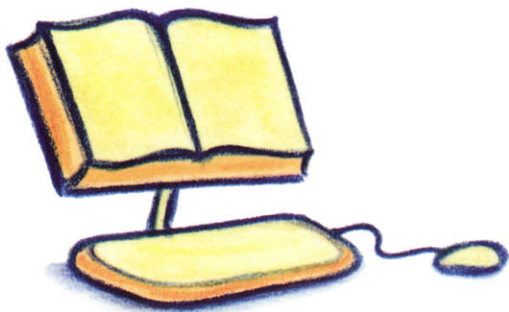




paramètres

Marie-Claude L'Homme

# La terminologie : principes et techniques



Extrait de la publication  
Les Presses de l'Université de Montréal

## **LA TERMINOLOGIE : PRINCIPES ET TECHNIQUES**



MARIE-CLAUDE L'HOMME

**LA TERMINOLOGIE :  
PRINCIPES ET TECHNIQUES**

**Les Presses de l'Université de Montréal**

Extrait de la publication

*Données de catalogage avant publication (Canada)*

L'Homme, Marie-Claude

La terminologie : principes et techniques  
(Paramètres)

Comprend des réf. bibliogr. et un index.

ISBN 2-7606-1949-4

1. Terminologie. I. Titre. II. Collection.

P305.L46 2004 401'.4 C2004-940869-0

Dépôt légal : 3<sup>e</sup> trimestre 2004

Bibliothèque nationale du Québec

© Les Presses de l'Université de Montréal, 2004

Les Presses de l'Université de Montréal remercient de leur soutien financier le ministère du Patrimoine canadien, le Conseil des Arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

**IMPRIMÉ AU CANADA EN AOÛT 2004**

# TABLE DES MATIÈRES

Remerciements	11
Liste des abréviations et des symboles	13
Conventions typographiques	14
<b>Introduction</b>	<b>15</b>
Un manuel de terminologie ou de terminographie ?	15
Un manuel de terminographie ou de terminotique ?	16
Revisiter la terminologie et la terminographie	18
De nouvelles pratiques	19
<b>1. Terminographie, terminologie et terminotique</b>	<b>21</b>
Qu'est-ce que la terminographie ?	21
Les modèles théoriques de la terminologie	23
<i>L'optique classique « conceptuelle »</i>	25
<i>Les reproches faits à l'optique conceptuelle</i>	29
<i>L'optique lexico-sémantique</i>	32
<i>Une optique meilleure que l'autre pour la terminographie ?</i>	36
Les données terminologiques	38
Les sept travaux du terminographe	45
La terminotique	47
Conclusion	50
Suggestions de lectures	51

<b>2. Le terme dans le texte spécialisé</b>	<b>52</b>
Qu'est-ce qu'un terme ?	52
Le domaine comme paramètre classificateur des sens	53
Fréquence et répartition	57
Les types de termes	58
Le nom : partie du discours centrale en terminologie	59
Prédicats sémantiques et actants sémantiques	62
Des critères pour identifier les termes	64
Réalizations dans les textes spécialisés	66
Distinguer les sens dans un texte spécialisé	69
Le terme dans la phrase et dans le texte	72
<i>Des formes qui varient</i>	73
<i>Des termes complexes à repérer</i>	75
<i>Les termes d'une phrase à l'autre</i>	79
Conclusion	81
Suggestions de lectures	81
<b>3. Structures terminologiques</b>	<b>83</b>
Importance des structures terminologiques	83
Classements thématiques	84
Représentations conceptuelles	86
Relations lexico-sémantiques classiques	90
<i>Relations taxinomiques</i>	91
<i>Synonymie et quasi-synonymie</i>	94
<i>Antonymie et incompatibilité</i>	96
<i>Relations méronymiques</i>	98
Autres relations lexico-sémantiques	102
Forme et sens	107
Combinatoire et classes de termes	111
Collocations	112
Relations interlinguistiques	115
Conclusion	117
Suggestions de lectures	117
<b>4. Le corpus spécialisé</b>	<b>119</b>
Le texte en terminographie	119
Corpus « général » et corpus spécialisé	123
<i>Critères de sélection des textes composant un corpus spécialisé</i>	125
<i>Taille et équilibre</i>	128
Recherche de textes en format électronique	129
Corpus bilingues et multilingues	131

