Meta
Journal des traducteurs
Translators' Journal

META

Approche lexicale des registres en langues de spécialité

Danielle Candel and Pierre Lafon

Volume 39, Number 4, décembre 1994

Hommage à Bernard Quemada : termes et textes

URI: https://id.erudit.org/iderudit/002196ar
DOI: https://doi.org/10.7202/002196ar

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

Candel, D. & Lafon, P. (1994). Approche lexicale des registres en langues de spécialité. *Meta*, *39*(4), 807–815. https://doi.org/10.7202/002196ar

Article abstract

This article shows some results of an experiment carried out on an Lsp corpus. Four texts were selected and classified from the most general to the most specialized. They were analyzed with the help of a statistical program, in view to testing the initial classification and to seeing whether significant differences exist between the texts.

APPROCHE LEXICALE DES REGISTRES EN LANGUES DE SPÉCIALITÉ

DANIELLE CANDEL* ET PIERRE LAFON**
*CNRS-INaLF, Paris, France
**CNRS-INaLF, ENS de Fontenay-Saint-Cloud, Saint-Cloud, France

Résumé

Cet article décrit les résultats d'une expérience réalisée pour contribuer à l'étude des registres dans les textes de spécialité. Quatre textes, classés au départ du général vers le spécialisé, ont été traités par un logiciel d'analyse statistique. On tente, par ce type d'analyse, de tester la fiabilité du classement initial, et de faire apparaître des différences marquantes entre ces textes.

Abstract

This article shows some results of an experiment carried out on an Lsp corpus. Four texts were selected and classified from the most general to the most specialized. They were analyzed with the help of a statistical program, in view to testing the initial classification and to seeing whether significant differences exist between the texts.

Registres de spécialisation ou de technicité et niveaux de langue participent, dans les langues de spécialité, aux processus de la variation linguistique. Pour les besoins de la présente étude, nous avons réuni un corpus de faible dimension, qui devait tendre à l'homogénéité et être représentatif d'un domaine de spécialité. Ce corpus se compose de quatre sous-parties distinctes, se prêtant à une étude comparative[1].

1. PRÉSENTATION GÉNÉRALE

1.1. Démarche adoptée

Nous avons souhaité proposer ici une étude de l'influence de la spécialisation sur le lexique des textes. L'hypothèse la plus vraisemblable nous semblait être que les textes les plus spécialisés comportaient le vocabulaire le moins étendu et que les textes les plus généraux proposaient un vocabulaire plus riche ou plus varié[2]. Pour tenter de vérifier cette hypothèse, il fallait constituer un corpus dont la seule variable de production soit le degré plus ou moins grand de spécialisation.

Nous avons choisi d'étudier la langue d'un scientifique à travers plusieurs de ses écrits. Paul Germain est un scientifique éminent, spécialiste d'un domaine de la physique, reconnu comme tel par la communauté scientifique. Longtemps professeur de mécanique à l'École Polytechnique, il est depuis 1975 Secrétaire perpétuel de l'Académie des Sciences. Grand spécialiste de la mécanique, il œuvre pour le développement et l'évolution de la science, dont il s'efforce constamment d'amplifier la diffusion. À sa culture profondément et essentiellement scientifique s'ajoute une préoccupation que l'on retrouve dans plusieurs de ses textes pour l'avenir de la langue française. Il est membre du Conseil supérieur de la langue française, dont Bernard Quemada est le Vice-président.

Paul Germain a eu l'occasion d'écrire des textes et de prononcer des discours où il exprime ses idées sur la langue des sciences, aussi bien que des textes sur la science, et des manuels d'enseignement supérieur de haut niveau.

1.2. Le corpus

Il est formé de textes saisis exhaustivement ou partiellement, ressortissant à divers degrés de spécialisation. Nous avons réuni les quatre textes suivants, en les classant spontanément dans une échelle qui nous semblait aller du plus général au plus particulier :

1. *Langue française et mutation du monde*, 1987, pp. 17 à 29 ;
2. *Le français dans les sciences et les techniques*, 1990, pp. 2 à 27 ;
3. *La science interpellée*, 1989, pp. 1 à 11 ;
4. *Mécanique. École Polytechnique*, 1986 (fragments)[3].

