Meta
Journal des traducteurs
Translators' Journal

META

La topographie des termes

Philippe Thoiron and Daniel Serant

Volume 34, Number 3, septembre 1989

1. Actes du Colloque Les terminologies spécialisées : Approches quantitative et logico-sémantique et 2. Actes du Colloque Terminologie et Industries de la langue

URI: https://id.erudit.org/iderudit/004485ar
DOI: https://doi.org/10.7202/004485ar

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

Thoiron, P. & Serant, D. (1989). La topographie des termes. *Meta*, *34*(3), 435–442. https://doi.org/10.7202/004485ar

LA TOPOGRAPHIE DES TERMES

PHILIPPE THOIRON
Centre de Recherche en Terminologie et Traduction
Université Lumière Lyon 2
DANIEL SERANT
Département de Mathématiques
Université Claude Bernard Lyon 1
Lyon, France

Nous présentons ici le début d'un travail qui a pour but d'étudier dans quelle mesure une approche quantitative d'un texte de spécialité peut aider à mettre en évidence les éléments dont le contenu informationel doit être considéré comme important. Nous voulons étudier aussi, mais dans un stade ultérieur, dans quelle mesure des critères de nature statistique peuvent contribuer à l'établissement d'une typologie des textes scientifiques, puisqu'il est maintenant bien admis qu'il n'existe pas un discours scientifique, mais plusieurs (cf. LOFFLER-LAURIAN, 1983 : 8-20).

Le critère quantitatif qui vient immédiatement à l'esprit, s'agissant de textes, est évidemment celui de la fréquence des mots (ou des termes). On a fondé beaucoup d'espoir sur la fréquence, dans de nombreux domaines des études linguistiques, allant de l'établissement du français fondamental (ou du Basic English) à l'indexation automatique. On sait que les déceptions ont été nombreuses, probablement à la mesure des espoirs excessifs qu'on avait suscités, et les inconvénients d'une utilisation exclusive du critère de fréquence sont bien connus (cf. par ex. comment JUILLAND a pu y remédier). La situation, en ce qui concerne les textes scientifiques, est assez bien résumée par deux positions apparemment contradictoires. Pour GALISSON (1978 : 112) «la fréquence n'est pas un critère pertinent pour différencier les termes techniques des termes courants». Pour HUIZHONG (1986 : 93-103) il est possible d'identifier les termes techniques et scientifiques en se fondant exclusivement sur leur comportement statistique. La contradiction ne tient que si l'on assimile «comportement statistique» et «fréquence». Or, on sait bien (et le travail de HUIZHONG en témoigne) que ceci est impossible : à fréquence égale, deux items lexicaux peuvent avoir des comportements statistiques différents variant, par exemple, avec leurs sous-fréquences dans les textes du corpus. De même, dans un texte, deux items de même fréquence peuvent avoir des valeurs très différentes en fonction notamment de la topographie des occurrences répétées, c'est-à-dire de leur position au fil du texte. On peut estimer qu'un texte où les items sont très espacés est moins «répétitif» qu'un texte où les mêmes items employés avec la même fréquence se trouvent regroupés en blocs (cf. les «rafales» de LAFON).

Le phénomène de la répétition est d'ailleurs souvent évoqué à propos des textes scientifiques. Il est bien sûr lié à celui de la fréquence mais ne lui est pas équivalent puisqu'il y ajoute, implicitement le plus souvent, la notion de proximité. On dit bien, en effet, qu'un texte est répétitif quand on y retrouve, à intervalles rapprochés, plusieurs occurrences des mêmes vocables. C'est donc le concept de topographie qui réapparaît.