<i>Corpus alignés</i>	131
<i>Corpus comparables</i>	133
Étiquetage	134
<i>Pourquoi étiqueter les corpus ?</i>	134
<i>Réalisation de l'étiquetage</i>	136
Conclusion	140
Suggestions de lectures	140
<b>5. Le corpus spécialisé : interrogation</b>	<b>142</b>
Aborder les corpus sous des angles différents	142
Concordanciers : recherche, affichage et gestion des textes	143
Interrogation d'un corpus unilingue	146
<i>Indexation</i>	146
<i>Recherche d'une partie de terme</i>	151
<i>Recherche de deux ou trois termes</i>	152
<i>Opérateurs de proximité</i>	154
<i>Recherche de contextes « riches en connaissances »</i>	155
<i>Organisation des contextes</i>	157
Interrogation de corpus alignés	160
Interrogation de corpus étiquetés	162
Conclusion	165
Suggestions de lectures	165
<b>6. Extraction de termes</b>	<b>166</b>
Automatiser le plus possible...	166
Quels sont les indices sur lesquels peut s'appuyer un extracteur de termes ?	167
Repérer des termes fréquents par comparaison à un autre corpus	169
Extraire des chaînes de caractères apparaissant « souvent » ensemble	173
<i>Rechercher des segments répétés</i>	174
<i>Mesurer le degré d'association</i>	175
<i>Les apports des stratégies d'extraction de mots apparaissant souvent ensemble</i>	179
Extraire des séquences de parties du discours	180
<i>Chercher des patrons typiques</i>	180
<i>Isoler des termes au moyen de frontières</i>	183
<i>Les apports des stratégies d'extraction de séquences de parties du discours</i>	184
Combiner les séquences de parties du discours à la fréquence	185
Ordonnement des termes extraits	188
Problèmes courants des extracteurs de termes	192



Rectifier le tir de différentes manières	195
Conclusion	199
Suggestions de lectures	200
<b>7. Trouver d'autres données terminologiques</b>	<b>201</b>
Repousser les limites...	201
Extraction de collocations	202
Extraction bilingue	207
Trouver de l'information sémantique	210
<i>Établir des relations à partir de la forme</i>	211
<i>Établir des relations à partir du contexte</i>	217
Utiliser des ressources externes	223
Conclusion	225
Suggestions de lectures	226
<b>8. Organisation des données terminologiques</b>	<b>227</b>
Gérer des données terminologiques	227
Les conditions	228
Environnement « contraint » :	
le système de gestion de bases de données	230
Environnements « flexibles » : documents structurés	233
L'environnement hypertexte	236
Environnements spécialisés : logiciels de terminologie et banques de terminologie	238
Environnements intégrés : extraction et gestion de données terminologiques	240
Les modèles d'encodage des données terminologiques	240
<i>Le modèle « plat »</i>	244
<i>Le modèle relationnel</i>	247
<i>Le modèle hiérarchique</i>	254
Conclusion	257
Suggestions de lectures	258
<b>Bibliographie</b>	<b>259</b>
<b>ANNEXE A :</b>	
Liste des dictionnaires et banques de terminologie cités	268
<b>ANNEXE B :</b>	
Liste des logiciels et programmes cités	270
Index des noms d'auteurs	272
Index des notions	274

*À Camille et Denis,*

*C'est sans doute à vous que je dois cette propension  
à me lancer dans des entreprises un peu folles.*

*Page laissée blanche*

## REMERCIEMENTS

Je remercie en tout premier lieu les personnes qui ont accepté de donner leur avis de manière tout à fait désintéressée sur une version ou l'autre de ce travail, à savoir Patrick Drouin, Francis Grossmann, Igor Mel'čuk, Alain Polguère, Jean Quirion et Juan C. Sager. J'ai tenté d'intégrer l'ensemble de leurs suggestions. Les oublis, omissions ou distractions ne sont imputables qu'à moi-même.

