faster clock speeds than does the
```

Figure 6 : Concordances pour *CPU* extraites d'un corpus d'informatique

La figure 6 montre une série de contextes (sous forme d'index KWIC) extraits d'un corpus de textes d'informatique au moyen d'un concordancier.

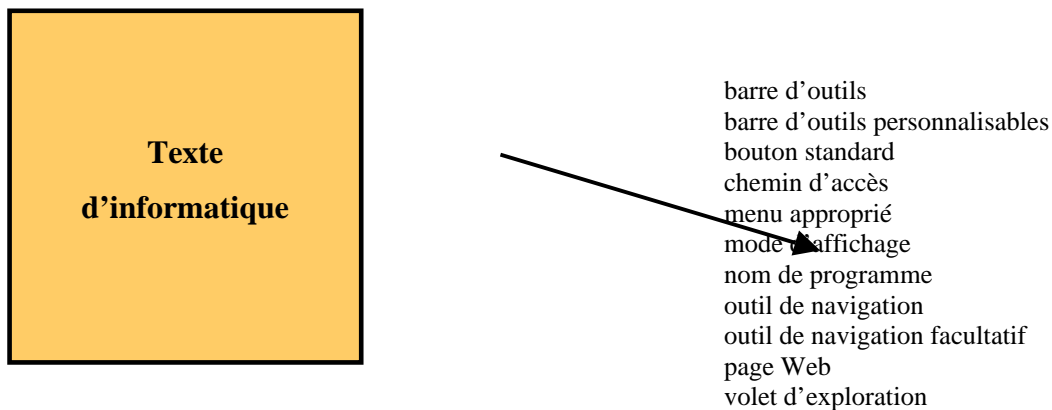


Figure 7 : Termes complexes extraits de d'un corpus de textes d'informatique

La figure 7 montre une liste produite par un extracteur de termes fictif à la suite du traitement d'un texte d'informatique. En réalité, ces logiciels recherchent surtout des termes complexes et proposent à l'utilisateur une liste de candidats-termes, c'est-à-dire de groupes de mots susceptibles d'être révélateurs des connaissances véhiculées dans le texte traité.

Nous ne reviendrons pas sur la première application que nous supposons largement connue. La seconde, en revanche, sera décrite à la section 3.2.

D'autres renseignements peuvent être extraits des textes en format électronique. Ces nouvelles applications seront considérées dans un deuxième groupe que nous appellerons *version enrichie* et sont illustrées à la figure 8.

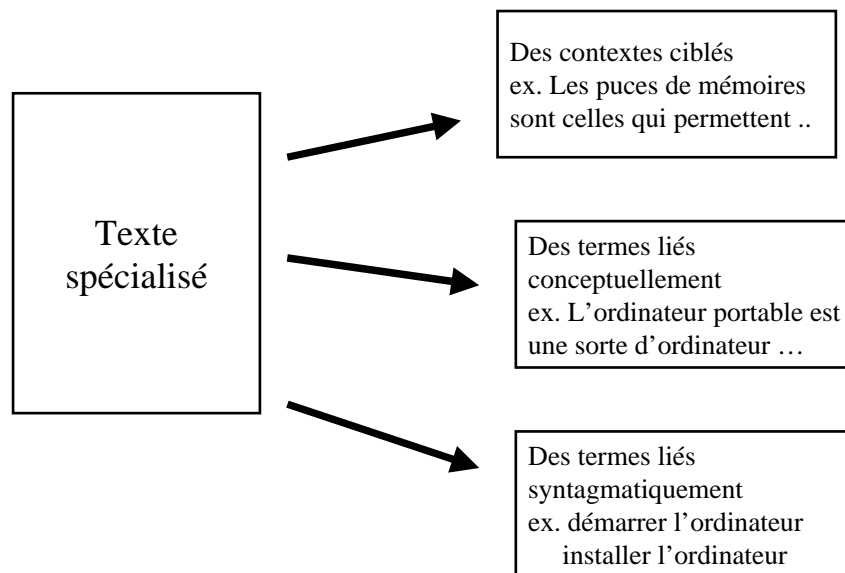


Figure 8 : Extraction d'informations de corpus : la version enrichie

Comme le montre la figure 8, les contextes dans lesquels figurent les termes peuvent être exploités pour dénicher de l'information complémentaire :

- des contextes ciblés : par exemple, des contextes contenant des définitions ou des éléments explicatifs plutôt que tous les contextes pris indifféremment ;
- des indices sur l'organisation conceptuelle d'un domaine : plus précisément sur les liens conceptuels partagés par des termes d'un domaine ;

- des affinités entre un terme et d'autres mots : des collocations ou des cooccurrents<sup>1</sup>.

Il y a différentes raisons pour distinguer la version de base et la version enrichie comme nous l'avons fait ci-dessus. Le premier groupe d'applications est plus largement diffusé que le second, a donné lieu à des logiciels commerciaux et a fait l'objet d'évaluations. Les applications faisant partie du second groupe ne sont pas encore véritablement sorties des laboratoires de recherche. Toutefois, elles présentent un intérêt pour la terminologie et c'est la raison pour laquelle nous nous y attarderons à la section 3.3.

## 3.2 Extraction de termes complexes

L'extraction de termes s'appuie sur les présupposés suivants (ou sur une partie d'entre eux selon la stratégie adoptée).

- a) Les textes spécialisés comportent beaucoup de termes qui servent de véhicules privilégiés des connaissances spécialisées;
- b) Un terme significatif sera utilisé à plusieurs reprises dans un texte spécialisé;
- c) La très grande majorité des termes sont de nature nominale;
- d) La plupart de ces termes sont complexes, c'est-à-dire qu'ils sont composés de plusieurs mots par ailleurs utilisés isolément (ex. *pression artérielle*; *intelligence artificielle*, *aigle à tête blanche*)
- e) Les termes complexes se construisent au moyen d'un nombre fini de séquences de catégories grammaticales. En effet, la plupart des termes complexes français se composent d'un nom modifié par (autrement dit, il s'agit d'un sous-ensemble de syntagmes nominaux) :
  - un adjectif : ex. *intelligence artificielle*, *haute tension*;

---

<sup>1</sup> Nous verrons également que les cooccurrents peuvent servir d'indices pour dégager des « familles conceptuelles ».

