Laboratoire d'observation de l'usage terminologique

Josette Rebeyrolle, Ludovic Tanguy, Amélie Josselin-Leray
CLLE, UMR 5263 CNRS et Université Toulouse Le Mirail
Maison de la Recherche
5, allées A. Machado 31058 Toulouse cedex

NOTE Ce texte fait suite à une communication de 2007 :
J. Rebeyrolle, D. Bourigault, C. Fabre, A. Josselin-Leray et L. Tanguy (2007). Un laboratoire d'observation de l'usage du vocabulaire recommandé par les instances officielles françaises, *Colloque "Prescriptions en Langue",* Paris.
Le présent texte a été rédigé à la suite sous forme de chapitre d'ouvrage pour un projet de livre sur la prescription, qui n'a jamais été mené à son terme. Le présent texte est finalisé au vu de ses auteurs, mais n'a pas subi de processus de révision.

1 INTRODUCTION

En France, c'est la délégation générale à la langue française et aux langues de France (DGLFLF) qui est chargée de planifier et modeler le changement linguistique dans la société. Plus exactement, le processus d'aménagement linguistique est enclenché par la Commission générale de terminologie et de néologie. Dès leur publication au *Journal officiel*, les termes qu'elle recommande deviennent obligatoires pour les administrations et services de l'État, conformément à l'article 11 du décret du 3 juillet 1996 relatif à l'enrichissement de la langue française. Dans les situations de concurrence linguistique, notamment entre des termes étrangers et des termes français, le but de la commission est d'influencer l'usage dans le sens des termes français. Une fois les prescriptions publiées, reste à évaluer leur influence sur l'usage des locuteurs. C'est à cette tâche qu'entend contribuer le travail présenté ici.

Plus précisément, nous nous attacherons à proposer une méthode permettant d'apprécier l'enracinement dans l'usage des termes proposés par les instances officielles et à fournir des outils permettant de mesurer les effets de ces prescriptions notamment sur les pratiques langagières des services de l'État qui, comme on vient de le dire, sont contraints d'employer les termes recommandés publiés au *Journal officiel*. Pour ce faire, nous nous intéressons spécifiquement aux différents sites Web des institutions, en tant que medium aisément accessible et exploitable de leurs productions, qu'ils s'agissent de textes officiels ou de communications à visée plus générale. La méthode prend appui sur une étude antérieure, conduite en 2001, qui rend possible l'observation de la variation des usages dans le temps.

Rappelons qu'en 2001, à la demande de la Délégation Générale à la Langue Française, devenue, cette année là, Délégation Générale à la Langue Française et aux Langues de France (DGLFLF), une étude de l'implantation des termes recommandés dans le domaine de l'économie et des finances a été conduite par D. Bourigault et L. Tanguy[1]. L'objectif de cette étude était de déterminer, dans ce domaine, le degré d'implantation des termes proposés par la commission spécialisée de terminologie et de néologie en matière économique et financière. L'observation de l'usage portait sur un domaine restreint abordé

[1] Pour une comparaison avec le domaine de l'informatique, on se reportera au travail de D. Candel (2005).

via les termes qui font l'objet d'un article produit par la commission spécialisée dans ce domaine[2].

En suivant les mêmes principes méthodologiques, nous avons proposé une nouvelle mesure de l'implantation en 2007. Pour présenter ces principes et les résultats qu'ils ont produits, nous procèderons en quatre temps. D'abord, nous exposerons les principes de la méthode que nous avons élaborée (§2). Ensuite, nous indiquerons les grandes lignes et les principaux résultats de la première étude que nous avons conduite sur la terminologie du domaine de l'économie et des finances (§3). Puis, nous comparerons les effets de la prescription dans le temps en observant l'implantation des termes de ce domaine six ans plus tard (§4). L'originalité de notre contribution réside dans le fait qu'elle bénéficie des résultats d'une enquête antérieure lui donnant les moyens de défricher le terrain de l'analyse diachronique pour éclairer l'évolution de l'emploi des termes.