Ce concept de répétition a pu servir à distinguer entre un discours littéraire, faisant appel à la synonymie et à la paraphrase pour des raisons de stylistique normative, et des discours scientifiques ou techniques où l'on reprend volontiers le même terme, qui est souvent monoréférentiel et tend à être normalisé (cf. TUKIA 1983 : 34-43 et RAFALO-

VICH 1983 : 45-59). On laissera de côté la synonymie et la monoréférentialité (développées par d'autres, y compris dans ce colloque) pour revenir sur deux points, l'un concernant le terme, l'autre les discours techno-scientifiques. S'agissant du terme, la question est de savoir s'il tend vraiment à être répété, et comment il est répété. Quant aux textes techno-scientifiques, sont-ils répétitifs ? Disons tout de suite, mais sans développer, que ces deux questions sont liées. En effet, compte tenu de la diversité des discours techno-scientifiques qui est mise en évidence dans le cadre d'une typologie, il serait assez surprenant que le statut du terme, par rapport au phénomène de la répétition, fût uniforme.

On voit donc que, à travers le phénomène de la répétition et plus particulièrement de la topographie des formes répétées, on cherche à étudier aussi bien les parties du texte, que représentent les termes, que l'ensemble du texte lui-même. Mais une telle étude est de type essentiellement comparatiste. Parler de répétition, c'est finalement (comme on le voit chez TUKIA et RAFALOVICH) comparer le «mot courant» au «terme technique», le discours littéraire aux discours techno-scientifiques. Or, à supposer que le «mot courant» soit une entité simple à définir, à supposer qu'un concept unitaire de «discours littéraire» ait un sens alors même qu'on insiste sur l'existence des discours techno-scientifiques, il restera à quantifier le phénomène de la répétitivité et à apprécier le sens des écarts entre les différentes valeurs de répétitivité.

C'est une proposition de quantification de la topographie que nous présentons ici, avec une illustration sur plusieurs textes scientifiques, afin de caractériser d'une part les éléments (et notamment les termes) et d'autre part l'ensemble (le texte lui-même).

I. ÉLABORATION DES INDICES DE MESURE DE LA TOPOGRAPHIE

Nous avons été amené à procéder en deux étapes :

a — Trouver, pour un item lexical donné, un indice qui reflète la topographie de ses formes répétées.

b — Trouver les moyens d'établir une distinction entre deux textes ayant des topographies de formes répétées différentes.

1. CONSTRUCTION D'UN INDICE DE TOPOGRAPHIE DES FORMES RÉPÉTÉES POUR UN TERME DONNÉ

Il s'agit, dans un premier temps, de trouver, pour un item de fréquence fixée, un indice de topographie des répétitions qui permette de discriminer entre les deux situations extrêmes suivantes :

1 — L'éparpillement, où le même item se répartit à peu près uniformément tout au long du texte

2 — Le regroupement ou effet de blocs où les intervalles sont cette fois très courts et dont la situation limite se réalise lorsque les items répétés se suivent immédiatement.

La méthode que nous proposons s'inspire de techniques utilisées dans la famille de tests dits «tests des suites» et qui servent à déceler, entre autres, l'existence de tendances ou d'anomalies dans une série statistique. Nous avons pour objectif d'établir, entre les différentes topographies, une hiérarchie qui s'appuie principalement sur les effets de blocs et ceci (tout au moins dans un premier temps) indépendamment de la longueur du texte dont est issue la séquence.

1.1. L'indice T.R.(s) :

Le texte T est représenté comme un segment de longueur N. Chaque forme peut occuper l'une des N positions possibles (numérotées de 1 à N) sur le segment. On s'intéresse à la topographie des répétitions d'un même item A de fréquence n dans T. Supposons par exemple qu'un texte ait une longueur de 100 mots et que le terme A soit répété 6 fois (n = 6) aux positions 1, 3, 5, 50, 52, et 55. On notera d_j la distance entre

deux occurrences successives de A. Cette distance est mesurée par la différence entre les positions des deux occurrences successives. Ainsi, dans notre exemple, d_j est égal à 2 puis 2 puis 45 puis 2 et enfin 3. La topographie des répétitions du terme A dans le texte T se résume donc dans la suite $(d_1, d_2,..., d_{n-1})$ des (n-1), soit ici 6 — 1 = 5, valeurs successives de d_j, soit ici la suite 2, 2, 45, 2, 3. On remarquera dès maintenant que cette suite est invariante par translation de l'ensemble des localisations de A dans le texte.