Ma conception de la terminologie s'est considérablement modifiée au cours des années et il serait impossible de citer toutes les personnes qui y ont contribué. J'ai une pensée particulièrement chaleureuse pour tous ceux (chercheurs, professeurs, terminologues praticiens, étudiants, linguistes et informaticiens) qui, en posant une question ou en formulant une critique, m'ont amenée à remettre en question certains présupposés. Ceux que je n'aurai pas cités sauront certainement trouver des traces de nos discussions dans les pages du manuel.

J'adresse des remerciements particuliers aux membres du groupe Observatoire de linguistique Sens-Texte (OLST). L'influence de la Théorie Sens-Texte sur mes travaux et, en particulier de son volet lexicologique – la lexicologie explicative et combinatoire –, dépasse largement les quelques références explicites apparaissant dans ce manuel.

J'aimerais également remercier des collègues européens avec qui j'ai eu l'occasion de travailler sur des thématiques liées à la terminologie. Je pense aux collègues de la Maison de la recherche de Toulouse-le-Mirail, de l'Université de Nantes et de l'Institut universitaire de linguistique appliquée (IULA) de Barcelone.

Enfin, ma participation à la rédaction de la revue *Terminology* et les discussions avec ses auteurs et les membres de ses comités éditorial et scientifique me rappellent constamment l'étendue et l'intérêt des problèmes relevant de la terminologie et les applications extrêmement nombreuses de ce domaine.

## Liste des abréviations et des symboles

ADJ	adjectif
ADJAC	adjacent : opérateur indiquant une relation de contiguïté entre les conditions
ADV	adverbe
ART	article
CARD	cardinal
CONJ	conjonction
CONJUG	conjugué (dans <i>verbe conjugué</i> )
COORD	coordination (dans <i>conjonction de coordination</i> )
DÉT	déterminant
ET	opérateur combinant deux conditions nécessaires
FÉM	féminin
fréq.	fréquence
HTML	Hypertext Markup Language
IM	information mutuelle
IND	indicatif
INF	infinitif
MASC	masculin
<i>n</i>	nombre indéfini
NC	nom commun
OU	opérateur combinant deux conditions dont l'une ou l'autre peut être remplie
P	probabilité
PART	participe
PASSÉ	passé
PERS	personne
PONCT	ponctuation
POSS	possessif
PLUR	pluriel
PRÉP	préposition
PRÉS	présent
PRON	pronom
SAUF	opérateur excluant la condition qui lui succède
SGBD	système de gestion de bases de données
SGML	Standard Generalized Markup Language
SING	singulier
SUB	subordination (dans <i>conjonction de subordination</i> )
Syn.	synonyme
VERBE	verbe
XML	eXtensible Markup Language

## Conventions typographiques

<i>mot</i>	forme linguistique
« mot »	sens ou concept
<b>MOT</b>	terme ou unité lexicale, c'est-à-dire forme linguistique assortie d'un sens spécifique ; la particularité du terme est que son sens est envisagé par rapport à un domaine de spécialité
<b>MOT<sub>1</sub></b> , <b>MOT<sub>2</sub></b>	distinction des sens pour une même forme linguistique lorsque le contexte l'exige
<b>mot</b> <sub>étiquette</sub>	mot assorti d'une étiquette morphosyntaxique
<b>mo*</b> ou <b>*ot</b>	mot tronqué
<b>*phrase</b>	phrase inacceptable
<b>?phrase</b>	phrase dont l'acceptabilité peut être mise en doute
<...>	balise ouvrante (utilisée dans les documents structurés)
</...>	balise fermante (utilisée dans les documents structurés)
[ <b>domaine</b> ]	domaine de connaissance auquel est associé le sens d'un terme
{ <b>champ</b> }	champ dans une base de données
[...]	citation ou contexte tronqué

Lorsqu'une notion importante est expliquée dans le texte, elle est mise en évidence au moyen d'italiques et de caractères gras : *terminologie*, *sens*, *terme*. Les notions importantes font l'objet d'une entrée dans l'index à la fin du manuel.

Un certain nombre de passages sont détachés du texte central et présentés de la manière suivante :

---

### Terminotique

La *Terminotique* a été créée pour faire référence à l'ensemble des activités liées à la description des termes dans lesquelles intervient une application informatique. Comme on le verra au chapitre 1, l'informatique est présente à presque toutes les étapes : de la collecte de termes et d'information sur ces termes à l'encodage des descriptions.