2 METHODOLOGIE

La méthode que nous avons utilisée lors des deux expériences, la première en 2001 et la seconde en 2007, consiste en une analyse automatisée d'un ensemble de documents issus du Web, afin d'y mesurer la présence des termes visés.

Dans ses grandes lignes, la méthode partage les principes proposés par Quirion (2003a). Partant du constat que les premiers travaux ayant porté sur l'implantation terminologique (cf. Depecker, L. et Mamavi, G. (eds). (1997) ; n°12 de la revue *Terminologies nouvelles*, 1994) se fondent davantage sur des impressions que sur des données fiables (sur des estimations approximatives de l'implantation), l'auteur propose une méthode permettant de construire des données robustes pour mesurer scientifiquement l'implantation terminologique. Considérant que les données dont on dispose sont des données impressionnistes qui "ne constituent pas un appui suffisamment solide pour déterminer le succès relatif d'une implantation" (Quirion, *op. cit.*, p.22), Quirion fournit des outils pour mieux quantifier les observations.

Bien que notre proposition s'inscrive à la suite des travaux de Quirion, les visées de notre propre travail sont plus modestes notamment parce que nous ne cherchons pas à déceler les raisons des succès ou des échecs des recommandations officielles. Notre objectif est d'abord de contribuer à l'élaboration d'une méthode permettant de mesurer l'implantation de terminologies. Contrairement à Quirion, nous suivons les principes méthodologiques de la linguistique de corpus et rejetons la technique de l'échantillon qu'il préconise (2003a). Comme Bowker et Pearson (2002, p.49), nous pensons en effet que "If you decide to choose an extract at random, you might accidentally eliminate a part of text that could be very interesting for your study. Therefore, it is a good idea to use full texts when compiling LSP[3] corpora."

2.1 Utilisation du Web comme corpus

Notre méthode se concentre exclusivement sur des données issues du Web. En ce sens, elle s'inscrit dans une mouvance croissante en linguistique et en traitement automatique des langues, qui utilise cette masse de données électronique facilement accessible, mais qui soulève un ensemble de questions pratiques et méthodologiques.

---

[2] Ces fiches sont disponibles sur la base FRANCETERM de la DGLFLF : `http://franceterme.culture.fr`

[3] *Language for Specific Purposes*, i.e. langue de spécialité

Plus précisément, notre approche concerne ce qu'on appelle désormais "*Web **for** corpus*" (i.e. en utilisant le Web pour construire un corpus traditionnel en visant des sites particuliers), par opposition à une utilisation globale du Web par le biais de moteurs de recherche généralistes ou spécialisés ("*Web **as** corpus*"), comme présenté par Hundt et al. (2007). En effet, puisque nous nous intéressons à un domaine spécialisé (l'économie et les finances), en posant une question précise sur l'implantation des termes par des institutions officielles, nous avons dû sélectionner un ensemble de sites particuliers, ce qui nous permet d'obtenir des résultats contrastés en fonction de la nature (publique vs non-publique) de ceux-ci.

La technique employée consiste donc à identifier une liste de sites Web (ou de parties de sites), et d'en aspirer automatiquement le contenu des pages afin de constituer un corpus numérique de plusieurs millions de mots. Les pages ainsi obtenues sont dépouillées de leur contenu non-textuel, et normalisées pour permettre une analyse par des outils automatiques.