1.1.1. Le seuil s :

La suite $(d_1, d_2,..., d_{n-1})$ des (n-1) distances sera transformée en une suite 0 et de 1 à l'aide d'un test d'inégalité entre d_j et une valeur seuil s. Cette valeur peut être liée aux capacités mémorielles (cf. le «memory span» des psychologues anglais et américains). Si $d_j <= s$ on ramènera d_j à 1, si dj > s, d_j sera ramené à 0. Dans l'exemple ci-dessus, la binarisation de la suite de d_j donne les résultats suivants, pour s valant successivement 1, 2 et 3.

	2	2	45	2	3
seuil = 1					
test	?2<=1	?2<=1	?45<=1	?2<=1	?3<=1
résultat	0	0	0	0	0
seuil = 2					
test	?2<=2	?2<=2	?45<=2	?2<=2	?3<=2
résultat	1	1	0	1	0
seuil = 3					
test	?2<=3	?2<=3	?45<=3	?2<=3	?3<=3
résultat	1	1	0	1	1

On s'intéressera, au sein de cette suite de 0 et/ou 1, aux sous-suites ininterrompues (désormais SSI) de 1. En effet, il est évident que, outre le nombre total de 1 dans la suite, le paramètre «longueur des SSI» doit jouer un rôle crucial dans l'étude de la répétition.

1.1.2. Le coefficient TR(s) :

Le coefficient TR(s) (TR(s) pour «topographie des formes répétées au seuil s») a été conçu pour croître avec trois paramètres :

- le nombre de répétitions de l'item A
- le nombre de SSI de 1
- a longueur des SSI de 1

On désigne par m_k le nombre de SSI de 1 de longueur k. Avec notre exemple on aura, au seuil 1, zéro SSI de 1 ; au seuil 2 on aura une SSI de 1 ayant la longueur 2 et une autre ayant la longueur 1 ; au seuil 3 on aura deux SSI de 1 ayant chacune la longueur 2.

On définit TR(s) par :

$TR(s) = \sum k^2 m_k$ (pour k allant de 1 à l'infini)

Si n, nombre de répétitions de A est égal à 1 on pose par convention que TR(s) = 0.

Le calcul de TR(s) pour notre exemple se fait donc comme suit :

seuil = 1 0 SSI k = 0 m = 0
d'où TR(1) = 0
seuil = 2 1 SSI de longueur 1 (donc k = 1, m_k = 1)
et 1 SSI de longueur 2 (donc k = 2, m_k = 1)
d'où TR(2) = $(1^2*1)+(2^2*1) = 5$
seuil = 3 2 SSI de longueur 2 (donc k = 2, m_k = 2)
d'où TR(3) = $2^2 + 2^2 = 8$

1.2. L'indice TRCum :

Il n'y a aucune raison de donner à s une seule valeur choisie arbitrairement. On a donc calculé TR(s) pour toutes les valeurs possibles de s et cumulé les TR(s) ainsi obtenus. Avec l'exemple donné ci-dessus on aurait, pour les premières valeurs de s, obtenu 0 + 5 + 8 = 13. Toutefois, il faut observer que, dans une opération de cumul simple, tous les TR(s) ont la même importance. Ainsi, telle SSI de 1 obtenue avec s = 1 a le même poids qu'une SSI obtenue avec s = 15. Or, on est en droit de considérer qu'une topographie de type AAA doit être traitée différemment d'une topographie de type A .... A .... A . C'est à cet effet que, pour mieux faire ressortir l'effet de blocs, les TR partiels ont été pondérés, dans l'opération de cumul, au profit des valeurs de s les plus faibles. Les coefficients de pondération des TR(s) seront donc des fonctions décroissantes des valeurs du seuil s.