- un syntagme prépositionnel contenant un nom : ex. *robinet de commande, traitement de la demande*;
- un syntagme prépositionnel contenant un verbe : ex. *machine à coudre*;
- un autre nom : ex. *imprimante laser, page Web*;
- n'importe quelle combinaison des séquences ci-dessus : ex. *temps de conduction auriculaire*.

Pour faire identifier les termes complexes automatiquement, on fait appel à des techniques regroupées traditionnellement dans deux catégories : les stratégies *linguistiques* ou *statistiques*. Les premières associent des informations linguistiques à des chaînes de caractères ou font appel à des connaissances sur la langue traitée (au moins minimales) et recherchent, le plus souvent, des suites de catégories grammaticales. Les secondes effectuent des calculs sur les chaînes de caractères et s'appuient sur le fait que des termes significatifs sont employés forcément plus d'une fois dans un texte. Dans la plupart des outils existants, toutefois, connaissances linguistiques et calculs statistiques sont combinés selon différentes modalités.

Les stratégies linguistiques se fondent avant tout sur le fait que les termes complexes sont des syntagmes nominaux composés de suites de catégories grammaticales régulières. Ils repèrent donc des séquences de mots correspondant à des patrons préalablement définis. Par exemple, si on demande à un extracteur de termes de retenir les suites « nom + adjectif » et « nom + prép. + nom », il s'arrêtera sur deux séquences dans la phrase suivante et les placera dans une liste proposée à l'utilisateur.

*L'ordinateur portable est en général plus coûteux que l'ordinateur de bureau.*

*ordinateur de bureau  
ordinateur portable*

Une variante à cette première stratégie linguistique consiste à envisager le problème à l'envers. Plutôt que d'exploiter des connaissances « en positif », c'est-à-dire chercher des syntagmes nominaux correspondant à des suites de catégories grammaticales précises, on

exploite des connaissances « en négatif ». Cette technique, mise au point par Bourigault (1993) consiste à appliquer au texte une série de règles de découpage. Les coupes sont pratiquées là où le programme rencontre une unité qui ne peut pas faire partie de termes complexes, comme un signe de ponctuation fort, un pronom ou un verbe conjugué. Pour certains mots, le logiciel recourt à d'autres éléments de la phrase. Par exemple, certaines prépositions sont éliminées si elles sont suivies d'un possessif, les participes passés sont éliminés s'ils sont suivis d'une préposition. L'exemple ci-dessous montre comment les syntagmes nominaux sont délimités à partir de frontières.

*Le circuit d'aspersion de l'enceinte de confinement / assure le / maintien / de sa / température nominale de fonctionnement / après une / augmentation de pression (Bourigault 1993).*

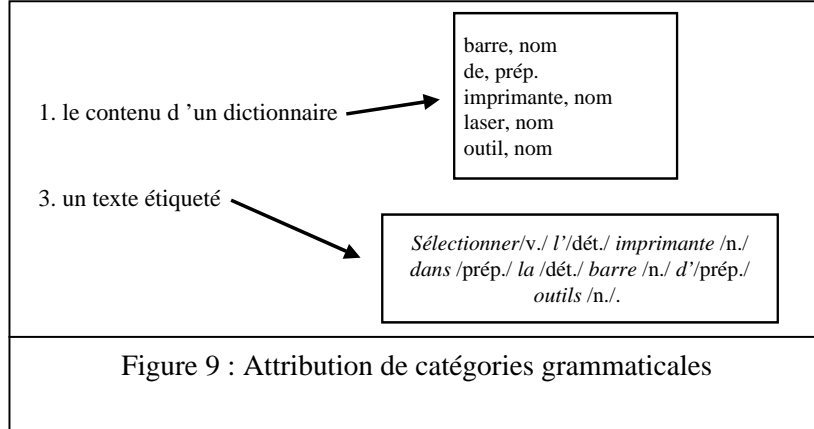
Les deux techniques linguistiques décrites ci-dessus s'appuient sur une reconnaissance des catégories grammaticales des mots dans le texte dépouillé. Cette reconnaissance se fait de deux manières :

- par la consultation d'un dictionnaire dans lequel on associe, à chacun des mots simples, sa catégorie grammaticale (certains mécanismes de désambiguïsation sont parfois implantés pour régler les problèmes posés par des mots ambigus) ;
- par la consultation d'un texte étiqueté renfermant les mentions explicites des catégories grammaticales (la désambiguïsation est alors faite par l'étiqueteur)<sup>2</sup>.

La figure 9 illustre ces deux méthodes d'explicitation des catégories grammaticales.

---

<sup>2</sup> Signalons que Nomino (David et Plante 1990) fait exception. Il effectue une analyse morphologique des mots contenus dans le texte dépouillé et c'est sur cette base qu'il attribue des catégories grammaticales.



Les stratégies statistiques (ou probabilistes) s'appuient principalement sur le fait qu'un terme significatif sera forcément utilisé plus d'une fois dans un texte spécialisé. Il existe plusieurs méthodes statistiques appliquées à l'extraction terminologique qui sont pour la plupart fondée sur un principe central, appelé *information mutuelle*. Grosso modo, ce principe veut que l'association récurrente de deux mots ne peut être que le fruit du hasard et est forcément significative.

Concrètement, les occurrences des mots d'un texte (c'est-à-dire les chaînes de caractères non interrompues) sont examinées de la manière suivante : si un mot X apparaît plus fréquemment dans l'entourage d'un mot Y qu'ailleurs dans le texte, alors X et Y forment une combinaison significative.