Bien qu'il soit très facile par cette méthode d'obtenir de grandes quantités de données, un ensemble de considérations spécifiques aux données du Web nécessitent des précisions. Tout d'abord, les données issues du Web ne sont pas aisément identifiables, en ce sens que leur auteur véritable, leur date et leur contexte de création sont dans le cas général inconnus. Nous nous contenterons donc dans ce qui suit de ne considérer, pour les différents textes de notre corpus, que le site dont ils sont issus, sans distinguer notamment les différents types ou genres dont ils relèvent. Un autre problème, spécifique à notre approche diachronique, provient du fait qu'il n'est pas possible de dater un texte issu du Web : la seule indication chronologique accessible est la date de modification du texte, qui peut traduire différents états de fait (modification mineure, voire même simple restructuration du site sans changement de contenu). De même, la possibilité de données dupliquées dans un tel corpus n'est pas à écarter : les textes électroniques ont une tendance naturelle à se recopier, que ce soit au sein d'un même site ou d'un site à l'autre. L'adresse électronique (URL) d'une page Web n'est pas un identifiant permettant de garantir son unicité ni son origine. Un exemple de cette situation complexe est le cas où un site institutionnel inclut une revue de presse, avec des extraits d'articles de journaux. Avec notre méthode, ces textes sont quand même considérés comme relevant de l'institution en question.

Toutefois, ces différents problèmes sont minoritaires et sont en grande partie résolus par une sélection semi-manuelle des parties de sites. Enfin, quelques obstacles techniques ont limité notre corpus. Certains sites ou parties de site ne sont pas accessibles par des moyens automatiques, pour différentes raisons : protection des données contre une aspiration automatique, commercialisation de certains textes (notamment pour les journaux), formats de fichiers difficilement exploitables. Par ailleurs, certains sites hébergent des pages écrites dans une langue autre que le français, qui ont dû être détectées par des méthodes automatiques spécifiques.

Quoiqu'il en soit, pour ce type d'étude, l'utilisation du Web présente de nombreux avantages qui l'emportent sur ces limitations : les données accessibles sont variées, récentes, facilement exploitables, et permettent des volumes de données que tout autre étude sur corpus ne permettrait pas d'atteindre dans un temps réduit et pour un coût raisonnable.

Nous avons ainsi pu constituer, pour chacune des deux expériences décrites par la suite, des corpus spécialisés de plusieurs dizaines de millions de mots.

2.2 Termes étudiés

Les termes visés par cette étude sont ceux qui ont été publiés au Journal Officiel par la Commission Générale de Terminologie et de Néologie de la DGLFLF, et accessibles directement sur le site FRANCETERME (http://franceterme.culture.fr). Cette base de données est organisée en différentes thématiques, dont celles qui nous concernent : l'économie et les finances. Les termes sont présentés sous forme de fiches dont un exemple est reproduit dans la figure 1.

***Journal officiel* du 12/05/2000**

aide de caisse

Domaine : Économie et gestion d'entreprise
Définition : Employé d'un magasin en libre-service chargé d'assister la clientèle aux caisses de sortie.
Note : Le terme « *caddie-boy* » est impropre.
Équivalent étranger : bag-boy (en), bagger (en), bag-girl (en), bagman (en)

Figure 1: Exemple de fiche terminologique de la base FranceTerme

Les fiches des domaines concernés ont été automatiquement téléchargées et analysées. Comme on le voit dans l'exemple ci-dessous, en plus du terme-vedette (terme recommandé), apparaissent la date de sa publication au journal officiel, les équivalents étrangers (avec une identification de la langue d'origine, dans nos domaines il s'agit exclusivement de l'anglais) et le cas échéant des termes impropres.

Au final, les données de départ dont nous disposons sont : 637 termes dont 307 recommandés, 313 équivalents étrangers et 17 impropres. 25% des termes sont des termes publiés en 1987, et plus de 35% des termes publiés en 1998 et 2000. La plupart de ces termes sont complexes (442 sur 637).

Les termes les plus polysémiques ont été exclus de la recherche, comme par exemple "accord", "noyau", "écart", "rattachement", "division", "notation", "épreuve", "manager[4]", parce que leur apparition dans un texte n'indique par nécessairement qu'il s'agit de leur acception précise dans le domaine économique et financier.

