

Évaluation d'un échantillon d'index

Thierry Lepage and Caroline de Schaetzen

Volume 42, Number 2, juin 1997

Lexicologie et terminologie II (1) et Traduction et post-colonialisme en Inde

Translation and Postcolonialism: India (2)

URI: <https://id.erudit.org/iderudit/003465ar>

DOI: <https://doi.org/10.7202/003465ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Lepage, T. & de Schaetzen, C. (1997). Évaluation d'un échantillon d'index. *Meta*, 42(2), 328–346. <https://doi.org/10.7202/003465ar>

Article abstract

This article describes a study of the quality of the indexes contained in a sample of specialized works for possible use in computerized extraction of specialized vocabularies.

ÉVALUATION D'UN ÉCHANTILLON D'INDEX*

THIERRY LEPAGE ET CAROLINE DE SCHAEZEN

Institut Libre Marie Haps, Centre de terminologie de Bruxelles, Bruxelles, Belgique

Résumé

Le présent article décrit une étude de la qualité des index d'un échantillon d'ouvrages spécialisés en vue d'une exploitation possible pour l'extraction automatique de vocabulaires spécialisés.

Abstract

This article describes a study of the quality of the indexes contained in a sample of specialized works for possible use in computerized extraction of specialized vocabularies.

1. POURQUOI ÉVALUER LES INDEX ?

L'index est la liste alphabétique de toutes les occurrences des entrées d'index avec leurs numéros de pages dans le livre. En 1993, le Centre de Terminologie de Bruxelles (CTB) a mené, pour le ministère belge des Technologies nouvelles et de l'Emploi de la Région wallonne de Belgique, une recherche sur l'exploitation possible des index d'ouvrages pour l'**extraction automatique de vocabulaire spécialisé**. Cette recherche a exploré les obstacles à l'utilisation d'un programme écrit par le CTB. Ce programme consulte les index des livres disponibles sur support informatique ; il se reporte aux pages dont les numéros sont cités à côté de chaque expression reprise dans cet index. Comme un être humain le ferait, ce programme lit donc les index ; comme lui, il lit également les phrases, qui, aux pages du livre mentionnées dans cet index, comportent les mots listés par **ordre alphabétique** dans l'index. Le programme recopie enfin les phrases du livre comportant ces termes, après avoir mis ces termes comme vedettes de ces «fiches» terminologiques. À la fin de chaque fiche terminologique de ce nouveau document, le programme recopie également les références bibliographiques du livre. Il travaille donc comme un terminographe qui, au fil de ses lectures d'ouvrages scientifiques ou techniques, note les termes et leur contexte. Une seule différence: c'est l'index du livre et non une intelligence du caractère terminologique (ou probablement terminologique) des mots et syntagmes qui sert de guide à ce petit programme. Pourquoi ne pas utiliser plutôt un générateur de concordances, qui dresse la liste de tous les mots d'un document et recopie lui aussi leur **contexte d'occurrence** ? L'avantage du **lecteur d'index**, qui est un outil complémentaire du générateur de concordances, est d'exploiter l'étape la plus difficile de l'indexation, la présélection des entrées d'index effectuée par l'auteur ou l'indexeur du livre. Pour les manuels didactiques de certaines collections d'ouvrages de spécialité, l'index est très complet et très articulé et il comporte en fait tous les termes utiles de l'ouvrage. Qui plus est, il ne comporte qu'eux. Comme le concordancier, le lecteur d'index permet de **paramétrer le contexte** à copier: un nombre précis de phrases ou de mots à gauche et/ou à droite de l'expression qui doit figurer dans l'index.

Ce sont les éventuels obstacles à l'utilisation intensive d'un tel «lecteur d'index» qui faisaient l'objet de la recherche menée pour la Région wallonne. Ces obstacles sont juridiques (l'exploitation dérivée d'une copie électronique d'un ouvrage est soumise au droit d'auteur), linguistiques (les index ne sont pas des glossaires exhaustifs ; les définitions du livre, pour les contextes définitoires des entrées d'index qui sont des termes, sont parfois

loin d'être exactes et exhaustives, l'intérêt des contextes des termes faisant l'objet d'une entrée dans l'index peut être limité) et pratiques (il n'existe pratiquement pas de corpora de textes spécialisés et aucun réseau international de terminologie n'a inclus à son programme d'activités la constitution progressive de tels corpora). Dans le cadre de cette recherche, la qualité d'un échantillon d'index spécialisé a été évaluée. Le présent article rend compte des résultats de cette évaluation.

Un index est-il assimilable à un glossaire, trié et assez exhaustif, du vocabulaire spécialisé d'un ouvrage ? Quels sont les types de documents qui se prêtent le mieux au dépouillement automatisé de leur index ? Tel est le premier but de cette étude.

Indépendamment de son utilité pour la terminographie assistée, un jugement sur les index n'est, en troisième lieu, pas inintéressant pour les auteurs des livres de spécialités. Les **commandes d'indexation** offertes par tous les logiciels de traitement de texte permettent en effet de créer rapidement des index de livres puis d'améliorer le contenu et la présentation de ces index. Le traitement de texte permet dès lors à tous les auteurs de livres spécialisés de rendre au lecteur un service supplémentaire : un index de leur propre vocabulaire de spécialistes. Le savoir-faire qu'exige l'indexation doit cependant être assimilé en même temps que la connaissance des commandes qui l'automatisent. Or l'index est un produit terminographique, même s'il diffère à maints égards des glossaires, vocabulaires, dictionnaires. Nous émettons donc en conclusion quelques préceptes linguistiques à l'endroit des indexeurs professionnels et occasionnels.

2. OUVRAGES ANALYSÉS

Le nombre d'ouvrages analysés est peu élevé. Cette étude devrait donc être refaite.