À quelques rares exceptions, les outils d'extraction de termes combinent deux stratégies. On parle alors d'extraction faisant appel à des stratégies hybrides ou mixtes : les extracteurs génèrent une liste de termes à partir d'informations linguistiques et épurent la liste au moyen de calculs statistiques; soit, au contraire, ils établissent une première liste de termes au moyen de calculs sur des chaînes de caractères et exploitent par la suite de l'information linguistique. Par exemple, ACABIT (Daille 1994) recherche des suites de



catégories grammaticales dans des textes étiquetés (connaissances linguistiques) et fait des calculs statistiques sur les termes préalablement extraits (stratégie probabiliste)<sup>3</sup>.

Peu importe la stratégie adoptée, les listes produites par les extracteurs ne sont pas parfaites. Elle comporte des suites qui ne sont pas celles qu'on recherche (ces erreurs sont regroupées sous le générique *bruit*). Par ailleurs, les extracteurs passent outre à des termes pourtant corrects. Ce deuxième groupe d'erreurs est appelé *silence*. Les concepteurs ne peuvent corriger toutes les erreurs, compte tenu de la difficulté liée à la tâche d'extraction, mais tentent de trouver des méthodes pour réduire le silence ou le bruit en fonction d'une application donnée.

Deux mesures, empruntées au domaine de la recherche d'information, servent à calculer la performance des extracteurs de termes. La *précision* évalue la proportion de termes corrects présents dans la liste produite par le programme; le *rappel* mesure la proportion de termes relevés par rapport aux termes présents dans le texte traité.

Depuis quelque temps, plusieurs travaux tentent de raffiner l'extraction de termes qui se limite souvent à la production d'une liste de termes complexes. Voici un aperçu de quelques aspects qui ont retenu l'attention :

- A) Extraction de termes simples : les termes simples sont extrêmement difficiles à repérer puisque rien dans leur forme ne permet de les distinguer des autres unités lexicales présentes dans les textes spécialisés. Toutefois, on croit que leur fréquence d'utilisation peut servir d'indicateur fiable. Une technique simple consiste à présenter les mots simples ordonnés par ordre décroissant de fréquence (avec exclusion des mots grammaticaux) en tenant pour acquis que les mots en tête de

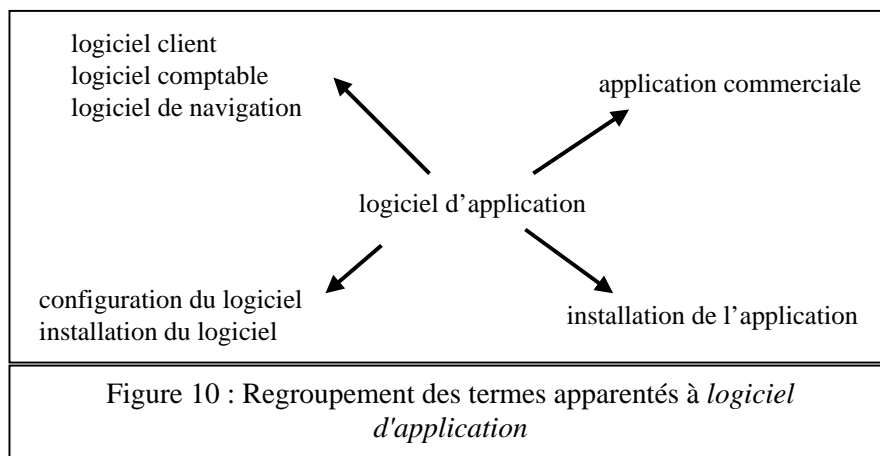
---

<sup>3</sup> Les logiciels d'extraction de termes se distinguent également par un certain nombre de caractéristiques que nous ne pouvons passer en revue : prise en compte ou non de la flexion, ordonnancement des termes dans les listes (tri alphabétique ou fréquence par ordre décroissant), appel à une ressources externe, comme un dictionnaire ou un thésaurus, etc. (Encore une fois, nous renvoyons le lecteur à l'Annexe A pour des références sur le mode de fonctionnement d'extracteurs particuliers).

liste sont révélateurs du contenu d'un texte. Une seconde stratégie consiste à comparer la fréquence des mots simples présents dans un corpus (par exemple, un corpus de textes informatiques) à celle de mots apparaissant dans un corpus de référence (par exemple, un corpus de textes journalistiques). Les mots dont la fréquence est nettement plus élevée dans le corpus de base risquent fort d'être des termes.

- B) Regroupements de termes préalablement extraits : différentes techniques tentent de rassembler des termes extraits en fonction de leur parenté formelle ou sémantique. Il est parfois difficile de retrouver manuellement les termes apparentés dans une liste qui les présente par fréquence décroissante ou par ordre alphabétique.

Une première technique consiste à retrouver des termes qui ont des composantes communes. La figure 10 montre comment les termes apparentés à *logiciel d'application* sont proposés à l'utilisateur. L'outil recherche tous les autres termes comportant *logiciel* (comme tête ou modificateur) et fait de même pour *application*. Cette liste peut être produite par simple comparaison graphique des termes entre eux.



D'autres travaux prennent en compte la variation terminologique, c'est-à-dire les multiples réalisations du terme qui peuvent aller d'un changement graphique (ex. *système expert, système-expert*) ou de transformations morphosyntaxiques (ex. *ulcère de cornée, ulcère cornéen; réseau pour données, réseau à données* (Daille 1995)) à des modifications syntaxiques beaucoup plus importantes, comme la coordination (ex. *artère rénale et artère coronaire* dans *artères fémorales, rénales et coronaires*) ou les transformations syntaxiques (ex. *concentrate measurement* dans *measured COHb concentration, measured the concentration* ou *measuring the concentration* (Jacquemin 2001)). L'appariement des variantes est nettement plus ardue à réaliser que le regroupement des termes ayant des composantes communes et repose sur un appareillage linguistique élaboré.