---

Il s'agit d'explications supplémentaires sur des notions introduites dans le texte central.

## INTRODUCTION

*Toute activité – technique, manuelle ou conceptuelle – se dote de termes spécifiques, des mots empruntés au langage commun ou construits de toutes pièces. En maths aussi. Cela est nécessaire dès que l'on veut être précis, efficace, opératoire. Un cours de maths est un cours de langue ! (Entrevue de Denis Guedj accordée à L'Actualité, 15 mai 2001.)*

### **Un manuel de terminologie ou de terminographie ?**

Dans les années 1970, Alain Rey a proposé de faire le départ entre les volets appliqué et théorique de la terminologie et de les étiqueter respectivement *terminographie* et *terminologie*. La terminographie regroupe les diverses activités d'acquisition, de compilation et de gestion des termes. La terminologie se penche sur les questions fondamentales que soulève l'étude des termes et propose un cadre conceptuel pour les appréhender. Rey reproduisait ainsi la distinction entre lexicographie et lexicologie.

Bien entendu, en terminologie comme ailleurs, la théorie et la pratique sont loin d'être mutuellement exclusives. La pratique fait souvent évoluer une théorie dans une direction qu'elle n'avait pas empruntée jusque-là. Les modèles théoriques, quant à eux, permettent aux praticiens d'examiner un objet sous un éclairage particulier.



Cependant, en terminologie peut-être davantage qu'ailleurs, on éprouve plus de difficulté à scinder théorie et pratique. Jusqu'à présent, la première a surtout servi à expliquer la seconde. Autrement dit, la terminologie fournit des repères fondamentaux aux différentes activités qui relèvent de la terminographie. Il est possible de concevoir une lexicologie sans lexicographie. La terminologie, en revanche, se définit surtout par rapport à des applications relevant de la terminographie.

Ainsi, ce manuel portera principalement sur la terminographie, c'est-à-dire sur un ensemble de pratiques dont l'objet commun est de décrire des termes. Bien entendu, ces pratiques gagnent à être soutenues par des modèles théoriques adéquats, donc par la terminologie. Mais nous tiendrons pour acquis qu'une théorie valable est le résultat de l'observation et de la description d'une masse importante de données terminologiques et qu'elle a tout avantage à s'appuyer sur des descriptions concrètes de termes.

Le chapitre 1 balisera avec plus de précision les domaines respectifs de la terminographie et de la terminologie. Il montrera également que les termes apparaissent sous un jour différent en fonction du modèle théorique choisi.

## **Un manuel de terminographie ou de terminotique ?**

La terminographie accueille des applications informatiques de plus en plus nombreuses. Une information extrêmement riche désormais accessible en format électronique ainsi que différents traitements automatiques (ou pilotés par l'humain) viennent faciliter la collecte et l'analyse des termes.

Déjà, en 1991, Auger *et al.* faisaient un parallèle entre les méthodes qualifiées de *classiques* et de nouvelles techniques intégrant des outils informatiques. Sager (1990) et Cabré (1992) accordaient eux aussi une grande importance aux nouvelles technologies<sup>1</sup>. De nombreux chapitres font référence à l'avènement des corpus en format électronique ; aux mérites des outils qui arrivent à s'y frayer un chemin pour débusquer un terme ou un contexte ; et, enfin, à la richesse d'une organisation de données terminologiques réalisée sur support électronique. Le concept de « terminotique » commençait à prendre forme.

1. Signe des temps, la dernière édition du *Manuel pratique de terminologie* (Dubuc 2002) comporte un chapitre additionnel consacré à l'informatique.

---

## Terminotique

La **Terminotique** a été créé pour faire référence à l'ensemble des activités liées à la description des termes dans lesquelles intervient une application informatique. Comme on le verra au chapitre 1, l'informatique est présente à presque toutes les étapes : de la collecte de termes et d'information sur ces termes à l'encodage des descriptions.