2.1. Critères de choix

Les ouvrages ayant servi à notre échantillonnage d'index sont essentiellement :

- des ouvrages didactiques, notamment des ouvrages de base dans une discipline scientifique et/ou technique, par exemple destinés à des étudiants de première année universitaire, ainsi que des ouvrages publiés dans des collections de vulgarisation de haut niveau. On suppose en effet que ce type d'ouvrages est bien structuré ;
- des ouvrages de collections faisant l'objet d'une politique d'édition bien précise et pour lesquels les conclusions sur définitions et index sont donc les plus susceptibles d'être utiles ;
- des livres de linguistique, car nous supposons que cordonnier serait bien chaussé... ;
- des ouvrages et manuels dont le feuilletage a permis de constater la présence de termes définis.

2.2. Liste des ouvrages étudiés

BELISLE, C. et G. JOUANNADE (1988) : *Communication Visuelle*, Paris, Éditions d'Organisation.

Collectif (1977) : *Troubles du langage : diagnostic et rééducation*, Bruxelles, Pierre Mardaga.

COSTERMANS, J. (1980) : *Psychologie du langage*, Bruxelles, Pierre Mardaga.

D'HONDT, Th. et M. JADOUL (1988) : *Een eerste stap in de Programmeerwereld (Deel II)*, Schoonhoven, Academic service.

DE BROCK, E. O. (1989) : *De Grondslagen van de Semantische Databases*, Schoonhoven, Academic service.

DE GROOFF, D. (1991) : *Media Lexicon*, Leuven, Lannoo - Mediacentrum.

De Gruyter (Ed.) (1989) : *Dictionaries — An International Encyclopedia of Lexicography*, Berlin/New York, De Gruyter.

- DE SCHAETZEN, C. (1990): *Word 4 et 5 avancé. Les macro-commandes*, Alleur, Marabout.
- DUMAS DE RAULY, Th. (1987): *Choisir et utiliser les supports visuels et audiovisuels*, Paris, Éditions d'Organisation.
- FISHMAN, J.-A. (1972): *Advances in the Sociology of Language*, Paris/The Hague, Mouton.
- GARDNER, H. (1980): *Gribouillages et dessins d'enfants*, Bruxelles, Pierre Mardaga.
- GIBSON, E.-J. et H. LEVIN (1976): *Psychology of Reading*, Cambridge, Massachusetts, MIT Press.
- HAMERS, J.-F. et M. BLANC (1983): *Bilingualité et Bilinguisme*, Bruxelles, Pierre Mardaga.
- IBM (1991): *Manuel de référence DOS 5.0*, IBM.
- KORNELIS, I. J. et R. A. MOLLEMA (1989): *Controle van de informatieverwerking*, Alphen-aan-den-Rijn/Brussel, Samson.
- LEROT, J. (1983): *Abrégé de linguistique générale*, Louvain-la-Neuve, Cabay.
- MOREAU, M.-L. et M. RICHELLE (1981): *Acquisition du langage*, Bruxelles, Pierre Mardaga.
- MYKLEBUST, H. R. (1964): *Psychology of Deafness*, New York/London, Grune and Stratton.
- OLDFIELD, R. C. et J. C. MARSHALL (1968): *Language*, Harmondsworth, Penguin.
- PAUL, B. (1988): *Efficiënter programmeren in Basic*, Antwerpen, Maklu.
- PIALOUX, P., VALTAT FREYSS, M. G. et F. LEGENT (1975): *Précis d'orthophonie*, Paris, Masson et cie.
- REY-DEBOVE, J. (1971): *Étude linguistique et sémantique des dictionnaires français contemporains*, Amsterdam, Mouton.
- SHARPLES, M., HOGG, D., HUTCHISON, Ch., TORRANCE, S. et D. YOUNG (1989): *Computers and Thought*, Cambridge, Massachusetts, MIT Press.
- VAN EIJCK, J. et E. THIJSEN (1989): *Logica voor Alfa's en informatici*, Academic service.
- ZGUSTA, L. (1971): *Manual of Lexicography*, The Hague/Paris, Mouton.

3. GRILLE D'ÉVALUATION DES INDEX

Nous avons conçu une grille destinée à vérifier l'exhaustivité des index proposés et, dans la négative, à sérier leurs carences. L'attention s'est également portée sur l'environnement, dans le corps du texte, des termes spécialisés de l'index (définition, exemple, collocations, renseignements complétant les définitions).

Nous avons profité de l'évaluation des index pour la terminographie pour étudier toutes les caractéristiques des index choisis. La raison en est, une fois encore, qu'aucune étude de ce type n'a été menée, ou très peu ; or les index contribuent à la diffusion du contenu des ouvrages. La figure 1 donne la grille :

1.	Auteur(s) :
2.	Titre :
	
	
3.	Collection :
4.	Éditeur :
5.	Année d'édition :
6.	Ville d'édition :
7.	Nombre total de pages :
8.	Présence d'un index ? (O/N) :

9. Nombre de pages d'index :
10. Nombre de pages d'annexes ou d'appendices autres que l'index :
11. Domaine abordé par le livre :
12. Index couvrant un domaine (*biffez ce qui ne convient pas, svp*) :
 trop large
 d'extension correcte
 trop étroit
13. Index omettant des entrées importantes, pourtant contenues dans le texte (*O/N*) :
14. Nombre d'entrées de l'index :
15. Nombre d'entrées d'index par rapport au nombre de pages de texte (%) :
16. Entrées secondaires ; tertiaires (*O/N*) :
17. Annexes séparées pour les (*biffez ce qui ne convient pas, svp*) :
 combinaisons de touches (*)
 termes de matériel (*)
 messages-système (*)
 index lexical
 liste de commandes (*)
 liste de fonctions (*)
 liste onomastique
 glossaire / lexique explicatif
 noms de menus (*)
 noms de fichiers (*)
 bibliographie
 messages d'erreurs (*)
18. Entrées d'index relatives à des (*biffez ce qui ne convient pas svp*) :
 combinaisons de touches (*)
 termes relatifs au matériel (*)
 messages-système (*)
 noms de commandes (*)
 noms de fonctions (*)
 entrées onomastiques
 entrées lexicales
 noms de menus (*)
 noms de de fichiers (*)
 entrées d'informatique théorique (*)
 noms de logiciels (*)
 opérateurs booléens (*)
19. Classification alphabétique (*O/N*) :
20. Classification systématique (*O/N*) :
21. Indexation en nid (*O/N*) :
22. Entrée de renvoi dans l'index pour tous les termes significatifs des entrées syntagmatiques (*O/N*) :
23. Entrée de renvoi pour les synonymes (*O/N*) :
24. Entrée de renvoi pour les cohyponymes (*O/N*) :
25. Entrée de renvoi pour les hypéronymes (*O/N*) :
26. Entrées secondaires, tertiaires (*O/N*) :
27. Répétition du premier terme d'un syntagme pour toutes les entrées d'index à même terme initial (*O/N*) :
28. Références de pages majoritairement (*biffez ce qui ne convient pas, svp*) :
 uniques
 multiples
29. Contenu le plus fréquent de la première page citée (*biffez ce qui ne convient pas svp*) :
 définition de l'entrée
 contexte explicatif pour l'entrée
 simple de l'entrée
 simple attestation de l'entrée
30. Différenciation typographique (mise en gras, italique, notamment) des types d'entrées (*O/N*) :
31. Différenciation typographique du numéro de page contenant la définition de l'entrée d'index (*O/N*) :
32. Différenciation typographique, au sein du corps du texte, de l'entrée d'index (*O/N*) :
33. Différenciation, au sein du corps du texte, de la définition des entrées d'index (*O/N*) :