- C) Extraction bilingue : certains travaux portent sur l'extraction de termes et leur équivalent à partir de textes alignés (Gaussier 2000, entre autres) souvent dans le but d'assister la traduction. Les techniques mises au point font une extraction de termes au moyen de stratégies similaires à celles qui ont été décrites plus haut et cherchent par la suite à établir des équivalences en fouillant les phrases alignées dans un bitexte.

### 3.3 Exploitation de l'information contextuelle

Comme nous l'avons mentionné à la section 3.1, le terminographe peut extraire des contextes présents dans un corpus spécialisé à partir d'un terme qu'il saisit au clavier. S'il travaille avec un concordancier qui établit des correspondances entre chaînes de caractères, il effectue lui-même l'interprétation de l'information qui est donnée dans les contextes.

De nouvelles applications informatiques tentent d'exploiter l'information présente dans les contextes de diverses manières. Nous en examinerons trois dans cette section.

- A) Recherche de contextes ciblés : Le terminographe praticien sait que les contextes ne renferment pas tous des renseignements d'égale valeur sur les termes qu'ils renferment. Certains contextes, comme ceux que nous avons reproduits ci-dessous, sont nettement plus « utiles » que d'autres, car ils renferment des indices sur le sens du terme.

A) contextes définitoires

*Elle affecte la rétine -- le tissu nerveux au fond de l'oeil qui transmet les messages visuels au cerveau (red\_diab)*

*a green substance in leaves called chlorophyll. (contexte cité dans Pearson 1998)*

B) contextes contenant des termes liés conceptuellement

*Un disque virtuel est une partie de la mémoire configurée comme une disquette.  
ZMSPC2*

*Compost contains nutrients, nitrogen, potassium and phosphorous (contexte cité dans Meyer 2001).*

*Sulphur dioxide is a gas given off by some fuels (contexte cité dans Pearson 1998).*

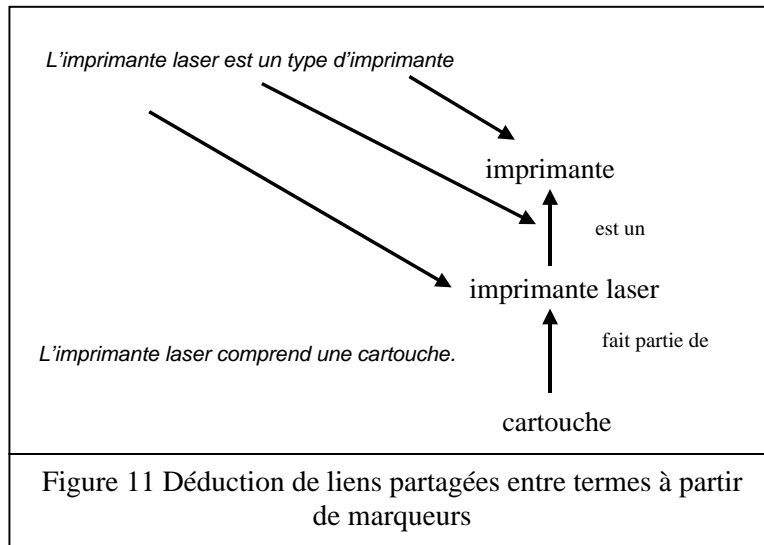
Les travaux de Pearson (1998) et d'autres (notamment Meyer 2000) ont défini des méthodes de ciblage des contextes « riches en connaissances » (knowledge-rich contexts, Meyer 2000). L'idée générale consiste à repérer des marqueurs explicites (des mots ou des suites de mots) qui servent d'indices annonçant qu'un contexte livre des informations sur un terme. Quelques exemples de marqueurs sont donnés ci-dessous :

*Anglais : is called, known as, also called, is defined as, consists of, contains, is part of, is a, is a type of, is a form of, e.g., i.e., :, (\*)*

*Français : est appelé, se définit de la manière suivante, fait partie de, est constitué de, contient, est un, est un type de, ex., c'est-à-dire, :, (\*)*

- B) Détection automatique de liens conceptuels : des chercheurs (Jouis 1995; Morin 1999) ont élaboré des programmes pour repérer automatiquement, dans les contextes, des indices de liens conceptuels (des liens d'hyponymie ou de méronymie, par exemple). Ils utilisent comme Pearson (1998) et Meyer (2000) des marqueurs explicites. Toutefois, dans leur travaux, les marqueurs sont utilisés pour déduire automatiquement les liens conceptuels. La figure 11 illustre

comment deux phrases peuvent être analysées par un programme de ce genre pour tirer des conclusions sur les liens partagés entre les termes.