Les textes 1, 2 et 3 ont été retranscrits exhaustivement[4]. Le texte 4, qui provient de deux importants volumes du *Cours de Mécanique* à l'École Polytechnique, est constitué de fragments : on a retenu les introductions et conclusions de chapitres, de manière à privilégier les passages à portée plus générale, au détriment des pages d'exposés trop mathématiques, spécialisées et techniques. Ces dernières comportent davantage de formules, de calculs et de tableaux, et s'avéraient plus difficilement exploitables dans le cadre de la présente étude.

Ces écrits sont tous rédigés par le même auteur en l'espace de quatre ans, ce qui leur confère un degré relatif d'homogénéité. Ils ont en commun plusieurs éléments. Ce ne sont pas des textes hautement spécialisés et qui risquaient de présenter un aspect fortement néologique, par l'emploi de nombreux termes techniques de pointe. Les trois premiers participent bien plutôt d'une vulgarisation de haut niveau. D'autre part, ils ont en commun d'être des textes d'«écrit oral» (plutôt que d'«oral écrit»), ce sont tous des versions écrites de textes prononcés ou de cours professés. Le texte 1 est une conférence, le texte 2 est un rapport présenté oralement (qui «reprodui[t] intégralement la présentation faite oralement [...] du rapport rédigé [...])[5], le texte 3 est une «Lecture» en séance solennelle à l'Académie des Sciences, et le quatrième est un «Cours». Les quatre textes étudiés tendent donc à constituer un corpus relativement homogène.

Ils sont cependant aisément différenciables les uns des autres. Ils traitent de sujets relatifs à la science, mais à des degrés divers : il s'agit d'une part de la langue des sciences (textes 1 et 2[6]), d'autre part, de la science en tant que telle (texte 3), et plus précisément, de la mécanique (texte 4). Ils peuvent être distingués non seulement par les thèmes qu'ils exposent, mais aussi par les publics auxquels ils s'adressent : respectivement, les participants au colloque portant sur «Le français, langue des sciences et des techniques», les participants à la Séance plénière du Conseil supérieur de la langue française, les personnes présentes à la séance solennelle de l'Académie du 27 novembre 1989, et les élèves de l'École Polytechnique. Ils revêtent, enfin, des formes éditoriales différentes : un recueil d'articles dans un volume broché à couverture souple, deux fascicules brochés, et les deux importants volumes du Cours[7].

Ce corpus se compose en tout de 41 229 occurrences, correspondant à 5 548 formes différentes, et on y dénombre 2 772 hapax. Les mots outils de très grande fréquence[8] ne retiendront pas notre attention dans les pages qui suivent, pas plus que les auxiliaires et les auxiliaires de mode.

1.3. Outil d'analyse

Le logiciel d'analyse utilisé est celui du laboratoire «Lexicométrie et textes politiques» de Saint-Cloud. Avant tout, il permet le relevé et le tri alphabétique des occurrences de formes et de segments à l'intérieur d'un corpus aussi bien que des sous-corpus qui le composent. Il traite également les dépouillements obtenus par diverses méthodes statistiques (spécificités, analyse factorielle, etc.).

2. COMPARAISON MACROSCOPIQUE DES QUATRE SOUS-CORPUS

Cette approche laisse de côté dans un premier temps le contenu des vocabulaires utilisés. Nous ne prenons en considération que les chiffres relatifs à chaque fragment du corpus étudié. Nous nous trouvons donc ici dans une situation classique mais incommode, pour comparer l'étendue des lexiques mis en jeu dans nos quatre fragments, à cause notamment des différences de taille, puisque les longueurs en nombre d'occurrences passent de 3 763 dans *Langue française et mutation du monde* à 17 864 dans le *Cours de mécanique.*

Charles Muller a proposé une méthode, devenue classique depuis, qui permet d'effectuer cette comparaison. Il n'est pas nécessaire de reprendre ici le détail de la méthode et des calculs, nous renvoyons le lecteur à l'exposé initial très clair de Charles Muller[9].

On sait que le nombre de formes différentes d'un texte résulte de l'addition des effectifs de toutes les classes de fréquence (1, les hapax, 2, 3, etc.): on trouvera au *Tableau 1* l'effectif des trois classes les plus nombreuses de chaque texte. La méthode consiste, en partant de la gamme des fréquences d'un fragment long, à estimer l'étendue du lexique qu'il aurait si sa longueur se réduisait à celle d'un fragment plus court. En prenant les textes paire par paire on établit ici facilement toutes les comparaisons, soit: 4/3, 4/2, 4/1 ; puis 2/3, 2/1 ; enfin 3/1.