Figure 1

Les astérisques (*) de la grille indiquent que les options à cocher qu'elles suivent ne sont éventuellement applicables que pour les ouvrages portant sur l'informatique. Dans ces ouvrages de l'échantillon, nous avons voulu voir si l'index était général, donc s'il comportait tous les types de noms-termes désignés par l'expression suivie de l'astérisque, ou s'il était spécial, en l'occurrence purement lexical.

4. ANALYSE DES INDEX

Voici les résultats de cette analyse. Il va de soi que l'étroitesse de l'échantillon interdit l'attribution d'une valeur statistique aux chiffres donnés dans les tableaux ci-après.

4.1. Nombre et taille des index

Le nombre de pages des ouvrages analysés varie de 93 à 3 355. Première question, à laquelle répond le tableau 1 : les ouvrages possèdent-ils un index ?

RÉPONSE	NOMBRE D'OUVRAGES
oui	22
non	3
TOTAL	25

Tableau 1 :
Présence d'un index dans l'ouvrage

Parmi les trois ouvrages dépourvus d'index figure celui de L. Zgusta, dont le sujet est la lexicographie. Souvent, c'est l'éditeur qui décide de l'existence des index : tous les livres publiés par les *Éditions d'Organisation* sont sans index.

Le nombre moyen d'entrées d'index par nombre de pages du livre (annexes comprises) indique la taille relative de l'index. L'index est la portion congrue des ouvrages lorsque le nombre de pages total est compté, comme c'est souvent le cas. De plus, le pointage systématique de toutes les entrées d'index est un gros travail. L'auteur est souvent pressé par l'éditeur, à ce stade, en raison des délais de publication. Ce chiffre fluctue dès lors fortement en fonction des auteurs, des collections et des éditeurs.

Le nombre de pages d'index varie de 2 à 66 ; le nombre moyen de pages d'index n'est pas significatif car il dépend de la typographie et de la mise en pages (taille de l'interlignage, nombre de colonnes par page, police et corps, etc., qui diffèrent de ceux du corps de l'ouvrage). C'est que ces paramètres constituent autant de facteurs sur lesquels peut jouer le maquettiste pour respecter le calibrage prévu. Le nombre d'entrées d'index varie de 89 à 9 504, avec une moyenne de 1 016, ce qui n'est pas négligeable. Le nombre d'entrées d'index par page de texte (à l'exception des annexes bibliographiques, onomastiques, glossaires, lexiques, et autres appendices de même type) est très variable : 0,25 à 9,95, selon les auteurs. Un consensus semble établir la moyenne aux environs de **1,6 entrée par page de texte**. À première vue, les index existants ne sont donc pas trop maigres et pourraient être exploités avec profit par les terminographes... Les éditeurs qui font appel à des professionnels de l'indexation, notamment Microsoft et IBM, publient des ouvrages dont l'index est nettement plus élevé. La raison est que le lecteur doit absolument pouvoir

retrouver toutes les commandes, fonctions et combinaisons de touches du logiciel sans connaître leur libellé car le manuel d'utilisation est aussi un tutoriel.

Le tableau 2 répond à une seconde question : s'il n'est limité qu'à quelques sous-domaines, l'index reflète-t-il tout le domaine couvert par l'ouvrage, est-il limité à quelques sous-domaines ou, au contraire, inclut-il des termes ne relevant pas directement du domaine abordé par l'ouvrage ?

RÉPONSE	NOMBRE D'OUVRAGES
trop large	0
correcte	22
trop étroite	3
TOTAL	26

Tableau 2 :
Couverture par l'index du sujet traité dans l'ouvrage

Lorsqu'un index a été créé, le tableau 2 montre qu'il recouvre dans 95 % des cas l'ensemble du domaine ou des domaines abordés par l'auteur dans le corps de l'ouvrage, ce que confirme le tableau 3 que voici :

RÉPONSE	NOMBRE D'OUVRAGES
oui	2
non	21
sans index	3
TOTAL	26

Tableau 3 :
Omission dans l'index d'entrées importantes pour le sujet du livre

Y a-t-il des termes pertinents ou vitaux relevant du sujet, éventuellement définis ou explicités dans le corps de l'ouvrage, qui ne figurent pas dans l'index ? Le tableau 3 semble indiquer que non. Si cela se vérifiait par d'autres comptages, les index gagneraient donc à faire l'objet d'une exploitation terminographique.

4.2. Types d'entrées

Au lieu d'un index général, les auteurs des ouvrages de notre échantillon ont-ils préféré les index spéciaux (index, glossaire/lexique, liste de commandes et fonctions, liste onomastique, bibliographie) ?