- C) Repérage de groupes de mots significatifs : une nouvelle catégorie d'applications recherche dans les textes non plus des termes complexes mais d'autres associations de mots significatives comme les collocations ou les cooccurrents. En terminographie, ces groupes révèlent les unités lexicales avec lesquelles un terme se combine préférentiellement. Les phrases suivantes montrent le terme *accusation*, emprunté au domaine du droit, et ses cooccurrents :

*décide de porter une accusation différente, mais pour la même affaire; ou encore lorsque, l'accusé ayant été libéré à la suite de son enquête préliminaire, (Morel)*  
*Se référant au droit commun, il rappelle que dans une mise en accusation, on incorpore les faits reprochés et le droit spécifiquement enfreint. (Hebert 1984).*  
*Il en va de même lorsqu'une nouvelle accusation est portée après que l'avocat de la poursuite eut ordonné d'arrêter les procédures (Morel)*  
*Et la gravité de l'accusation peut avoir cet effet aussi (Chevre)*

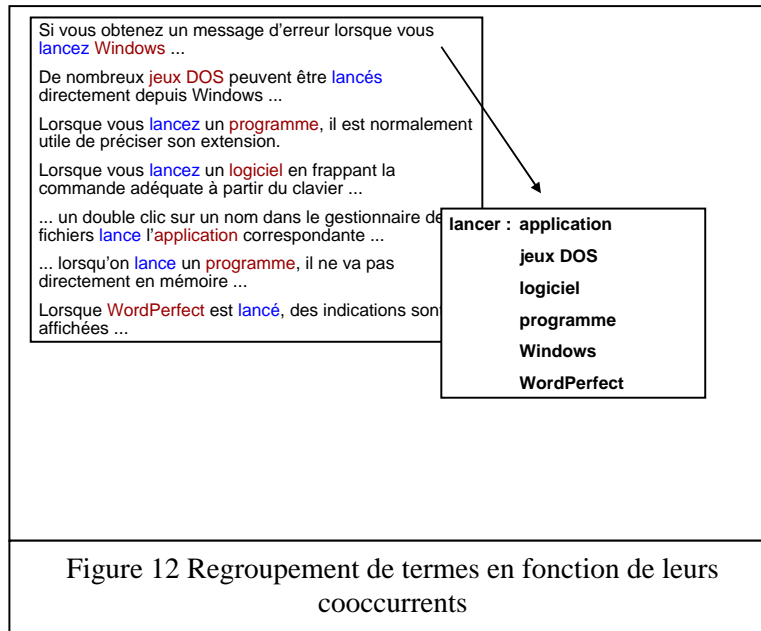
Certains outils, qui n'ont pas été mis au point spécifiquement pour l'exploitation de textes spécialisés, sont conçus pour extraire ces groupes de mots. Ils utilisent des techniques qui s'apparentent aux techniques d'extraction de termes complexes.

Smadja (1993) a recours à des calculs inspirés de l'information mutuelle assortis d'un marquage des fonctions des mots; Kilgariff et Kugwell (2001) repèrent des patrons reflétant les constructions recherchées et épurent ces premiers résultats au moyen de calculs probabilistes.

Il convient de souligner que le repérage des collocations est compliqué notamment par le fait que les mots les composant peuvent être permutés ou séparés par plusieurs autres mots. Enfin, notons que les cloisons entre extracteurs de collocations et extracteurs de termes ne sont pas étanches : les extracteurs de collocations relèvent des termes complexes; les extracteurs de termes complexes relèvent forcément des collocations. Ce chevauchement s'explique en grande partie par la similitude des techniques utilisées.

Mentionnons que des travaux récents se penchent plus précisément sur certains types de combinaisons de mots, notamment l'extraction de syntagmes verbaux (ex. *disséquer le plateau rocheux en chevrons, observer une charge excessive en trouble* (Bourigault et Fabre 2000)).

Dans certains programmes, les cooccurrents servent d'indicateurs sur la parenté conceptuelle des termes et sont utilisés pour regrouper des ensemble de termes. On se fonde sur le fait que si des termes ont les mêmes cooccurrents, il partagent des caractéristiques sémantiques communes ou appartiennent à la même classe. La figure 12 illustre cette technique : une liste de termes ayant un cooccurrent commun, à savoir *lancer*, est proposée à l'utilisateur et signalent une famille conceptuelle probable.



#### 4. Impact sur le travail du terminographe

Les descriptions de la section 3, bien que succinctes, montrent que les terminographes ont désormais accès à toutes sortes de techniques pour repérer et extraire des informations présentes sous différentes formes dans les textes spécialisés : extraction de termes (complexes, mais également simples); repérage de collocations, repérage de contextes ciblés, établissement de liens conceptuels entre termes ou construction de classes.

A priori, ces outils paraissent utiles, voire indispensables, dans un environnement de travail où l'accès à des textes spécialisés en format électronique fait partie du quotidien. De plus, ils permettent de dégager des données particulièrement utiles dans un contexte où les descriptions terminographiques sont de plus en plus riches (voir la section 2).

D'abord, les applications informatiques peuvent mouliner des volumes de textes imposants beaucoup plus rapidement que le terminographe. De plus, comme les textes sont

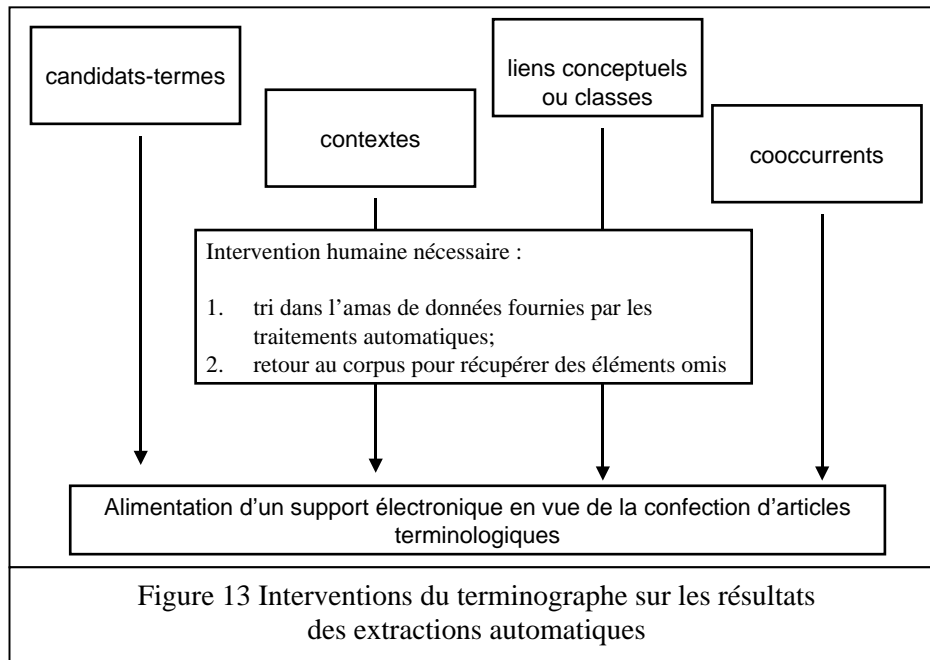
extrêmement riches en information, on peut penser que les traitements automatiques arrivent plus facilement que l'humain à faire le tri dans tout ce qui s'y trouve.