Tableau 1.
Présentation des quatre textes

sous-corpus	occurrences	formes différentes	basses fréquences
1 Langue française et mutation du monde	3 763 = 9 % du corpus total	1 239	fr.1 = 843 fr.2 = 192 fr.3 = 67 ...
2 Le français dans les sciences et les techniques	13 392 = 32 % du corpus total	2 408	fr.1 =1284 fr.2 = 418 fr.3 = 207 ...
3 La science interpellée	6 210 = 15 % du corpus total	1 619	fr.1 =1045 fr.2 = 240 fr.3 = 108 ...
4 Mécanique (Cours, École Polytechnique)	17 864 = 43 % du corpus total	3 204	fr.1 =1711 fr.2 = 547 fr.3 = 256 ...
corpus total : 1 + 2 + 3 + 4	41 229	5 548	fr.1 =2772 fr.2 =1403 fr.3 = 484 ...

Il ressort de ces calculs que la hiérarchie des richesses s'établit de la manière suivante :

Le texte 1, *Langue française et mutation du monde*, possède le vocabulaire le plus étendu de tous.

Le texte 2, *Le français dans les sciences et les techniques*, a le vocabulaire le plus restreint de tous.

Les textes 3 et 4, respectivement *La science interpellée* et le *Cours de mécanique* sont intermédiaires, mais 3 est légèrement plus étendu que 4.

On a donc pour nos quatre fragments le classement suivant: 1 est plus étendu ou plus varié que 3, lui-même plus étendu que 4, lui-même plus étendu que 2.

L' hypothèse que nous avions faite au début, selon laquelle plus un texte est spécialisé moins son lexique est étendu, s'avère ici partiellement fausse. S'il est bien vrai que le texte le plus varié correspond au thème le plus général, le texte le plus spécialisé, ici manifestement *Le cours de mécanique*, n'est pas le moins étendu[10].

En étudiant maintenant les formes du lexique qui entrent dans les différents registres, nous tenterons de mieux expliquer et d'exemplifier ce phénomène.

3. ANALYSE DES TEXTES

3.1. Méthode d'analyse

3.1.1. Analyse par indices de spécificité

L'analyse du «calcul des spécificités» est effectuée par le logiciel de Saint-Cloud: «méthode qui consiste dans la probabilisation de toutes les fréquences locales obtenues pour les formes et les segments d'un corpus selon la partition étudiée» (Tournier, ici même, note 8). La probabilité de rencontrer un mot ou une séquence de mots dans l'un des sous-corpus est évaluée en fonction de sa présence dans l'ensemble du corpus.

L'indice de spécificité est signalé par la marque d'abréviation «Sp». Cette marque est suivie du signe + si l'indice de spécificité est positif, du signe – si l'indice de spécificité est négatif. Ainsi, la marque «Sp+» signale que le mot ou le terme est «sur-employé» dans le texte par rapport à l'ensemble du corpus, et la marque «Sp-» indique que l'unité est «sous-employée» dans le texte par rapport au corpus total (v. encore Tournier, ici même, note 9, et Lebart et Salem 1988: 187-188). «Sp+» ou «Sp-» sont suivis d'un indice chiffré dont la valeur correspond à une spécificité plus ou moins grande. Ils vont, dans nos listages, de Sp+35 à Sp – 32.

3.1.2. Analyse par segments répétés

Le logiciel de Saint-Cloud offre la possibilité d'analyse en «segments répétés» (Lafon et Salem 1983; Salem 1987; Lafon 1994). Toutes données confondues, l'étude par segments répétés atteste, pour les formes répétées trois fois au moins dans le corpus total, un ensemble de 2 881 formes différentes. Ces données sont brutes, et un tri manuel s'impose, pour éliminer les segments qui ne semblent pas pertinents pour cette analyse (par exemple *dire en, donc être, entre la, lui est,* ou encore *et de la*).

C'est donc l'ensemble des unités rencontrées plus de trois fois dans le corpus, mots isolés et segments, qui sera pris en considération, après un tri manuel effectué sur les données.