RÉPONSE	NOMBRE D'OUVRAGES
combinaisons de touches	0
termes relatifs à du matériel	0
messages-système	0
index lexical	15
liste de commandes	2
liste de fonctions	1
liste onomastique	10
glossaire / lexique explicatif	1
noms de menus	0
noms de fichiers	15
bibliographie	0
S.R.	3

Tableau 4 :
Index séparés

Le tableau 4 montre que les types d'index séparés les plus courants sont les index onomastiques et les bibliographies.

RÉPONSE	NOMBRE D'OUVRAGES
combinaisons de touches	2
termes relatifs à du matériel	4
messages-système	0
index lexical	3
liste de commandes	3
liste de fonctions	2
liste onomastique	4
glossaire / lexique explicatif	2
noms de menus	5
noms de fichiers	5
bibliographie	2
opérateurs booléens	3
S.R.	3

Tableau 5 :
Présence d'entrées d'index pour des éléments non lexicaux

Le but de cette question est de connaître la nature des entrées présentes dans l'index. Le tableau 5 confirme que les index généraux dominent, ce qui ne fait pas réellement l'affaire de la terminographie...

Les entrées lexicales de nos index sont nominales plutôt que verbales, adjectivables, adverbiales ou d'autres catégories syntaxiques. Elles reflètent en cela les caractéristiques distributives du vocabulaire spécialisé.

Le sigle ou l'acronyme figure-t-il en entrée (assorti du numéro de page correspondant et éventuellement de sa forme complète), avec ou sans son équivalent en toutes lettres ? Ou bien le lecteur est-il renvoyé vers la forme complète du sigle ? L'acronyme n'apparaît pas souvent dans les index de notre échantillon ; lorsqu'il y figure, il constitue le plus souvent une entrée d'index individuelle et n'apparaît que rarement en compagnie de son syntagme complet. La figure 2 montre les présentations adoptées pour les sigles :

DOS 100
DOS (Disk Operating System) 100
DOS —> Disk Operating System
Disk Operating System, Dos 100

Figure 2

Ces différences de traitement sont préjudiciables à l'exploitation automatique des index mais sans doute aussi à leur cohérence.

4.3. Classement des entrées

Quel mode de classement l'auteur adopte-t-il pour les entrées ?

RÉPONSE	NOMBRE D'OUVRAGES
oui	23
non	0
s.r.	3
TOTAL	26

Tableau 6 :
Classement alphabétique des entrées d'index

Selon le tableau 6, le classement alphabétique domine en raison de sa commodité. Ce classement se prête à l'utilisation des index pour l'établissement de vocabulaires, toujours classés par ordre alphabétique. Ajoutons que les entrées d'index dont le caractère initial n'est pas une lettre (* £\$ @ #) ou est numérique peuvent être classées en tête ou en fin d'index. Une faible majorité des auteurs préfèrent les faire figurer en tête d'index. Le **classement alphabétique discontinu** ne tient pas compte de la ponctuation. C'est le cas

de la suite d'entrées suivante : *copie d'un document, copie d'une macro, copie sans formatage, copie, macro de-, copie, procédure de-*. Le **classement alphabétique discontinu** en tient compte, comme dans la série : *copie d'un document, copie d'une macro, copie, macro de-, copie, procédure de-, copie sans formatage*. L'édition admet les deux modes de classement.

RÉPONSE	NOMBRE D'OUVRAGES
oui	0
non	23
s.r.	3
TOTAL	26

Tableau 7 :
Classement systématique des entrées d'index

Le classement systématique est peu prisé des auteurs et indexeurs. Il se prête mal à l'utilisation des index pour l'établissement de vocabulaires, ces derniers étant également toujours classés par ordre alphabétique. Ajoutons que la majorité des index incluant des entrées non alphabétiques font précéder le *a* des caractères spéciaux non alphabétiques, et que près d'un quart d'entre eux mettraient ces caractères derrière le *z*.

Dans la séquence de leurs entrées d'index, certains auteurs distinguent les deux modes de classement, par exemple le classement alphabétique venant en premier lieu puis les termes indexés en nid, comme en *a*), à la figure 3. Par contre, d'autres auteurs mêlent indistinctement les entrées obtenues à partir de ces deux types de classements différents, ce que montre *b*), et qui paraît moins logique.

<i>a)</i> copie d'un document copie d'une macro copie sans formatage copie, macro de- copie, procédure de- <i>b)</i> copie d'un document copie d'une macro copie, macro de- copie, procédure de- copie sans formatage
--

Figure 3

Il arrive, dans les index de manuels d'informatique (2 cas sur 7, dans notre échantillon), que des mots anglais dits «réservés» parce que obligatoires et figurant dans une commande, soient traduits en français. Certains auteurs insèrent alors séparément ces deux entrées

dans leur index en les faisant référer directement à un numéro de page précis, ou à une autre entrée apparentée ou synonymique (figure 4) :

boucle WHILE —> boucle indéfinie.
boucle TANT QUE —> boucle indéfinie.

Figure 4

4.4. Renvois

Le tableau 8 montre la structuration des renvois :

RÉPONSE	NOMBRE D'OUVRAGES
oui	17
non	6
s.r.	3
TOTAL	26

Tableau 8 :
Indexation en nid

Si une indexation en nid a été prévue, et il semble que ce soit le cas pour notre échantillon, le terminographe ne devra pas répéter ce travail d'inversion et de renvois.

Lorsque l'index n'est pas en nid, c'est souvent le calibrage des publications qui l'explique : dans un index comportant une majorité de syntagmes, l'ajout d'entrées en nid allonge trop l'index (jusqu'à un tiers d'entrées en plus).

RÉPONSE	NOMBRE D'OUVRAGES
oui	2
non	21
s.r.	3
TOTAL	26

Tableau 9 :
Renvois significatifs pour tous les constituants principaux des syntagmes

Cette indexation en nid est-elle complète ? Il est extrêmement rare (2 sur 17 ouvrages comportant des entrées d'index en nid, d'après le tableau 9) que les auteurs fassent une

entrée d'index pour chaque composante d'un syntagme. Un contrôle des renvois inverses s'imposera donc aux terminographes utilisant notre programme.