Cependant, même si ce qui est mentionné dans le paragraphe précédent est vrai, la question de l'utilité de ces techniques d'extraction est plus complexe et nécessite quelques précisions. Il n'est pas possible, à ce stade, d'entrevoir tous les impacts que ces outils ont sur la terminographie, puisque leur utilisation n'est pas généralisée. En outre, comme nous l'avons déjà mentionné, certains logiciels sont utilisés par les terminographes, mais n'ont pas été mis au point pour eux.

Nous allons tout de même nous attarder sur certaines conséquences méthodologiques qu'entraîne l'introduction des techniques d'extraction dans la chaîne de travail terminographique présentée à la figure 5.

Mentionnons en premier lieu qu'aucun de ces outils ne produit de résultats parfaits. Le terminographe doit intervenir de deux manières sur les résultats produits : 1) il élimine, de la liste des résultats, des éléments indésirables; 2) il retourne au corpus pour retrouver un élément qu'un traitement automatique n'a pas relevé. La figure 13 montre comment le terminographe est appelé à intervenir.





Or, ces deux tâches peuvent exiger beaucoup de temps en fonction de la qualité des résultats produits. Pour donner un aperçu de cette qualité, examinons les performances de logiciels d'extraction de termes complexes. Nous avons reproduit, à la figure 14, les résultats d'évaluations menées par des utilisateurs.

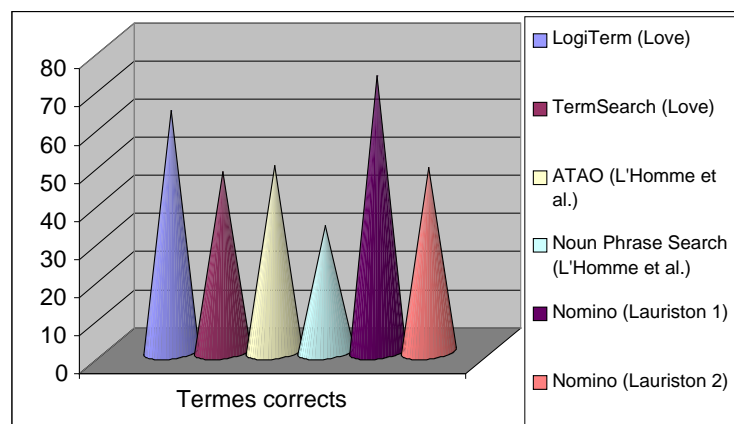


Figure 14 Précision de logiciels d'extraction de termes

Ces résultats, bien que sommaires, montrent que, dans le meilleur des cas, le terminographe devra éliminer, d'une liste produite par un logiciel d'extraction, 30 % des

unités retenues. Une des listes évaluées nécessite une élimination de 70 % des unités retenues, ce qui représente un travail considérable. Mentionnons, par ailleurs, que les résultats présentés ici ne portent que sur la précision (on ne tient pas compte du rappel).

L'utilisation de logiciels d'extraction de données a des impacts méthodologiques d'une autre nature. En recherche terminographique classique, le repérage des termes et le relevé de contextes dans lesquels ils figurent se faisaient simultanément. Le terminographe parcourait la documentation, relevait un terme potentiel ainsi que la phrase dans laquelle il apparaissait. En recherche automatisée, le repérage de contextes succède à l'extraction de termes. Le retour au contexte est souvent indispensable pour prendre une décision sur la nature « terminologique » d'une suite de mots.

De plus, en recherche classique, une grande partie des connaissances spécialisées étaient acquises lors du repérage des termes et de contextes. En recherche assistée par ordinateur, cette acquisition doit se faire à un autre moment (idéalement, avant l'application d'un traitement automatique). Par ailleurs, les décisions relatives au statut terminologique d'une suite de mots dans une liste produite par un extracteur nécessitent d'excellentes connaissances du domaine et de sa terminologie. Ces connaissances doivent être encore plus fines lorsqu'il s'agit de prendre une décision sur l'établissement d'une relation conceptuelle ou l'à-propos d'une collocation.

Une autre question centrale demeure et n'a pas encore été prise en considération dans les travaux d'évaluation : les outils d'extraction terminologique permettent-ils au terminographe de gagner du temps ? Autrement dit, est-il plus rapide de générer des listes de candidats-termes automatiquement et de les épurer par la suite que de relever des termes manuellement ?

On dispose de renseignements sur la qualité des listes produites par les extracteurs de termes (qui sont comparées à des listes humaines) et si ces évaluations permettent de connaître les forces et les faiblesses des logiciels et indiquent aux concepteurs où ils

doivent intervenir, elles ne donnent pas d'indication sur le « gain de temps ». Par ailleurs, on ne connaît pas avec précision l'accueil réservé à ces outils par les terminographes. Contrairement aux concordanciers qui sont immédiatement appréciés, l'extraction de termes complexes semble recevoir un accueil plus mitigé.

Quand aux autres applications abordées aux sections 3.2 et 3.3, elles n'ont pas encore fait l'objet d'évaluations systématiques. On n'a pas encore d'idée précise sur la qualité des résultats produits par un extracteur de collocations ou un programme qui génère des classes de termes, par exemple. Toutefois, il est permis de penser que ces extractions sont nettement plus difficiles à mettre en œuvre que l'extraction de termes complexes puisqu'elles recherchent des informations nettement plus difficiles à isoler sans ambiguïté.