3.2. Résultats de l'analyse

Nous orientons notre analyse d'abord en fonction des indices de spécificité, et dans un second temps seulement, en fonction du classement fréquentiel. La recherche par indices de spécificité présente en effet l'avantage de nous renseigner sur le contraste entre la fréquence du terme dans un fragment et celle qu'il a dans le corpus total.

Un rapide parcours des relevés par coefficients de spécificité donne les mots-clés définis soit positivement soit négativement au *Tableau 2*. Ce tableau propose aussi, à titre de comparaison, les résultats des fréquences.

Tableau 2.
Unités lexicales les plus représentatives du corpus,
par leur présence ou par leur absence dans les sous-corpus

Texte	Unité linguistique (précédée de l'indice de spécificité extrême)	Fréquence dans le texte (par rapport à la fréquence totale dans le corpus)	Fréquence la plus élevée
1 : *Langue française et mutation du monde*	Sp.+13 : **langue** Sp. – 05 : **mécanique**	41 (sur 126) 0 (sur 100)	**langue** (41)
2 : *Le français dans les sciences et les techniques*	Sp.+20 : **pays** Sp. – 18 : **mécanique**	54 (sur 60) 0 (sur 100)	**scientifique** (109)
3 : *La science interpellée*	Sp.+23 : **science** Sp. – 09 : **langue**	63 (sur 114) 0 (sur 126)	**science** (63)
4 : *Cours de Mécanique*	Sp.+35 : **mécanique** Sp. – 32 : **langue**	99 (sur 100) 0 (sur 126)	**mécanique** (99)

Les unités qui ont attiré notre attention sont celles relevant du vocabulaire scientifique ou technique ou du vocabulaire général[11]. Les études terminologiques traitent le plus souvent de la classe des substantifs. Nous avons tenu à accorder une égale importance aux verbes du corpus, aux mots de liaison et de subordination (autres que les très fréquents *et, que, qui...*), aux adjectifs, aux adverbes, aux locutions mais aussi à l'utilisation par l'auteur des marques de la première personne.

3.2.1. Analyse d'unités lexicales représentatives de chaque sous-corpus

Nous tentons d'évaluer successivement les unités rencontrées dans chacun des sous-corpus. Pour un complément d'information, nous évoquons brièvement les mots absents de chacun de ces textes et présents dans chacun des trois autres sous-corpus : pour ce faire, nous proposons des extraits de nos relevés.

a) Unités caratéristiques et unités absentes du sous-corpus 1

Tous relevés confondus, voici trois termes, attestés uniquement dans ce texte, et qui en sont les plus représentatifs : *industries de la langue* (Sp+8), *institutions scientifiques* et *programme mobilisateur* (Sp+4).

Il est intéressant par ailleurs de noter quelques unités, attestées dans les trois autres sous-corpus, et absentes de ce texte, comme *concept(s), connaissance(s), la connaissance scientifique, à titre d'exemple, prise en compte, en ce qui concerne, il convient de, rendre compte, si l'on veut.* Nous avons relevé également seize adverbes et locutions adverbiales.

b) Unités caratéristiques et unités absentes du sous-corpus 2

Parmi les unités les plus caractéristiques de ce texte, et attestées uniquement dans ce texte, nous avons retenu les substantifs et syntagmes substantivaux *centres* (Sp+12), *vie scientifique et technique* (Sp.+10), *laboratoires* (Sp.+7), affectés des indices de spécificité les plus élevés, puis *développement scientifique et technique, expression scientifique et technique, lignes de conduite, rayonnement de la science française, science mondiale* (Sp.+3) et la locution *à cet égard* (Sp.+3).