RÉPONSE	NOMBRE D'OUVRAGES
oui	4
non	19
s.r.	3
TOTAL	26

Tableau 10 :
Renvois à des cohyponymes

Le tableau 10 fait apparaître que quelques rares auteurs renvoient vers des notions apparentées selon un lien autre que la cohyponymie. Le tableau 11 révèle que les renvois à des hyperonymes ne sont pas nombreux non plus :

RÉPONSE	NOMBRE D'OUVRAGES
oui	5
non	18
s.r.	3
TOTAL	26

Tableau 11 :
Renvois à des hyperonymes

Il n'est pas tellement étonnant que la plupart des index ne comportent pas de renvois aux termes logiquement proches : ces renvois rallongent la liste des entrées, sans être vraiment indispensables... Les index examinés ne comportent que rarement des renvois de synonymie. Ce n'est pas uniquement dommage pour les applications terminographiques : le lecteur consultant l'index risque de penser à un synonyme plutôt qu'à la dénomination figurant en entrée et il ne retrouvera donc pas les pages du livre traitant de ce concept.

Dans les manuels d'informatique (2 manuels sur 7 dans notre échantillon), des mots anglais réservés parce que obligatoires dans une commande sont traduits en français. Certains auteurs insèrent alors séparément ces entrées dans leur index en les faisant référer directement à un numéro de page précis, ou bien en font une entrée de renvoi : *boucle WHILE -> boucle indéfinie*.

4.5. Entrées secondaires

L'index comporte-t-il plusieurs niveaux hiérarchiques ?

RÉPONSE	NOMBRE D'OUVRAGES
oui	12
non	11
s.r.	3
TOTAL	26

Tableau 12 :
Entrées secondaires et tertiaires

Selon les chiffres du tableau 12, la moitié des livres n'ont pas de sous-entrées et de sous-sous-entrées. Ce sont sans doute surtout des impératifs d'espace qui expliquent l'absence de sous-entrées.

RÉPONSE	NOMBRE D'OUVRAGES
oui	9
non	14
s.r.	3
TOTAL	26

Tableau 13 :
Répétition du terme générique dans toutes les entrées secondaires

Le tableau 13 révèle que tous les auteurs (ou indexeurs) ne répètent pas l'élément initial d'un syntagme faisant l'objet d'entrées secondaires, comme dans l'extrait de la figure 5 :

Formatage d'un texte formatage de caractères formatage de notes formatages de paragraphes
--

Figure 5

Certains remplacent cet élément par des espaces ou un tiret cadratin, comme dans l'exemple de la figure 6 :

Formatage d'un texte
 – de caractères
 – de notes
 – de paragraphes

Figure 6

Par souci de concision, une légère majorité d'auteurs ou d'indexeurs préfèrent ne pas répéter le premier terme d'un syntagme commun à plusieurs entrées consécutives. Un nombre assez important d'entre eux recourent encore à cet affichage redondant et peu esthétique. Tous les indexeurs ne semblent pas avoir saisi le caractère significatif de l'imbrication des retraits, assortie d'un symbole comme le tiret ou le tilde.

4.6. Informations dans le livre même

L'auteur cite-t-il uniquement la page où le terme repris en entrée est défini ou expliqué en détail, ou bien toutes les pages où apparaît ce terme ? Nous avons posé cette question parce que nous espérions que les numéros de pages uniques renverraient à une définition du concept désigné par l'entrée d'index, ce qui faciliterait la confection automatique de fiches définitoires.

RÉPONSE	NOMBRE D'OUVRAGES
oui	11
non	12
s.r.	3
TOTAL	26

Tableau 14 :

Renvoi depuis l'entrée d'index à plusieurs numéros de pages

Les références de pages sont à parts quasi égales multiples ou uniques, selon le tableau 14.

RÉPONSE	NOMBRE D'OUVRAGES
définition de l'entrée	3
contexte explicatif de la notion	19
désignée par l'entrée	3
contexte attestatoire (simple mention de l'entrée)	1
s.r.	
TOTAL	

Tableau 15 :

Contenu auquel renvoie le plus souvent la première page citée dans les index

Si toutes les entrées d'index étaient définies dans l'ouvrage et si la première page à laquelle figure une occurrence d'une entrée était celle qui définit ou explique en détail le concept recouvert par l'entrée, l'identification des définitions par le programme serait aisée.

Bien que les références de pages soient, à parts quasiment égales, multiples ou uniques, la première page dont le numéro est mentionné à droite d'une entrée d'index ne contient que rarement une définition du terme indexé. Elle contient le plus souvent un contexte explicatif ou se borne même à mentionner l'entrée, comme le font les contextes attestatifs. Le nombre de termes définis correctement et celui des termes mentionnés sans autre information atteignent des pourcentages identiques. Comme les concordanciers, notre programme récolte donc autant d'exemples et d'attestations que de définitions. L'utilisateur qui ne voudrait relever que les définitions peut quant à lui supprimer les numéros de pages renvoyant à des contextes, avant de lancer notre programme ou de supprimer les fiches contextuelles à l'issue du déroulement du programme.

4.7. Indications typographiques

Les questions relatives à la typographie sont importantes car le programme du CTB peut détecter ces typographies lorsqu'il dispose de la version mise en forme du manuscrit. En effet, comme il est écrit dans le langage de macro-commandes d'un logiciel de traitement de texte, il repère ces mises en évidence typographiques en vérifiant la présence du gras et de l'italique lors de sa recherche des occurrences de chaque entrée d'index.

Certains auteurs différencient en premier lieu la typographie des différents types d'entrées de leur index. C'est ainsi que des auteurs de manuels d'informatique mettent en italique les entrées faisant l'objet d'un renvoi, en majuscules les entrées dont le concept est une commande de logiciel, en gras les entrées dont le concept est une fonction de logiciel, etc.