Nous avons choisi une décroissance de type «décroissance lente» et défini TRCum par

$$TRCum = c^*\Sigma TR(s)/s^2 \text{ où la constante c est égale à } 6/\pi^2$$

La constante c est une constante de normalisation : la somme de la série de terme général est égale à $\pi^2 / 6$ et TRCum est encore maximum pour $(n-1)^2$. Nous avons normalisé le coefficient TRCum en le divisant par $(n-1)^2$. Cette normalisation a pour effet de minimiser l'importance du paramètre n. Le coefficient ainsi normalisé (appelé TRNOR) est alors très directement lié à la structure des localisations et peut servir d'indice «absolu» permettant de comparer les topographies de deux termes dont les nombres de répétitions seraient différents.

2. TOPOGRAPHIE DE L'ENSEMBLE DES ITEMS D'UN TEXTE

Si l'on s'intéresse non plus à l'item pris isolément mais au texte dans son ensemble, plusieurs opérations peuvent être conduites.

2.1. Mise en évidence des termes à topographie remarquables

Il s'agit ici de savoir si un terme de fréquence n présente ou non des localisations remarquables de ses formes répétées, soit par effet de blocs, soit, à l'opposé, par effet d'éparpillement. Comme chez LAFON, le modèle de référence à partir duquel on jugera de la déviance des topographies repose sur le choix de n positions «au hasard» prises dans l'ensemble (1, 2, ... N), (loi uniforme les combinaisons de n objets pris dans N).

Simulations

Sous ce modèle, la loi de la variable aléatoire TRCum est théoriquement calculable. Mais on aurait alors besoin d'évaluer TRCum pour toutes les combinaisons de n objets pris dans N ce qui nécessite une puissance de calcul dépassant largement celle d'un micro-ordinateur. On peut néanmoins, avec un micro-ordinateur, réaliser un grand nombre (par exemple quelques centaines) de simulations indépendantes et obtenir ainsi une estimation de la loi, et notamment de la moyenne et de la variance de TRCum. À partir de ces paramètres, et en acceptant l'hypothèse gaussienne, la méthode de rejet du modèle est bien connue.

Pour chacun des vocables de fréquence supérieure à 1 d'un texte de notre corpus, on a calculé TRCum ainsi que la moyenne des TRCum simulés sur 200 épreuves.

2.2. Mesure globale de la topographie des éléments d'un texte

Nous avons mis au point deux indices permettent d'évaluer, globalement, les caractéristiques d'un texte. Il s'agit, d'une part, de la somme des TRNOR (ΣTRNOR), qui reflète en effet presque exclusivement les topographies des termes, et d'autre part, de la

somme des TRCum (ΣTRCum), qui prend en compte à la fois, la fréquence des termes et leur topographie. Toutefois, dans la présente communication, nous laisserons de côté cet aspect de notre recherche qui a pour but d'évaluer le rôle que peuvent jouer des critères statistiques dans l'élaboration d'une typologie des discours scientifiques.

Afin d'apprécier la signification des valeurs prises par tous nos indices, nous avons mis au point des programmes qui permettent de générer aléatoirement des textes ayant la même distribution de fréquences que le texte étudié.

II. APPLICATIONS

Nous avons appliqué la méthode à titre expérimental à un corpus de textes (cf bibliographie) empruntés aux domaines suivants :

- médecine (plusieurs textes sur le SIDA),
- écologie (les eaux douces),
- mathématiques.

Les textes d'écologie sont en anglais, les autres en français et tous ont été lemmatisés.

Nous avons traité le problème de manière formelle (en nous intéressant aux termes et non aux notions) et dans le cadre du texte. Nous n'avons pris en compte la topographie de formes répétées que lorsqu'elles sont absolument identiques (i.e. lymphocyte est différent de lymphocyte T4). Les observations faites concernent donc le fonctionnement d'un terme donné (ou de plusieurs) dans un texte donné. Il serait dangereux d'en déduire quoi que ce soit concernant la valeur «absolue» du (ou des) terme(s).

Pour chacun des éléments de chacun des textes du corpus (et pas seulement pour les termes) nous avons calculé les valeurs suivantes qui apparaissent dans des tableaux récapitulatifs comme le tableau 1 : la fréquence, TRCum, TRNOR (tous trois calculés directement sur le texte), TRNORSI et sa variance (qui proviennent de la série des simulations) et l'écart réduit z (qui fournit une échelle permettant de mesurer l'écart entre TRNOR et TRNORSI, c'est-à-dire, pour chaque terme, entre la valeur observée et la valeur estimée).