---

Aujourd'hui, la situation a évolué considérablement. Les applications informatiques se sont diversifiées et, surtout, elles sont devenues accessibles au plus grand nombre. Actuellement, toute recherche portant sur des termes fait appel à une forme ou une autre de traitement informatique, si bien que la distinction entre *terminotique* et *terminographie* ne se justifie que dans un contexte pédagogique. Ainsi, ce manuel mettra un accent particulier sur les activités terminographiques réalisées ou réalisables à l'aide de l'ordinateur. Nous verrons au chapitre 1 de quelle manière la terminographie peut tirer profit des outils informatiques.

---

## Terminotique ou terminologie computationnelle ?

Aujourd'hui, **terminologie computationnelle** (sur le modèle de l'anglais *computational terminology*) côtoie souvent *terminotique* dans la littérature spécialisée. Toutefois, les deux expressions ne recouvrent pas tout à fait les mêmes réalités, telles que nous les envisageons. La terminologie computationnelle constitue un domaine de recherche du traitement automatique de la langue dont les acteurs – informaticiens, spécialistes des sciences de l'information, linguistes et, bien entendu, terminologues – mettent au point divers traitements automatiques appliqués aux textes spécialisés.

La terminologie computationnelle et la terminotique ont donc des objectifs différents. La première se penche sur les traitements automatiques qui auront vraisemblablement une utilité en terminographie, mais ils ne sont pas conçus exclusivement pour cette fin. La seconde veut intégrer des outils informatiques dans la confection de dictionnaires spécialisés. Cet objectif pourra l'amener à mettre au point des formes de traitements automatiques, mais cela ne constitue pas sa raison d'être principale.

Bien que leurs objectifs divergent, chaque discipline profite des avancées de l'autre. Ainsi, bien que ce manuel accorde une grande place à la terminotique, il fera souvent référence à des travaux de terminologie computationnelle.

---

## Revisiter la terminologie et la terminographie

La terminologie et la terminographie ont beaucoup changé au fil des ans et le rythme de ce changement s'est accéléré récemment. Cette situation s'explique, entre autres, par le recours systématique à de grands corpus de textes et par l'utilisation de traitements informatiques qui modifient en profondeur le travail du terminographe.

D'abord, l'exploration automatisée des corpus révèle certaines dimensions qui échappaient plus facilement à l'observation. Les outils de traitement automatique permettent de localiser les données rapidement et, surtout, de les ordonner de différentes manières. C'est un peu comme si l'utilisation d'un appareil de mesure plus perfectionné mettait en évidence ce que l'appareillage précédent occultait.

---

### Produire des descriptions plus riches

Le recensement des cooccurrents dans les dictionnaires spécialisés (*BOURSE* 1986, *COOC INTERNET* 2000) est attribuable en grande partie à la possibilité de retrouver rapidement les unités lexicales apparaissant dans l'environnement du terme. Cette recherche est désormais facilitée par le concordancier.

Par ailleurs, un dictionnaire comme le *Dictionnaire d'apprentissage du français des affaires* (*DAFA* 2000), qui rend compte de la polysémie et qui dresse des listes de dérivés morphologiques et de collocations, n'aurait jamais été envisageable sans le recours à des outils d'interrogation de corpus en format électronique.

Enfin, l'émergence du concept de « base de connaissances terminologique » (Meyer *et al.* 1992), à savoir une base terminologique dans laquelle les relations entre les termes sont explicitées de façon systématique, doit beaucoup à la possibilité d'implanter ces liens dans les bases de données actuelles et d'en garder la trace.

---

De plus, l'utilisation d'outils informatiques apporte des éclairages différents sur les termes et leur environnement linguistique. En effet, la plupart d'entre eux permettent de rompre avec la lecture linéaire qui caractérisait la recherche classique et pratiquent des coupes sélectives dans les textes. La recherche de chaînes de caractères dans un corpus électronique force le terminographe à s'attarder sur des formes identiques mais qui ne véhiculent

pas le même sens. Cette même recherche montre que la forme d'un terme subit des transformations dans un texte. Une liste de concordances révèle des régularités quant aux unités lexicales qui apparaissent dans l'environnement du terme. Les notions de polysémie, d'ambiguïté ou de variation, pour ne citer que quelques cas, prennent des dimensions qu'elles n'avaient pas auparavant, puisque c'est l'humain avec toutes ses connaissances du monde qui tranchait sans toujours le faire consciemment.