Un ensemble d'autres termes, bien que qualifiés de «banals» par le logiciel d'analyse, et qui ne sont donc pas affectés d'un coefficient de spécificité, ont eux aussi la particularité d'être présents seulement dans le texte 2. C'est le cas des substantifs et syntagmes substantivaux *centres de formation, centres de recherche, communauté scientifique, bon niveau, milieux scientifiques, points forts, prise de conscience, public scientifique, spécialistes, techniciens, théories nouvelles,* des verbes et syntagmes verbaux *conduire l'action, faisant appel, il est certain que, il est clair, il faut encore,* et des locutions *ceci précisé de (très) haut niveau, de mise au point, en la matière.*

Certaines unités sont attestées dans les trois autres sous-corpus, et non dans ce texte, comme les adjectifs ou pronom de la première personne *mes, moi, mon.*

c) Unités caratéristiques et unités absentes du sous-corpus 3

Au nombre des unités les plus caractéristiques du texte 3, et attestées dans ce seul texte, nous avons relevé les substantifs ou syntagmes substantivaux *nos sciences* (Sp.+18), *le monde de l'entendement* (Sp.+6), *espace de liberté* (Sp.+5), *des connaissances scientifiques, l'image de la science* (Sp.+4) et, avec une référence au domaine précis de spécialisation, *biologie, monde mathématique* (Sp.+3) ; signalons enfin l'expression *noyau dur et pur* (Sp.+3) et le pronom personnel *vous* (Sp.+3).

D'autre part, parmi les unités attestées dans les trois autres sous-corpus et absentes de celui-ci, notons les verbes et syntagmes verbaux *il s'agit de, on constate, on sait,* et plusieurs adverbes, dont *actuellement, bientôt, maintenant,* les locutions *dans une large mesure, de toute évidence, en petit nombre, en premier lieu* et *quelque peu.*

d) Unités caratéristiques et unités absentes du sous-corpus 4

Parmi les unités les plus caractéristiques du texte 4, et présentes dans ce seul texte, citons d'abord *fluides* et *mécanique* (Sp.+16), affectées d'un indice de spécificité particulièrement élevé. Citons ensuite les substantifs ou syntagmes substantivaux *milieux continus* (Sp.+11), *un milieu* (Sp.+7), *corps rigide,* (Sp.+6), *efforts intérieurs, équations générales, loi(s) de comportement* (Sp.+4), *écoulements de fluides, énoncés fondamentaux, fluide parfait, (petites) perturbations, puissance(s) virtuelle(s), schématisation, système S* (Sp.+3), et, avec une référence précise au domaine de spécialisation, *mécanique classique* (Sp.+8), *mécanique des fluides, mécanique des milieux continus* (Sp.+4) et *mécanique des solides* (Sp.+3). Notons aussi les adjectifs *élastique(s), visqueux* (Sp.+4), *virtuelle(s)* (Sp.+3), et les locutions *en chaque point, en équilibre* (Sp.+3).

De nombreux autres termes, qualifiés de «banals» par le logiciel d'analyse, sont attestés dans le texte 4. Ainsi, on a relevé les substantifs ou syntagmes substantivaux *analyse fonctionnelle, analyse numérique, caractère opératoire, conditions initiales, couche limite, écoulement(s) de fluides, efforts de liaison, efforts intérieurs, énoncés fondamentaux, équations aux dérivées partielles, équations générales, état local, évolutions adiabatiques, évolutions isothermes, fluides classiques, fluides visqueux, l'instant T, liaisons internes, loi fondamentale, lois de conservation, matériaux nouveaux, milieu déformable, milieux curvilignes, mouvement virtuel, petits mouvements, nature physique, ondes de choc, propriétés mécaniques, propriétés physiques, quantité de mouvement, référentiels, repos, signification physique, structures élastiques, symétrie, systèmes dynamiques, taux, tenseur des contraintes, théorèmes généraux, théorie classique, turbulence, unité conceptuelle, viscosité,* et, avec une référence précise au domaine de spécialisation : *dynamique des fluides, génie chimique, génie civil, géométrie euclidienne, mécanique classique, mécanique des matériaux, résistance des matériaux, thermodynamique.* Nous avons relevé également des verbes et locutions verbales, comme *ce qu'on appelle, on peut dire, nous supposerons,* et des locutions, comme *à chaque instant, en ce sens que, en contact, en équilibre, en la matière, par rapport (à).*

Quelques unités seulement, attestées dans les trois autres sous-corpus, sont absentes du texte 4, parmi lesquelles les formes *gouvernement, préoccupations, recommandations,* les formes verbales *accepter, appartiennent, semble-t-il* et les locutions *au premier rang (de), bien sûr* et *tel ou tel.*

À l'issue de cette analyse, une première conclusion s'impose. Le traitement de nos textes par le logiciel de Saint-Cloud a permis une récolte fructueuse d'unités désignationnelles complexes, de syntagmes dénominatifs (v. Quemada 1978 : 1182). C'est très nettement le texte 4 qui, en l'occurrence, est le plus riche de ce point de vue. Vient ensuite le texte 2. Les segments de textes répétés spécifiques d'un sous-corpus particulier sont apparemment plus rares dans les textes 1 et 3.