1. OBSERVATIONS GLOBALES

Beaucoup de TRNOR sont inférieurs au TRNORSI (cf. les $Z < 0$). La cause en est simple. Dans le modèle qui a été utilisé pour les simulations, seul le «hasard» est intervenu. Or, le hasard n'a ni syntaxe ni mémoire, de sorte que les items peuvent, par exemple, être adjacents. Les textes constitués ainsi ont donc tendance à présenter des effets de blocs plus nombreux que les textes réels. Il faudra donc considérer comme d'autant plus significatives les déviations positives (donc $z > 0$) et s'intéresser spécialement aux termes concernés.

Si TRCum est assez fort (en tête de liste de classement par valeurs décroissantes) et TRNOR non remarquable (correspondant à une valeur de Z comprise entre 0 et — 1 par exemple), on a affaire à un terme fréquent mais dont les occurrences ne sont pas regroupées (toujours, rappelons-le, par référence à un modèle de répartition aléatoire sur le texte). Il s'agit d'un terme qui dans le texte en question a une portée générale : c'est un terme dont nous dirons qu'il est, dans ce texte, «générique» du domaine sans donner à ce mot l'acception restrictive qu'il prend dans les études sur les taxinomies populaires (cf. BERLIN *et al.*, 1973 : 216 ; WIERZBICKA : 1985 ; CRUSE : 1986). Les termes «génériques» et «spécifiques» sont pris ici au sens large et renvoient à une distinction binaire commode au sein de la hiérarchie entre les éléments situés plutôt vers le sommet et ceux qui sont situés plutôt vers la base.

Les exemples sont nombreux : il s'agit par exemple de *SIDA, virus, système immunitaire, protéine, Lymphocyte T, Lymphocyte T4*, etc. dans le texte de J. LAURENCE

consacré au SIDA, de *virus, SIDA, sang, cellule, lymphocyte, défense*, dans un texte de l'Institut Pasteur (1987), rédigé sous la direction de Luc MONTAGNIER et destiné au grand public (cf. sa recommandation par les principales associations de parents d'élèves). *Virus* (et *viral*), *SIDA* se retrouvent dans les mêmes conditions dans un autre texte de MONTAGNIER (1985) traitant de l'étiologie virale du SIDA. Les termes *phosphorus loading, algae, water, plant* présentent eux aussi ces caractéristiques dans un texte consacré à la teneur en phosphore de l'eau des lacs. De même pour *lake, fish, population, species, water, char*, dans un texte consacré à une espèce (species) de poisson (fish), l'omble chevalier (char) et plus particulièrement au sous-groupe lacustre (BURGIS et MORRIS, 1987). Dans un texte d'introduction à l'analyse convexe en mathématique, on trouve *analyse, calcul, application, mathématique, théorie, optimisation, solution, variation*, par exemple.

Si TRNOR est remarquable, et quelle que soit la valeur de TRCum,on a évidemment un terme dont les répétitions forment un effet de bloc (ou de «rafale» dans la terminologie de LAFON). Ses occurrences sont très rapprochées et se trouvent en un ou plusieurs endroits du texte. Il s'agit souvent de termes plus spécifiques dans le domaine, correspondant à des notions plus fines. Ainsi, dans le texte présentant l'omble chevalier les termes *spawn, line*, ... qui concernent des aspects plus limités, ou si on préfère, des sous-thèmes, dans le texte (ici le frai, la pêche, ...). Dans le texte de LAURENCE *acide aminé, génome, HTLVI, interleukine 1, lymphocyte T suppresseur, transcriptase inverse, TAT*. Dans celui de l'Institut Pasteur, *Système immunitaire, code, ARN, ADN, VIH, test, génétique, anticorps, protéine, lymphocyte T, lymphocyte B, séropositif, cellule T4, VIH 1, etc. Dans celui de MONTAGNIER anticorps, protéine, antigène, HTLV 1, porteur, rétrovirus, électrophorèse, syndrome, HTLV III*, etc. Dans le texte consacré au phosphore *macrophyte, weed, concentration, mix*, par exemple. Dans le texte de mathématique *dualité, fonction, primal, sous-différentiabilité* par exemple.