L'utilisation généralisée d'outils informatiques remet en question les modèles et méthodologies qui ont fait pourtant une unanimité durable. Ce consensus est ébranlé par une série de questions soulevées par l'examen de textes tout venant et pour lesquelles des modèles plus anciens n'apportent pas toujours les réponses. Ces questions seront au cœur du chapitre 1.

Le nouvel éclairage jeté sur les données terminologiques nous amène également à revisiter certains concepts fondamentaux de la terminologie. Le chapitre 2 définit le « terme » et dresse une liste de difficultés liées à son repérage dans le texte. Le chapitre 3 se penche sur les différents liens qui existent entre les termes. Dans ces deux chapitres, nous rappellerons des concepts décrits dans les manuels de terminologie. Toutefois, nous devons lorgner du côté d'autres disciplines (comme la lexicologie, en particulier la sémantique lexicale, et le traitement automatique de la langue) afin de clarifier des concepts sur lesquels la terminologie s'est peu attardée jusqu'à maintenant.

## **De nouvelles pratiques**

Après avoir revu les notions fondamentales de la terminologie, nous aborderons des thèmes qui relèvent plus directement de l'informatique. Le chapitre 4 présente les différentes manières d'élaborer un corpus spécialisé en format électronique et le chapitre 5 se penche sur les moyens mis en œuvre pour l'interroger. Le chapitre 6 porte sur l'extraction automatique de termes et souligne les difficultés de cette entreprise. Le chapitre 7 décrit des techniques qui vont au-delà de l'acquisition de termes et montre qu'il est maintenant possible de dégager de l'information de nature sémantique des corpus spécialisés. Le chapitre 8 aborde la question de l'organisation des données terminologiques et traite, notamment, des bases de données et des documents structurés.

**Introduction à la victimologie**

JO-ANNE WEMMERS

**Introduction aux relations  
internationales**

DIANE ÉTHIER avec

la collaboration de

MARIE-JOËLLE ZAHAR

**Lexicologie et  
sémantique lexicale**

*Notions fondamentales*

ALAIN POLGUÈRE

**Le modèle ludique**

*Le jeu, l'enfant avec déficience  
physique et l'ergothérapie*

Troisième édition

FRANCINE FERLAND

**La politique comparée**

*Fondements, enjeux  
et approches théoriques*

MAMOUDOU GAZIBO et JANE JENSON

**Pour comprendre le nationalisme  
au Québec et ailleurs**

DENIS MONIÈRE

**Préparer la relève**

*Neuf études de cas sur  
l'entreprise au Québec*

LOUISE ST-CYR et FRANCINE RICHER

**La psychocriminologie**

*Apports psychanalytiques  
et applications cliniques*

DIANNE CASONI et LOUIS BRUNET

**La radio à l'ère de la convergence**

En collaboration avec la chaîne  
culturelle de Radio-Canada

**Le régime monétaire canadien**

*Institutions, théories et politiques*

Nouvelle édition

BERNARD ÉLIE

**Savoir entreprendre**

*Douze modèles de réussite*

Études de cas

LOUIS JACQUES FILION

**Séduire par les mots**

*Pour des communications  
publiques efficaces*

JEAN DUMAS

**Le système politique américain**

Nouvelle édition

Sous la direction d'EDMOND ORBAN et

MICHEL FORTMANN

**Les temps du paysage**

Sous la direction de PHILIPPE

POULLAUEC-GONIDEC, SYLVAIN

PAQUETTE et GÉRALD DOMON

**Traité de criminologie empirique**

Troisième édition

Sous la direction de MARC LE BLANC,

MARC OUMET et DENIS SZABO

**L'univers social des adolescents**

MICHEL CLAES

**Violences au travail**

*Diagnostic et prévention*

Sous la direction de FRANÇOIS COURCY,

ANDRÉ SAVOIE et LUC BRUNET

**Les visages de la police**

*Pratiques et perceptions*

JEAN-PAUL BRODEUR

 **AGMV Marquis**  
MEMBRE DE SCARRINI MEDIA  
**Québec, Canada**  
**2004**

Extrait de la publication