RÉPONSE	NOMBRE D'OUVRAGES
oui	7
non	16
s.r.	3
TOTAL	26

Tableau 16 :

Typographie différente d'après le type de concept désigné par les entrées d'index

La différenciation typographique des entrées, renvois, commandes, etc., d'index n'est utilisée que dans 7 des index examinés (tableau 16). Dans les index généraux, donc hétérogènes, qui constituent la majeure partie de notre échantillon, les termes ne sont donc malheureusement pas distingués typographiquement des autres types d'entrées ; si cette tendance était générale, le programme ne pourrait donc pas sélectionner uniquement le vocabulaire sur base d'index faits par les auteurs.

Si la définition du terme constituant l'entrée d'index ne se trouve pas à la première page figurant à côté de cette entrée, une typographie différente peut distinguer cette référence de page des autres : *hierarchical database* 123, 125-6, **129**, 134.

RÉPONSE	NOMBRE D'OUVRAGES
oui	1
non	22
s.r.	3
TOTAL	26

Tableau 17 :

Typographie du numéro de la page à laquelle l'entrée est définie

Le plus souvent, indique le tableau 17, la **référence de page**, c'est-à-dire le numéro de page renvoyant à une véritable définition de l'entrée d'index ou même à un contexte explicatif, n'est pas dans une typographie qui le distingue des autres numéros ou références de pages, plus accessoires. Ici encore, mettre manuellement en exergue le numéro de page comportant la définition est plus lent qu'éliminer, à l'issue du déroulement du programme du CTB, les fiches contextuelles excédentaires générées par ce programme.

Dans le texte même du livre, les termes qui sont des entrées d'index sont-ils revêtus d'attributs typographiques particuliers ?

RÉPONSE	NOMBRE D'OUVRAGES
oui	0
non	23
s.r.	3
TOTAL	26

Tableau 18 :

Typographie des occurrences des entrées d'index, au sein du corps du livre

Aucun auteur ou indexeur ne confère aux termes-entrées une typographie différente de celle du reste de la page, au sein de l'ouvrage. On déduira donc du tableau 18 que le programme du CTB ne peut repérer les termes techniques repris en vérifiant la typographie, donc en faisant l'économie de la consultation de l'index.

5. EXPLOITATION DES INDEX

Voici les résultats pour notre programme de ce sondage dans les index d'ouvrages :

- le nombre d'ouvrages indexés est élevé et l'informatisation des auteurs devrait encore accroître ce nombre ;
- pour que le nombre des entrées d'index s'accroisse, les éditeurs devraient leur faire de la place. Le corps et l'interlignage des index devraient dès lors être réduits dans

plus d'ouvrages, ce qui permettrait de grossir l'index tout en respectant le calibrage du livre. Quant aux auteurs, ils devraient consacrer plus de temps au pointage des entrées d'index. L'éditeur qui ne confie pas l'indexation à des indexeurs professionnels devrait donner à l'auteur des indications sur le nombre moyen souhaité pour les pages ou les entrées d'index ;

- les indexeurs professionnels devraient entrer en contact avec les milieux de la terminologie. Non seulement leur travail facilite la compilation automatique de fiches terminologiques de travail termes, mais leur activité n'est pas si différente de l'établissement de la micro-structure d'un vocabulaire.

6. RECOMMANDATIONS

Le nombre d'ouvrages indexés est heureusement élevé, et l'informatisation des auteurs accroîtra encore leur nombre. Pour une plus grande maniabilité et clarté des index, les auteurs et les éditeurs qui leur donnent des directives devraient être attentifs aux points ci-après :

- le nombre des entrées d'index devrait être accru. Le corps et l'interlignage des index devraient être réduits, pour grossir l'index tout en respectant le calibre prévu pour le livre. Quant aux auteurs, ils devraient consacrer plus de temps au pointage des entrées dans le corps de leur manuscrit. L'éditeur qui ne fait pas appel aux services d'un indexeur devrait donner des indications écrites à ses auteurs sur le nombre moyen de pages ou d'entrées d'index utiles ;
- les index à niveau multiple et les index spéciaux sont plus cohérents, parce que les entrées non lexicales en sont absentes. Ils devraient être plus fréquents ;
- la mise en évidence typographique du numéro de la page comportant la définition ou le contexte définitoire d'une entrée est pratique pour le lecteur ;
- les avantages du classement systématique pour certaines applications littéraires ou linguistiques devraient être gardés à l'esprit ;
- l'emplacement dans l'index des entrées non alphabétiques ou numériques doit être pesé ;
- les entrées hyponymiques et les renvois sont incontournables ;
- le mélange entre entrées d'index disposées en nid et syntagmes à même terme initial est une erreur, susceptible de se produire lors d'un tri par traitement de texte ;
- la différenciation typographique des différents types d'entrée d'index et une typographie particulière pour le numéro de page renvoyant à la définition accélèrent la consultation des index ;
- le traitement de texte offre une fonction d'indexation et de gestion des index. Les auteurs la maîtrisent mal, lorsqu'ils en connaissent l'existence ;
- une «nidification» de l'index peut être opérée semi-automatiquement, donc très rapidement avec l'assistance d'un programme de quelques lignes ;
- on peut générer des index multiples, donc spéciaux, grâce à la fonction de génération de tables du traitement de texte ;
- il est possible de dissocier entrée d'index et texte du corps du texte : lorsque l'auteur ou l'indexeur désire pointer un mot ou une expression dans l'ouvrage pour lui donner le statut d'entrée d'index, il suffit de sélectionner la première occurrence de cette chaîne de caractères et de lui donner le statut d'entrées par la commande requise du traitement de texte concerné. En WORD 6, le logiciel de traitement de texte copie alors cette entrée, mettant alors en texte caché le second exemplaire à côté de l'occurrence pointée, dans le corps du texte même. L'utilisateur peut modifier à son gré cette entrée d'index. Il n'y a donc aucune raison pour que les index créés avec

un traitement de texte ne soient pas lemmatisés, n'aient pas de sous-entrée ou de renvoi synonymique, etc. ;

- le traitement de texte peut, dans ses versions récentes, générer un index unique pour plusieurs documents ;
- il peut également faire un index sur base d'une liste préalablement dressée par l'auteur ou l'éditeur.