2. INDICES DE TOPOGRAPHIE ET HIÉRARCHISATION DU DOMAINE

On pourrait, sur la foi de ces observations, estimer que les indices de topographie reflètent la bipartition d'un domaine entre générique et spécifique. Mais un examen attentif des résultats montre que, à l'intérieur d'un même domaine, un terme donné peut avoir des topographies très différentes selon les textes, voire selon les différents fragments du même texte. Un exemple suffira. *Lymphocyte T* a une topographie remarquable (effet de blocs) dans le texte de l'Institut Pasteur mais pas dans celui de LAURENCE. Ceci nous a conduit à nous interroger sur la relation entre nos indices et la hiérarchisation du domaine.

Quand, dans un texte, une notion est présentée et développée, l'auteur est confronté au choix suivant :

- ◆ répéter le même terme,
- ◆ trouver un synonyme, ou recourir à la paraphrase ou à l'anaphore.

Il n'y a pas de fatalité dans la répétition des termes. On peut trouver des synonymes en terminologie aussi (voir par exemple les 14 équivalents anglais de *Arctic char* cités dans le *Bulletin de Terminologie* No. 161 consacré à l'ichtyologie) mais ce n'est pas toujours possible. Pour ce qui est de la «paraphrase» c'est parfois possible, parfois souhaitable — notamment dans des textes à vocation didactique où la paraphrase est volontiers explicative. S'agissant de l'anaphore, on laissera de côté l'emploi des pronoms pour s'attarder un peu sur les anaphores «lexicales».

Lorsqu'un terme est intégré dans une hiérarchie, on peut toujours, logiquement, dans le cadre d'un texte, lui substituer un hyperonyme. C'est ce qui se pratique souvent, notamment semble-t-il dans les textes didactiques où le procédé sert alors à replacer le

terme dans la terminologie (hiérarchie) contribuant ainsi à une meilleure compréhension et une mémorisation plus efficace de celle-ci. C'est ainsi que *lymphocyte T4 suppresseur* peut être repris par *lymphocyte T4* ou *lymphocyte* voire *cellule.*

Si ce procédé fonctionne bien, et assez fréquemment, avec les termes spécifiques du texte, ceux-ci n'en ont pas le monopole. On peut trouver le même phénomène avec les génériques d'un texte car la hiérarchisation est extérieure au texte, elle appartient au domaine lui-même. Ainsi avec *Artic Char* (qui est un générique dans le cadre du texte que nous avons), on a des reprises à l'aide de *population, species*, etc. Il semble que ceci puisse être le cas lorsque ce générique est subdivisé en ses spécifiques (cf. *migratory* and *landlocked* ici) mais alors la fréquence de ces termes est faible.

On arrive donc bien ici à une distiction entre deux catégories de termes : ceux qui ont une fréquence forte, alliée à un TRNOR faible et ceux qui ont un TRNOR fort, allié à une fréquence faible ou forte. Les premiers sont, dans le texte traité, plutôt génériques, les seconds plutôt spécifiques. Cette opposition peut être affinée et servir de base non pas à une partition mais à un repérage dans un espace à deux dimensions, FREQ et Z (cf. fig. 1).