## 5. Conclusion

En conclusion, il importe en tout premier lieu de souligner qu'une réflexion sur les différents aspects de l'intégration des outils d'extraction de données terminologiques s'impose. Nous disposons de renseignements sur la performance de certains logiciels d'extraction de termes, mais d'autres questions restent en suspens. Voici quelques pistes de réflexion<sup>4</sup> :

- évaluation du rapport temps extraction automatique – temps humain;
- évaluation de l'accueil réservé à certains outils d'extraction par les terminologues-terminographes;
- étude portant sur l'intégration des diverses techniques d'extraction dans un poste de travail terminographique.

---

<sup>4</sup> Certaines pistes ont été proposées lors du Colloque sur l'impact des nouvelles technologies sur la gestion terminologique qui a eu lieu à Toronto, le 18 août 2001. Nous aimerions remercier les participants qui nous les ont signalées.

Même si le recours à l'informatique, peu importe son intérêt, est incontournable en raison du volume croissant de textes électroniques, il convient de mieux définir sa place dans une séquence de tâches terminologiques. De nouvelles méthodes de travail doivent être étudiées et, inversement, on doit voir comment les logiciels peuvent mieux s'adapter aux tâches normalement réalisées par le terminographe. De plus, il faut souligner, que toutes les données terminologiques ne peuvent être extraites d'un corpus (ex. données chronologiques, géographiques ou d'usage) à moins que le corpus n'ait été préalablement caractérisé. L'expertise du terminographe est nécessaire à toutes les étapes et le recours à des informateurs spécialistes reste indispensable. Ce fait semble aller de lui-même mais on a tendance à l'occulter dans les discussions portant sur l'intégration de nouvelles technologies en terminographie.

Par ailleurs, les outils d'extraction (et d'autres outils, notamment les outils de consignation des données) amènent le terminographe à envisager d'autres types de données terminologiques. Cette nouvelle donne doit également être prise en considération dans une redéfinition des méthodes de travail.

Enfin, la formation en terminologie dispensée dans les universités, ne peut plus se faire sans tenir compte des nouvelles technologies. Elle doit être revue en profondeur et mettre l'accent sur la recherche terminographique telle qu'elle est réalisée aujourd'hui. Il ne s'agit plus de donner une formation traditionnelle en prévoyant un bref exposé sur quelques méthodes de recherche documentaire ou sur corpus. La formation doit intégrer des notions de traitement automatique de langue minimales, de gestion de corpus et d'organisation des données dans des bases de données ou au moyen de langages documentaires. Il faut également rendre le futur terminologue apte à exploiter l'information numérique et en tirer le maximum.

## Bibliographique

- Assadi, H. (1997). « Knowledge Acquisition from Texts: Using an Automatic Clustering Method Based on Noun-modifier Relationship », In Proceedings of the 35<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and 8<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (ACL–EACL '97), Madrid, pp. 504-506.
- Auger, P., P. Drouin et M.C. L'Homme (1991). « Un projet d'automatisation des procédures en terminographie », *Meta* 36(1), pp. 121-127.
- Binon, J., S. Verlinde, J. Van Dyck et A. Bertels (2000). Dictionnaire d'apprentissage du français des affaires. Dictionnaire de compréhension et de production de la langue des affaires, Paris : Didier.
- Bourigault, D. (1993). « An Endogenous Corpus-based Method for Structural Noun Phrase Disambiguation », In Proceedings of the 6th Conference of the European Chapter of the Association of Computational Linguistics (EACL '93), Utrecht, pp. 81-86.
- Bourigault, D. et C. Fabre (2000). « Approche linguistique pour l'analyse syntaxique de corpus », *Cahiers de grammaire* 25, pp. 131-151.
- Bourigault, D., C. Jacquemin et M.C. L'Homme (Eds.) (2000). *Recent Advances in Computational Terminology*, Amsterdam – Philadelphia, John Benjamins.
- Cabré, M.T., R. Estopa and J. Vivaldi (2000). "Automatic Term Detection: A Review of Current Systems", In Bourigault, D., C. Jacquemin et M.C. L'Homme (Eds.) (2000). *Recent Advances in Computational Terminology*, Amsterdam – Philadelphia, John Benjamins, pp. 53-87.
- Daille, B. (1994). *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques*, thèse de doctorat, Paris : Université Paris VII.
- Daille, B. (1995). « Repérage et extraction de terminologie par une approche mixte statistique et linguistique », *Traitement automatique des langues (TAL)* 36(1-2), pp. 101-118.
- David, S. et P. Plante (1990). « De la nécessité d'une approche morpho-syntaxique dans l'analyse des textes », *Intelligence artificielle et sciences cognitives au Québec* 3(3), pp. 140-154
- Drouin, P. et J. Ladouceur (1994). « L'identification automatique des descripteurs complexes dans les textes de spécialité », In Bouillon, P. et D. Estival (Eds.). *Proceedings of the Workshop on Compound Nouns. Multilingual Aspects of Nominal Composition*, 2-3 Dec. 1994, Université de Genève (Genève), pp. 18-28.
- Dubuc, R. (1986). *Manuel pratique de terminologie*, Brossard (Québec) : Linguatex.
- Gaussier, É. (2000). *General Considerations on Bilingual Terminology Extraction*, Bourigault, D., C. Jacquemin et M.C. L'Homme (Eds.) (2000). *Recent Advances in Computational Terminology*, Amsterdam – Philadelphia, John Benjamins, pp. 167-183.
- Habert, B., E. Naulleau A. Nazarenko (1996). *Symbolic Word Clustering for Medium-size Corpora*, In Proceedings of the 16<sup>th</sup> International Conference on Computational Linguistics (COLING '96), Copenhagen, pp. 490-495.
- Habert, B. et C. Jacquemin (1993). « Noms composés, termes, dénominations complexes : problématiques linguistiques et traitements automatiques », *Traitement automatique des langues (TAL)* 34(2), pp. 5-42.

- Jacquemin, C. (2001). *Spotting and Discovering Terms through Natural Language Processing*, Cambridge, MIT Press.
- Jouis, C. (1995). "SEEK, un logiciel d'acquisition des connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe", In *Actes des 6<sup>e</sup> Journées Acquisition, validation (JAVA 95)*, Grenoble, pp. 159-172.