3.2.2. La marque de la première et de la deuxième personnes d'après l'utilisation des pronoms personnels et adjectifs possessifs

Le *Tableau 3* présente la distribution fréquentielle des pronoms personnels et adjectifs possessifs de la première personne dans notre corpus. Il indique aussi les coefficients de spécificité enregistrés.

Tableau 3.
Pronoms personnels et adjectifs possessifs (1re personne)

unité	texte 1	texte 2	texte 3	texte 4
j' (18)	1	1 **(Sp. – 3)**	6	10
je (51)	16 **(Sp.+6)**	2 **(Sp. – 7)**	24 **(Sp.+8)**	9 **(Sp. – 5)**
m' (14)	1	1	3	9
ma (6)	1	1	2	2
me (17)	3	1	8 **(Sp.+3)**	5
mes (12)	5 **(Sp.+3)**	0 **(Sp. – 3)**	2	5
moi (6)	2	0	3	1
mon (6)	1	0	3	2
nos (82)	10	43 **(Sp.+4)**	26 **(Sp.+4)**	3 **(Sp. – 16)**
notre (82)	28 **(Sp.+10)**	26	19	9 **(Sp. – 10)**
nous (150)	27 **(Sp.+4)**	29 **(Sp. – 4)**	35 **(Sp.+3)**	59

Les unités *m', ma, moi* et *mon* sont, dans ces textes, d'un emploi toujours «banal». Cependant, alors que le texte 2 offre à la fois des indices négatifs et des indices positifs, les textes 1 et 3 présentent des indices de spécificité positifs. Seul le texte 4, qui atteste de nombreux emplois «banals» des formes de la première personne, ne propose par ailleurs que des indices de spécificité négatifs. ***Je (ou j')***, qui est sous-employé dans les textes 2 et 4, est sur-employé dans les textes 1 et 3. *Nous* est sous-employé dans le texte 2, et sur-employé dans les textes 1 et 3.

On n'est pas surpris d'une telle répartition : les exposés technoscientifiques, comme ceux du sous-corpus 4, tendent, en principe, à la neutralité et à l'objectivité. Rappelons

que les sous-corpus 1 et 3 émanent respectivement d'une communication à un colloque, et d'une communication lors d'une séance solennelle sous la Coupole, textes qui correspondent certainement à un plus grand engagement personnel de l'auteur que lorsqu'il expose un rapport sur la langue française à l'Hôtel Matignon ou lorsqu'il fait un cours de mécanique.

On note, à la fin de cette étude, que les deux textes dont le vocabulaire est le plus «étendu», s'apparentent aussi par la rareté des termes de spécialité, et par un usage commun des marques de la première personne.

Pour caractériser des textes de spécialité, il semble judicieux d'avoir recours à une méthodologie consistant à repérer l'usage répété de syntagmes et locutions complexes. Une telle démarche est plus probante qu'une évaluation globale du vocabulaire, qui tend à démontrer qu'un texte ne comporte pas un vocabulaire très étendu. Nous avions classé initialement nos textes dans un ordre de spécialisation croissante allant de 1 à 4. La vérification de cet ordre intuitif, au moyen de l'étude «macroscopique» du corpus, nous a fait corriger les données, et reclasser les sous-corpus dans l'ordre décroissant d'«étendue» ou de «variété» du vocabulaire suivant: 1 - 3 - 4 - 2. Or, comme l'a ensuite montré l'analyse détaillée comparative des sous-corpus, l'étude de l'usage répété de segments, de locutions et de dénominations syntagmatiques semble être une caractéristique particulièrement intéressante dans les textes de spécialité.

Cette étude ne représente bien entendu qu'une expérience limitée. Elle a pourtant permis d'obtenir quelques résultats, et nous espérons que la méthodologie décrite permettra d'en apporter d'autres.