CONCLUSION

Si nos indices sont utilisables dans la représentation de la topographie des termes, ils peuvent servir aussi à la localisation, dans un texte, ou dans un corpus, de fragments dans lesquels une notion est développée. Cette localisation peut être faite automatiquement puisque nos programmes attribuent à chaque terme les valeurs de position de ses occurrences dans le texte. On pourra donc extraire facilement le fragment de texte compris entre des bornes dépendant de ces valeurs de position. Il faut répéter ici que la topographie remarquable d'un terme n'est pas une condition nécessaire dans le repérage. On peut imaginer, en effet, qu'une notion soit abordée, puis développée, sans qu'aucun terme du domaine ait une topographie remarquable par effet de blocs. En revanche, une topographie remarquable constitue une condition suffisante puisqu'elle permet d'attirer l'attention sur des zones où la probabilité est forte de trouver les notions spécifiques du texte. C'est pourquoi nous pensons que la topographie des termes pourrait constituer un critère utile dans le cadre général d'un protocole de traitement automatique des textes.

Note

On ne s'intéresse dans cette étude qu'à la topographie des termes (on a bien sûr supposé résolu le problème de leur identification) mais les programmes réalisés permettent de traiter n'importe quelle catégorie.

BIBLIOGRAPHIE

BERLIN, B., BREEDLOVE, D. et RAVEN, P. : «General Principles of Classification and Nomenclature in Folk Biology», *American Anthropologist*, 75, pp. 214-242.

BUREAU DES TRADUCTIONS (1978) : Direction de la Terminologie, *Ichtyologie*, Ottawa : Approvisionnements et Service Canada.

BURGIS, Mary L. et MORRIS, Pat (1987) : *The Natural History of Lakes*, Cambridge, C.U.P.

CRUSE, D. A. (1986) : *Lexical Semantics*, Cambridge, C.U.P.

EKELAND, Ivar (1983) : *Le Calcul et l'imprévu*, Paris, Seuil.

EKELAND, Ivar et TEMAM Roger (1974) : *Analyse convexe et problèmes variationnels*, Paris, Dunod-Gauthier-Villars.

GALISSON, Robert (1978) : *Recherche de lexicologie descriptive : la banalisation lexicale*, p. 112, Nathan, Paris.

GODSHALK, G. L. et WETZEL, R. G. (1978) : «Decomposition in the Littoral Zone of Lakes», in GOOD, R. E. et WHIGHAM D. F. ed. Freshwater Wetlands, New York, Academic Press, pp. 131-143.

HUIZHONG, Yang (1986) : «A New Technique for Identifying Scientific / Technical Terms and Describing Science Texts», *Literary and Linguistic Computing*, Vol. 1, n° 2, pp. 93-103.

INSTITUT PASTEUR (1987) : sous la direction du Pr. Luc MONTAGNIER, SIDA : les faits, l'espoir, Paris.

JUILLAND, A. et al. (1970) : Frequency Dictionary of French Words, La Haye-Paris.

LAFON, P. (1984) : *Dépouillements et statistiques en lexicométrie*, Genève-Paris, Slatkine-Champion.
LAURENCE, Jeffrey (1986) : «Le système immunitaire et le Sida», *Pour la science*, février 1986, pp. 40-51.
LEIBOWITCH, J. (1983) : «Le syndrome d'immuno-déficit acquis (SIDA) : problèmes diagnostiques, thérapeutiques et étiopathologiques», *Le Concours médical*, 23/30-07-1983, pp. 3257-3264.
LOFFLER-LAURIAN, Anne-Marie (1983) : «Typologie des discours scientifiques : Deux approches», *Études de Linguistique Appliquée*, 51, pp. 8-20.
MOOD M., A. GRAYBILL, DUANE, *Introduction to the Theory of Statistics*, New York, McGraw-Hill.
MONTAGNIER, Luc (1985) : «L'étiologie virale du SIDA et son impact en Santé Publique», *J.A.M.A.*, vol. 10, pp. 414-417.
RAFALOVICH, Hilmar (1983) : «Négativité ou créativité des langues de spécialités allemandes», *Études de Linguistique Appliquée*, 51, pp. 45-59.
TUKIA, Marc (1983) : «Observations sur le vocabulaire, sur les marques d'énonciateur et sur la construction dans le discours scientifique», *Études de Linguistique Appliquée*, 51, pp. 34-43.
WIERZBICKA, Anna (1985) : *Lexicography and Conceptual Analysis*, Ann Arbor, Karoma.