On en conclura par l'intérêt des recyclages au traitement de texte orienté pour les applications éditoriales.

Note

- * Cet article est issu d'une communication présentée par l'auteur aux IV^{es} Journées scientifiques du réseau «Lexicologie, terminologie, traduction» de l'AUELF-UREF (Lyon, France, 28, 29, 30 septembre 1995).

RÉFÉRENCES

- BERNIER, B. (1979) : *Guide de présentation d'un travail de recherche*, Québec, Les Presses de l'Université du Québec.
- BOUVAIST, J.-M./UFR Communication (1986) : *Pratiques et métiers de l'édition*, Paris, Promodis.
- CAYER, M. et L. LEBEL-HAROU (1983) : *Guide de présentation des cahiers de Néologie en marche*, Québec, Éditeur officiel du Québec.
- Centre de terminologie de Bruxelles (1992) : *Dictionary of Office Automation*, Amsterdam, Elsevier.
- Cercle de la librairie (1993) : *Répertoire international des éditeurs et diffuseurs de langue française*, Paris, Cercle de la librairie.
- CHOUKEA, Y. (1988) : «Looking for Needles in a Haystack», *Proceedings of the RIAO Conference on User-oriented Contextual Based Text and Image Handling*, Cambridge (Mass.).
- CHURCH, K., GALE, W., HANKS, P. et D. HINDLE (1991) : «Using Statistics in Lexical Analysis», U. Zernik (Ed.), *Lexical Acquisition: Using On-line Resources to Build a Lexicon*, Lawrence Erlbaum.
- DAOUST, F. (1990) : *SATO. Système de base d'analyse de textes*, Montréal, Université du Québec à Montréal (non publié).
- DATTA, J. (1991) : «Full-text Databases as a terminological Support for Translation», *Terminology Work in Subject Field. Third INFOTERM Symposium Proceedings*, Vienna, INFOTERM.
- DE SCHAEZTEN, C. (1989) : *Guide complet de WORD 4*, Paris, Eyrolles.
- DE SCHAEZTEN, C. (1990) : *Guide complet de WORD 5 sur PC et PS*, Paris, Eyrolles.
- DE SCHAEZTEN, C. (1990) : *WORD 5 et 5 avancé. Les macro-commandes*, Alleur, Marabout.
- DE SCHAEZTEN, C. (1991) : «Checklist for the Assessment of Technical Dictionaries», *Le linguiste*, XXXVIII (3/4), Bruxelles, Institut Libre Marie Haps.
- DE SCHAEZTEN, C. (1992) : *La terminologie. Un apport possible à la réduction du nombre d'échecs en première candidature à l'Université. Rapport de recherche au Ministère de l'Éducation, de la Recherche et de la Formation de la Communauté Française*, Bruxelles, Centre de terminologie de Bruxelles, Institut Libre Marie Haps (non publié).
- DE SCHAEZTEN, C. (1992) : *WORD 2 pour WINDOWS. Guide complet*, Paris, Eyrolles.
- DEVILLERS, CH. SEM. (1990) : *Survey of Technical Standards for Lexical Resources, DOC-4*, Stuttgart, Université de Stuttgart (non publié).
- DG XIII (1993) : *XII Magazine*, Bruxelles, Commission des Communautés européennes, Direction générale XIII.
- DOUTRELEPONT, C. (1992) : «Quelques logiciels utiles aux traducteurs ou aux terminologues. Les hyper-textes et l'analyse des contextes», A. Clas et H. Safar (dir.), *L'environnement traductionnel. La station de travail du traducteur de l'an 2001*, Actes du colloque de Mons, Sillery (Québec), Presses de l'Université du Québec et AUELF-UREF.
- FELBER, H. (1984) : *Terminology Manual*, Paris, INFOTERM.
- FONTENELLE, T. (1992) : «Term Acquisition from a Corpus or from a Dictionary, a Comparison», H. Tomola, K. Varantola, T. Salmi-Tolonen et J. S. Chopp (Eds), *Euralex' 92 Proceedings I-II*, Tampere, Studia translologica.
- GUER, Y. L. (1991) : *EUROTRA-7. Feasibility and Project definition Study on the Reusability of Lexical and Terminological Resources in Computerized Applications. Intermediate Report*, (non publié).
- HOCKEY, S. (1988) : «Creating and Using Large Text Databases for Scholarly research in the Humanities. Some Practical Issues», Gignoni et Peters (Eds), *Computational Lexicology and Lexicography*, Pisa, L. Giardini Editori e Stampatori n° 1, coll. «Linguistica computazionale», vol. VI.
- IBM (1993) : *Education programma*, Bruxelles, IBM Education Center.