- Justeson, J.S. et S.M.Katz (1995). "Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text", *Natural Language Engineering* 1(1), pp. 9-27.
- Kilgariff, A. & D. Tugwell (2001). "Word Sketch: Extraction and Display of Significant Collocations for Lexicography", In *Workshop Proceedings. Collocation: Computational Extraction, Analysis and Exploitation*, Toulouse, pp. 32-38.
- Lauriston, A. (1994). "Automatic Recognition of Complex Terms: The Terminology Solution", *Terminology* 1(1), pp. 147-170.
- L'Homme, M.C., L. Benali, C. Bertrand and P. Lauduique (1996). « Definition of an Evaluation Grid for Term-extraction Software », *Terminology* 3(2), pp. 291-312.
- Love, S. (2000). *Benchmarking the Performance of Two Automated Term-extraction Systems*, Mémoire de maîtrise, Montréal : Université de Montréal.
- Meyer, I. (2000). "Extracting Knowledge-rich Contexts for Terminography: A Conceptual and Methodological Framework", ", In Bourigault, D., C. Jacquemin et M.C. L'Homme (Eds.) (2000). *Recent Advances in Computational Linguistics*, Amsterdam – Philadelphia, John Benjamins, pp. 279-302.
- Meyer, I., D. Skuce, L. Bowker and K. Eck (1992). "Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge base", In *Proceedings of the 16<sup>th</sup> International Conference on Computational Linguistics Coling '92*, Nantes, pp. 956-957.
- Meynard, I. (2000). *Internet. Répertoire bilingue de combinaisons lexicales spécialisées français-anglais*, Brossard (Québec) : Linguatex.
- Morin, E. (1999). « Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique », *Traitement automatique des langues (TAL)* 40(1), pp. 172-181.
- Otman, G. (1991). « Des ambitions et des performances d'un système de dépouillement terminologique assisté par ordinateur », *La banque des mots* 4, pp. 56-96.
- Otman, G. (1996). *Les représentations sémantiques en terminologie*, Paris : Masson.
- Pearson, J. (1998). *Terms in Context*, Amsterdam – Philadelphia, John Benjamins.
- Quillian, R. (1968). "Semantic Memory", In Minsky, E. (Ed.) *Semantic Information Processing*, Cambridge : MIT Press, pp. 227-270.
- Rondeau, G. (1984). *Introduction à la terminologie*, Chicoutimi (Québec) : Gaëtan Morin.
- Sager, J.C. (1990). *A Practical Course in Terminology Processing*, Amsterdam – Philadelphia, John Benjamins.
- Smadja, F. (1993). "Retrieving Collocations from Text: Xtract", *Computational Linguistics* 19(1), pp. 143-177.
- Verrier, D. (2000). *Recherche terminographique en traduction juridique réalisée au moyen d'outils informatiques*, Travail dirigé, Montréal : Université de Montréal.

## ANNEXE A

### Logiciels d'extraction de données terminologiques

#### Extraction de termes (complexes)

Acabit : français à l'origine avec versions mises au point pour d'autres langues (Daille 1994).

Lexter L : Extraction de termes complexes français (Bourigault 1993) (exploitations ultérieures de Lexter : constructions de classes, Assadi 1996); extraction de syntagmes verbaux, Bourigault et Fabre 2000).

NOMINO (anc. Termino) : extraction de termes complexes français (David et Plante 1990) et catégorisation des autres mots pleins.

TermPlus (anc. Notions) : extraction de termes complexes anglais et français avec exploration des termes reliés formellement (Drouin et Ladouceur 1994 et communications personnelles)

Terms : extraction de termes anglais (Justeson et Katz 1995).

#### Autres types d'extraction

Termes et leurs variantes : FASTR (Jacquemin 2001).

Construction de classes conceptuelles : LexiClass (Assadi 1997); ZELLIG (Habert *et al.* 1996)

Établissement de relations sémantiques entre termes : Promothée (Morin 1999); Seek-Java (Jouis 1995).

Extraction de collocations (avec une visée générale) : Xtract (Smadja : 1993); Word Sketch (Kilgariff & Tugwell 2001).

Extraction de syntagmes verbaux (Bourigault et Fabre 2000).

## Notice biographique

Marie-Claude L'Homme a obtenu un Ph.D. en linguistique de l'Université Laval en 1992 et travaille depuis 1994 à l'Université de Montréal au Département de linguistique et de traduction après avoir travaillé près de deux ans en entreprise comme terminologue-documentaliste. À l'Université de Montréal, elle est chargée de former les traducteurs aux outils informatiques d'aide à la traduction et est co-responsable du Laboratoire de linguistique-informatique du département. Sa recherche porte sur l'évaluation des outils informatiques pour traducteurs et terminologues et sur l'élaboration de descriptions terminologiques qui intègrent des modèles sémantiques formels. Elle a publié des articles dans *Terminology*, *Terminologies nouvelles*, *Cahiers de lexicologie* ainsi qu'un manuel intitulé *Initiation à la traductique* destiné à vulgariser les outils informatiques. Elle est, depuis 2001, co-éditrice de la revue *Terminology*.

## Biographical note

Marie-Claude L'Homme has obtained a Ph.D. in Linguistics at Laval University in 1992, and works since 1994 at the University of Montreal. She had worked almost two years in the private sector as a terminologist. At the University of Montreal, she is responsible for the training of translators to computer tools and is co-responsible for the Laboratoire de linguistique informatique (computational linguistics laboratory) of the department. Her research focuses on the evaluation of computer tools for translators and terminologists and on the development of terminological descriptions based on formal semantics. She has published papers in *Terminology*, *Terminologies nouvelles*, *Cahiers de lexicologie*, and a textbook entitled *Initiation à la traductique* designed to explain computer tools. She was appointed, in 2001, co-editor of the journal *Terminology*.