2.3 Recherche des termes

Techniquement, les traitements automatiques déployés sur le corpus pour la recherche des termes sont très limités par rapport à ce qu'il est possible d'envisager pour ce genre de tâche.

Si une analyse syntaxique complète du corpus aurait permis d'isoler tous les syntagmes et de comparer ceux-ci à notre liste, ce n'est dans la pratique ni nécessaire ni efficace. Tout d'abord, les termes complexes de la liste ne nous intéressent que s'ils apparaissent sous leur forme complète (on s'intéresse à *détente fiscale* et non à *détente*) et sans variations. Ensuite, une grande partie des termes de la liste correspondent à des mots étrangers, des néologismes, ou des termes techniques et sont donc difficiles à traiter par un analyseur automatique qui s'appuie sur un lexique généraliste. Même une simple opération de lemmatisation pose des problèmes pour les cas courants où la flexion implique une variation d'acception de certains termes ou éléments de termes (comme *affaires* dans

[4] La forme verbale étant recommandée mais le nom issu de l'anglais proscrit, on choisit d'exclure cette forme pour ne pas fournir de chiffres erronés.

"centre d'affaires" ou "femme d'affaires"), et conduirait au repérage de termes autres que ceux de notre liste cible.

Au final, le seul traitement linguistique automatique appliqué sur le corpus a été une segmentation en mots, nécessaire pour assurer l'efficacité de la recherche dans des textes dont la mise en page peut être complexe.

Le résultat de cette étape de recherche est représenté de la façon suivante : pour chaque sous-corpus (i.e. site Web) considéré, et pour chaque terme de la liste (trois catégories), le nombre d'occurrences de ce terme pour ce site est calculé.

Comme on le verra par la suite, les fréquences d'apparition sont très variables, et certains termes n'ont été repérés sur aucun des sites examinés.

3 USAGE DES TERMES ECONOMIQUES ET FINANCIERS EN 2001

L'étude de l'implantation des termes recommandés dans le domaine de l'économie et des finances réalisée en réponse à un appel d'offres de la DGLFLF en 2001 fixe le premier repère rendant possible quelques années plus tard une nouvelle mesure de l'implantation de ces termes.

3.1 Description du corpus

Les textes sélectionnés en 2001 sont extraits d'une diversité de sites Web. Selon les principes habituellement utilisés dans les travaux consacrés à l'implantation terminologique (cf. Auger, *in* Quirion, 2007), on distingue deux types de documents : ceux qui relèvent de l'administration publique et ceux qui relèvent d'institutions sociales (médias spécialisés, médias généraux, enseignement, etc.). Parmi les sites publics ou parapublics, sont concernés le portail de l'administration française (*Legifrance*), La Commission des opérations de bourse (COB), Le Conseil des marchés financiers (CMF), La Banque de France, Le Sénat, L'Assemblée nationale, Le Ministère des finances, Le Conseil constitutionnel, La Cour des comptes. Parmi les sites ne relevant pas d'organismes publics mais ayant un lien fort avec le domaine de l'économie et des finances, on compte des articles du quotidien spécialisé *Les Échos*, des articles financiers et économiques du quotidien régional *Ouest France*, des articles du quotidien économique *L'expansion*, des textes produits par des écoles de commerce, comme l'ESSEC ou HEC, des documents accessibles sur les sites de chambres de commerce et de l'industrie (CCI) de grandes villes de France.

Sachant que, pour chaque site, nous avons veillé à sélectionner uniquement les documents qui relèvent indiscutablement du domaine de l'économie et des finances, le corpus constitué compte, au final, plus de 23 millions de mots. Un tel filtrage ne peut se faire que par une longue étape de sélection manuelle des parties de l'arborescence des sites internet contenant des documents jugés pertinents pour le domaine étudié. Une fois le repérage manuel effectué, l'exploitation automatique a été faite. Les résultats produits sont exposés dans la section suivante.