Notes

1. Nous appellerons «corpus», ou «corpus total», l'ensemble textuel sur lequel porte l'étude, et «sous-corpus», «texte» ou «fragment», chacune des parties composant le corpus total.
2. Ce que recouvre cette qualification demeure, bien entendu, assez largement intuitif. Le degré de spécialisation d'un texte est lié au sujet aussi bien qu'au domaine dont relève ce texte. Il dépend, en même temps, du public, plus ou moins large, auquel l'auteur s'adresse. Nous adoptons les définitions de «langue de spécialité» et de «registre de spécialité» rappelées dans Candel 1994.
3. Voici les références précises correspondant à ce corpus :

   P. Germain, «Langue française et mutation du monde», *Le français, langue des sciences et des techniques*, Actes du colloque organisé par l'Extension de l'Université libre de Bruxelles, l'Ambassade de France et le Centre universitaire de Luxembourg, les 21 et 22 mars 1986 à Luxembourg, RTL Édition, 1987, pp. 17-29.
   P. Germain, *Le français dans les sciences et les techniques*, Rapport présenté à l'Hôtel de Matignon (rédigé avec le concours de plusieurs personnalités), Séance plénière du Conseil supérieur de la langue française, 19 juin 1990, 30 p.
   P. Germain, *La science interpellée*, Lecture faite en la séance solennelle du 27 novembre 1989, Académie des Sciences, Institut de France, 19 p.
   P. Germain, *Mécanique*, tome 1 et tome 2, École Polytechnique, Ellipses, 1986, 444 p. et 446 p.
   Pour le dernier texte, seuls les fragments suivants ont été pris en compte :
   tome 1 : *Introduction*, pp. 7-17, *Avant-propos*, pp. 23-24, introductions aux *chap. I* (p. 27), *II* (p. 44), *III* (pp. 81-82), *IV* (pp. 155-156), *V* (pp. 249-252), *VI* (pp. 331-332), *VII* (p. 401) ; tome 2 : introductions aux *chap. VIII* (pp. 11-12), *IX* (pp. 115-116), *X* (pp. 257-258), *XI* (pp. 367-368) ; *Conclusion générale* : pp. 417-419.
4. À l'exclusion, toutefois, des notes et références, des annexes et des bibliographies.
5. Voir la note préliminaire en p. 1 du texte 2.
6. Voir la correction proposée par P. Germain au titre de sa communication, au début de son texte : «Quelques remarques sur le destin de la langue française dans un monde en mutation sous le signe de la science et de la technique».
7. Sur les variétés d'usages, voir notamment Quemada 1978, p. 1147.
8. Comme *de* (2250), *la* (1390), *et* (1209), *des* (1065), *les* (1033), *à* (711), *en* (636), *le* (616), *une* (519)...
9. Voir Muller 1992, pp. 101 et suivantes.
10. Ce dernier point peut être sujet à caution : en effet le texte 4, de loin le plus long, est composé de plusieurs fragments. On a là deux biais possibles dans la comparaison.
11. C'est une lecture intuitive des unités qui a orienté ce partage.

RÉFÉRENCES

CANDEL, D. (1994) : «Sur l'introduction de mots nouveaux dans des contextes spécialisés», *Rhetoric and Stylistics Today*, Nordeuropäische Beiträge aus den Human- und Gesellschaftswissenschaften, Peter Lang, pp. 71-79.

LAFON, P. (1994) : «Relations syntagmatiques, recherche des cooccurrences et segments répétés», *Traitements informatisés de corpus textuels*, CNRS, INaLF, Paris, Didier Érudition, pp. 151-165.

LAFON, P. et A. SALEM (1983) : «L'inventaire des segments répétés d'un texte», *Mots*, n° 6, pp. 161-177.

LEBART, L. et A. SALEM (1988) : *Analyse statistique des données textuelles*, Dunod.

MULLER, C. (1992) : *Principes et méthodes de statistique lexicale*, collection Unichamp, Paris, Éditions Champion, (réimpression de l'édition Hachette de 1977).

QUEMADA, B. (1978) : «Technique et langage», *Histoire des techniques*, Collection La Pléiade, NRF, pp. 1146-1240.

SALEM, A. (1987) : *Pratique des segments répétés. Essai de statistique textuelle*, Paris, Klincksieck.

TOURNIER, M. (1994) : «*Classe* et *masse* dans le discours syndical : formes de surface, problème de fond», ici même.