- ISO (1982) : *Information Transfert. Standards Handbook 1*, Vienne, ISO.
- JOHNS, T. (1988) : «Implications et applications des logiciels de concordance dans la salle de classe», *Les langues modernes. Le point sur l'EAO*, n° 5, Paris, APLV.
- KEMBLE, I. R. (1991) : «Lexicography», W. Brierly et I. R. Kemble (Eds), *Computers as a Tool in Language Teaching*, Chichester, Ellis Horwood.
- KEY SUN CHOI YOUNG, S. HAN (1992) : «Syntactic Analysis Based Automatic Indexing for Korean Texts», *International Conference on Terminology, Standardization and Technology Transfert*, Pékin, Science press.
- KUKULSKA-HULME, A. (1993) : «From Words, through Rhetoric, to Knowledge Needs», K.-D. Schmitz (Ed.), *TKE' 83, Terminology and Knowledge Engineering*, Frankfort, Indeks Verlag.
- LABARRA, A. (1990) : *Histoire du livre*, Paris, PUF.
- LEROY, B. (1991) : *Gérer la fabrication d'un journal*, Éditions du Centre de formation et de perfectionnement des journalistes, Paris, Diffusion La Découverte.
- MCNAUGHT, J., NKWENTI-AZEH, B., MARTIN, W. et E. TEN PAS (1991) : *EUROTRA-7 Study DOC-11. Feasibility of Standards for Terminological Description of Lexical Items*, Final version (non publié).
- Ministère des Approvisionnements et services du Canada (1983) : *Guide du rédacteur de l'administration fédérale*, Ottawa, Secrétariat d'État du Canada.
- MUSTAFA-ELHADI, W. (1990) : «The Contribution of Terminology to the Theoretical Conception of Classification Languages and Documentation Indexing», International Society for Knowledge Organization (Ed.), *Tools for Knowledge Organization and Human Interface. Proceedings 1st International ISKO-Conference, Darmstadt, 14-17 August 1990*, vol. 1, Francfort, Indeks Verlag.
- NEGRE, Cl. et A. DU PONTAVICE (1990) : «Micro-édition et mise en page», *Le secrétariat de rédaction. De la copie à la maquette de mise en page*, Éditions du Centre de formation et de perfectionnement des journalistes, Paris, Diffusion La Découverte.
- Observatoire wallon des industries de la langue (1993) : *Des industries de la langue pour quoi faire ? technologies, usages et marchés*, A. Moulin (dir.), Liège, Université de Liège.
- PARKER, D. et B. STARRETT (1992) : *Technology Edge, Guide to CD-ROM*, Carmel, New Riders Publishing.
- PC Plus (1993) : «Tekstverwerker, dé favoriete toepassing», *PC Plus*, VNU, Amsterdam, Business Publications.
- PERRON, J. (1991) : «Présentation du progiciel de dépouillement terminologique assisté par ordinateur, TERMINO», *Les industries de la langue. Perspectives des années 1990. Actes du colloque*, tome II, Montréal, Office de la langue française et Société des traducteurs du Québec.
- PLOT, B. (1986) : *Écrire une thèse ou un mémoire en sciences humaines*, Champion, Unichamp.
- POIRIER, C. (1988) : «La place de l'informatique dans la préparation des dictionnaires de langue», P. Demers (dir.), *L'édition savante au Canada. Tendances actuelles et perspectives d'avenir*, Ottawa, Presses de l'Université d'Ottawa.
- POULAIN, M. (1988) : «Lectures et lecteurs», J. Vidal-Beneyto et B. Cassen (dir.), *Lire en Europe*, Strasbourg, Conseil de l'Europe, coll. «Études sur la culture et la communication».
- Publications du Québec (1986) : *Guide de présentation des manuscrits*, Québec, Gouvernement du Québec.
- REINKE, U. (1993) : «Towards a Standard Interchange Format for Terminographic Data», D. Schmitz (Ed.), *TKE' 83. Terminology and knowledge Engineering*, Frankfort, Indeks Verlag.
- REY-DEBOVE, J. (1971) : *Étude linguistique et sémiotique des dictionnaires français contemporains*, La Haye, Mouton.
- REZEAU, J. (1988) : «Que faire avec un outil professionnel en EAO des langues ?», *Les langues modernes. Le point sur l'EAO*, n° 5, Paris, APLV.
- RINT (1990) : «Harmonisation des méthodes en terminologie. Actes du séminaire. Talence, juin 1989 – Hull, décembre 1989, Terminologies nouvelles, Bruxelles, Ministère de la Culture de la Communauté française.
- ROGERS, S. (1992) : *ESPRIT. Information Processing Systems. The Synopses*, Bruxelles, Luxembourg, Commission of the European Communities, Directorate XIII.
- ROGERS, S. et S. TURIN (1992) : *ESPRIT. Basic Research. Projects, Actions, Working Groups and Network of Excellence*, Bruxelles, Luxembourg, Commission of the European Communities, Directorate XIII.
- ROLLIGND, L. (1993) : «Les ressources langagières. Un outil prodigieux pour les sciences de la connaissance», *XII Magazine*, Luxembourg, Commission des Communautés européennes.
- SINCLAIR, J. (1987) : *Looking UP. An Account of the COBUILD Project in Lexical Computing*, London, Collins Cobuild.
- SINCLAIR, J. (1991) : *Corpus, concordance, collocations*, Oxford, Oxford University Press.
- SOUFFLET, G. (1993) : *Cahiers pédagogiques*, n° 311, *Écrire avec l'ordinateur*, J. Crinon (dir.), Auxerre, CRAP.
- STEINACKER, I. (1992) : «The SYNTAX Textmill. An Intelligent Terminology Extractor», *Terminology Work in Subject Field. Third Infoterm Symposium. Proceedings*, Vienne, INFOTERM.
- TACKELS, S. (1988) : *Typographie et terminologie. Guide de présentation des travaux terminologiques*, Office de la langue française, Québec, Éditeur officiel du Québec.
- Technical Committee ISO/TC 97 (1986) : *Information Processing. Text and Office Systems. Standard Generalized Markup Language (SGML)*, ISO8879-(1986) (E), Vienne, ISO.

- TOIVO, R. (1988) : «Computerization of the Book Publishing Industry», P. Demers (dir.), *L'édition savante au Canada. tendances actuelles et perspectives d'avenir*, Ottawa, Presses de l'Université d'Ottawa.
- VAN DE BRAAK, P. (1988) : *How to Write Clear English*, Amsterdam, Academic Service.
- VANDERKAMP, R. (1988) : «The Impact of Computer Technology in the Production of Scholarly Journals and Books», P. Demers (Ed.), *Scholarly Publishing in Canada, Evolving Present, Uncertain Future*, Ottawa, University of Ottawa Press.
- VANSTEELANDT, A. (1993) : *Notes du cours d'initiation à la terminologie*, Bruxelles, Institut Libre Marie Haps (non publié).
- VESSILLIER-RESSI, M. (1988) : «Les sept paradoxes du métier d'auteur», J. Vidal-Beneyto et B. Cassen (dir.), *Lire en Europe*, Strasbourg, Conseil de l'Europe.
- WARWICK, S., HAJIC, J. et G. RUSSELL (1990) : «Searching on Tagged Corpora, Linguistically Motivated Concordance Analysis», *Electronic Text Research. Proceedings of the Conference*, Waterloo, UW Centre for the New OED and text Research.