3.2 Types de termes en usage en 2001

Le premier résultat obtenu lors de cette étude concerne la simple présence/absence de chacun des termes recherchés dans le corpus. Le tableau suivant rappelle le nombre de termes effectivement envisagés pour chaque catégorie, et indique le nombre de termes trouvés au moins une fois dans le corpus de 2001.

	Termes publiés au *Journal*	**Termes publiés ayant au**	**Termes publiés absents du**

	officiel	moins une occurrence dans le corpus de 2001		corpus de 2001	
Termes recommandés	307	161	(52%)	146	(48%)
Équivalents étrangers	313	86	(27%)	227	(73%)
Termes impropres	17	8	(47%)	9	(53%)
Total	637	255	(40%)	382	(60%)

Tableau 1 : nombre de termes repérés dans l'étude de 2001, par catégorie

Ce tableau montre que parmi les termes recommandés un peu plus de la moitié seulement est attestée dans le corpus sélectionné (52%). Il montre également que 27% seulement des termes étrangers présentent au moins une occurrence dans le corpus. Enfin, il fait apparaître le fait que nous n'avons pu observer que seulement 47% des formes considérées comme impropres par la commission spécialisée dans le domaine de l'économie et des finances. Le faible nombre de termes effectivement utilisés dans les textes extraits de la Toile peut s'expliquer par une diversité de facteurs.

Tout d'abord, le silence observé peut être expliqué par nos choix méthodologiques et techniques. La couverture de notre corpus est bien entendu limitée, et certains termes peuvent ne pas avoir été repérés, notamment à cause de variantes de surface (légitimes ou dues à des erreurs) ou de problèmes techniques.

Mais cette faible présence des termes étudiés peut également être expliquée par le manque d'adéquation entre la liste et le corpus. En effet, certains termes apparaissent dans la liste de référence comme relevant du domaine économique et financier au sens large, alors que notre corpus se concentre sur un discours plus technique. Il s'agit par exemple des termes *superette, magasin d'usine, démarchage téléphonique* ou encore *épinglette* ou *jardinerie*, dont les emplois se situent apriori dans des textes à visée plus générale. Il est donc important d'examiner plus précisément les différences entre les sous-corpus.

Ces premiers résultats bruts ne doivent pas être interprétés comme une preuve une faiblesse du processus d'implantation des termes recommandés ni une non-utilisation des termes étrangers.

3.3 Distribution des termes dans les textes de 2001

Afin de pouvoir comparer la distribution des occurrences des termes recommandés, étrangers et impropres dans chacun des sites examinés, nous avons procédé à un calcul du nombre d'occurrences normalisé pour chaque site pour 100 000 mots.

Corpus	**Termes recommandés**	**Termes étrangers**	**Termes impropres**
Sites publics			
Légifrance	109	0	0
COB	179	14	1
CMF	416	27	0
Banque de France	320	14	1
Sénat	153	5	2
Assemblée Nationale	155	2	3
Ministère des Finances	145	8	2
Conseil constitutionnel	90	0	4
Cour des comptes	99	0	0
Total sites publics	152	5	2
Sites non publics			
Les échos	272	191	1
L'expansion	162	38	13
Ouest-France	160	28	1
École de commerce	242	363	0
CCI	234	88	0
Total sites non puplics	255	166	2

Tableau 2 : nombre d'occurrences normalisées pour chaque site pour 100 000 mots

La figure ci-dessous réunit les résultats pour les sites publics et non publics. Nous examinerons toutefois successivement les deux sous-corpus pour voir comment se distribuent les trois types de termes dans les documents que nous avons extraits en 2001.

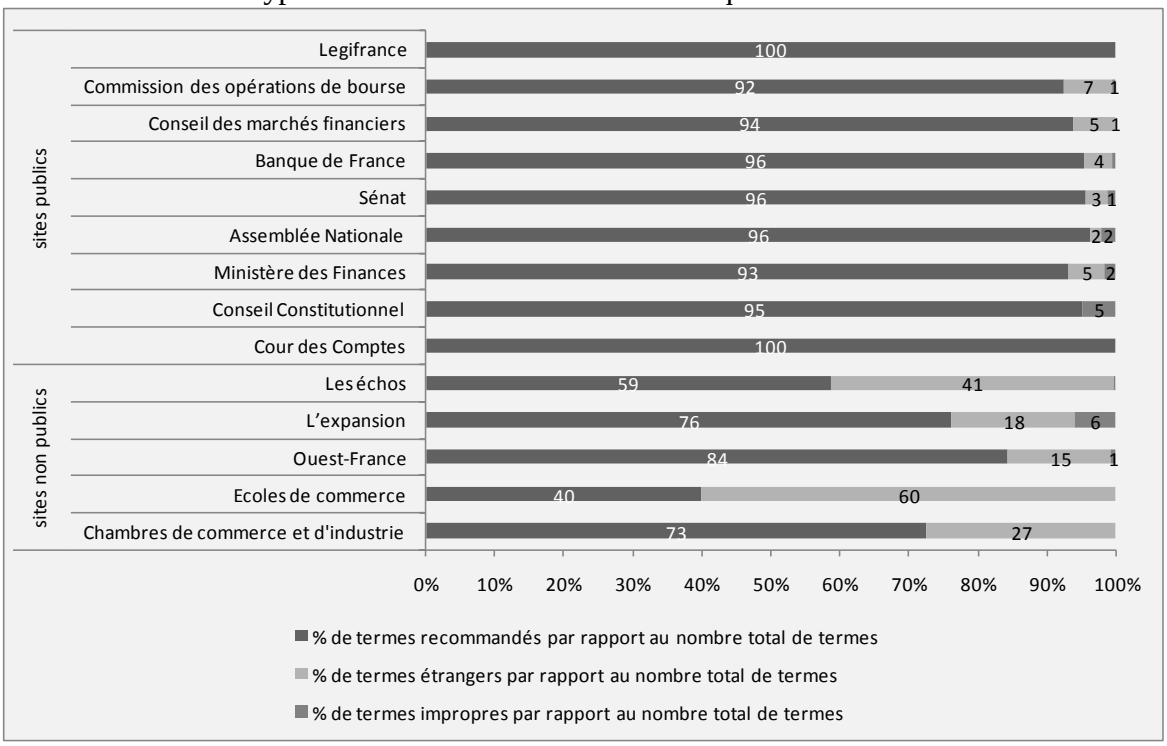

Figure 2 : distribution des trois types de formes (recommandées, étrangères et impropres) dans les documents de l'ensemble des sites sélectionnés en 2001

Dans les documents des sites publics, il est remarquable de noter l'usage quasi exclusif de la terminologie officielle (plus de 90% des termes employés dans les documents aspirés sur ces sites sont des termes recommandés). Contrairement à ce qu'on attendait, le pourcentage de termes étrangers n'est pas nul, sauf toutefois dans les documents accessibles sur le portail de l'administration française et sur le site de la Cour des comptes qui emploient, quant à eux, exclusivement des termes recommandés (100% des termes employés sur ces sites sont des termes recommandés). S'agissant des termes impropres, on constate qu'ils sont très peu employés quel que soit le site.

Si l'on observe ensuite la part des différents types de termes, dans les documents des sites non publics, on observe que les termes recommandés dominent partout, à l'exception des sites des écoles de commerce et du site du journal *Les échos* qui ont une part non négligeable de termes étrangers, respectivement 60% et 41%. C'est également le cas du site des CCI, même si c'est dans une moindre mesure (27%). Les termes impropres se trouvent quant à eux principalement dans l'hebdomadaire *L'expansion*.

Ces résultats montrent très clairement que les recommandations officielles sont suivies uniquement par ceux qui sont contraints de le faire.

Le graphique suivant propose une synthèse des résultats de l'étude d'implantation conduite en 2001 :

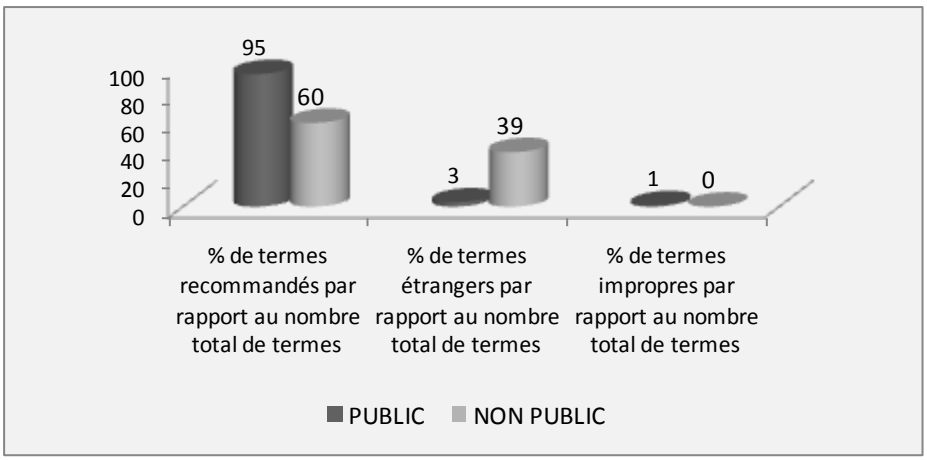

Figure 3 : Synthèse des résultats de l'implantation en 2001

De cette figure, on peut retenir trois éléments essentiels : 1) l'implantation des termes recommandés se fait majoritairement dans les documents des organismes publics ; 2) ailleurs, la concurrence des termes étrangers est très forte ; 3) l'emploi de termes impropres est extrêmement faible quels que soient les locuteurs. Dans les sites publics, on ne peut pas parler de concurrence entre formes prescrites et formes proscrites, alors que dans les sites non publics cette concurrence est sévère. Elle ne vient toutefois pas des termes impropres mais bien des termes étrangers dont l'usage persiste malgré les recommandations des instances de normalisation.

Au terme de ce premier examen, on peut se demander ce que deviennent, quelques années plus tard, les usages indésirables et, plus précisément, si les termes proposés pour remplacer les termes étrangers et les termes dits "impropres" sont utilisés notamment par ceux qui ne sont pas soumis à l'obligation de les employer. Autrement dit, les termes adoptés par les instances officielles finissent-ils par "passer" auprès des locuteurs ordinaires ? Autant de questions auxquelles une comparaison sur un nouveau corpus quelques années plus tard devrait permettre d'apporter une réponse.

4. USAGE DES TERMES ÉCONOMIQUES ET FINANCIERS EN 2007

Reconduire une étude six ans plus tard a d'abord pour but de répondre à l'un des objectifs fixé par l'étude de 2001 qui était de mettre en place une plate-forme automatique permettant de reproduire l'analyse afin d'obtenir une estimation de l'évolution de l'implantation des termes d'un même domaine. Notre objectif est notamment de voir si, parmi les termes recommandés qui n'apparaissaient pas dans le corpus en 2001, certains sont désormais employés et si ceux qui étaient faiblement employés se sont plus largement diffusés. On a vu en effet qu'en 2001 si la terminologie officielle est largement utilisée dans les sites publics, elle ne l'est pas aussi massivement ailleurs. On souhaite également établir si les termes que les instances voulaient bannir sont définitivement sortis de l'usage ou non.

4.1 Description du corpus

Mais, pour reconduire l'étude et répondre à ces questions, nous avons été confrontés à une difficulté. Une comparaison diachronique exige une même base empirique : même liste de termes et même corpus. Si la liste de termes ne pose aucune difficulté (nous n'avons pas, pour cette étude, pris en compte les termes apparus dans les recommandations de la

DGLFLF après 2001), il n'en va pas de même pour le corpus. En effet, plusieurs raisons nous ont empêchés de constituer un corpus parfaitement identique à celui de 2001.
Première raison : le monde change, et le Web traduit ces changements. Depuis 2001, certaines institutions ont ainsi disparu ou se sont réorganisées ; c'est le cas de la Commission des opérations de bourse et du Conseil des marchés financiers, qui ont fusionné. D'autres institutions se sont par contre développées, comme le Conseil d'analyse économique (CAE) du premier ministre.
Deuxième raison : le Web se développe et se démocratise. Ainsi, de nouveaux sites pertinents pour notre étude apparaissent. C'est le cas du site "service public", portail de l'administration française qui a vocation à s'adresser au grand public, à la différence du site Légifrance, qui répertorie les textes légaux. De même, certains sites se sont internationalisés, et possèdent une très grande partie de leur contenu rédigé en anglais, ce qui rend leur exploitation bien plus difficile, et risquerait de donner des résultats erronés.
Troisième raison : le Web se commercialise et son accès est protégé. Les techniques d'utilisation massive du Web se sont beaucoup développées en six ans, que ce soit à des fins scientifiques, ou commerciales. Ainsi, certains sites bloquent désormais l'aspiration automatique, afin de préserver leur accès réseau et/ou d'empêcher l'exploitation de leurs données. C'est notamment le cas des journaux en ligne, qui ont mis en place des accès payants à leurs archives. Nous avons toutefois pu conserver un accès privilégié à l'hebdomadaire économique *L'expansion*.
Pour ces différentes raisons, certains sites étudiés en 2001 ne font plus partie de notre second corpus, et de nouveaux sites y ont été ajoutés. Au-delà des trois raisons citées ci-dessus, nous avons également choisi de ne pas reprendre exactement les mêmes sites afin de permettre une vision diachronique. Le portail Légifrance n'a ainsi pas été pris en compte puisqu'il aurait conduit à une très grande redondance avec les textes de 2001, le contrôle des dates de publication étant techniquement très difficile, même pour les textes légaux.
Au final, notre nouveau corpus est comparable à celui de 2001, mais avec un ensemble de précautions. Sa taille est légèrement supérieure (30 millions de mots au lieu de 23), mais sa répartition est conforme à celle de la première étude.

4.2 Distribution des termes dans les textes de 2007

Avant d'examiner en détail les résultats de l'étude conduite six ans après celle que nous avons présentée dans la section 3.1, nous commencerons par quelques observations générales visant à donner un premier aperçu de la situation en 2007. D'abord, on est frappé de voir que toutes les formes publiées au *Journal officiel* ne sont toujours pas implantées en 2007. Les formes recommandées qui n'ont aucune occurrence dans le corpus sont donc des formes pour lesquelles on peut considérer que la prescription a échoué.

	Termes publiés au *Journal officiel*	**Termes publiés ayant au moins une occurrence dans le corpus de 2007**		**Termes publiés absents du corpus de 2007**	
Termes recommandés	307	112	(37%)	195	(63%)
Termes étrangers	313	99	(32%)	214	(68%)
Termes impropres	17	3	(18%)	14	(82%)
Total	637	214	(34%)	423	(66%)

Tableau 3 : nombre et types de termes étudiés dans l'étude de 2007

On retrouve un des principaux résultats de l'étude conduite en 2001 puisque dans les sites non publics la forte concurrence des termes étrangers persiste en 2007. En effet, comme en

2001, les concurrents des formes recommandées sont toujours principalement les termes étrangers qu'elles ont portant vocation à remplacer.
Comme pour la présentation des résultats de 2001, notre examen se fait en deux temps : on examine d'abord l'usage des termes dans les sites publics puis dans les sites non publics.

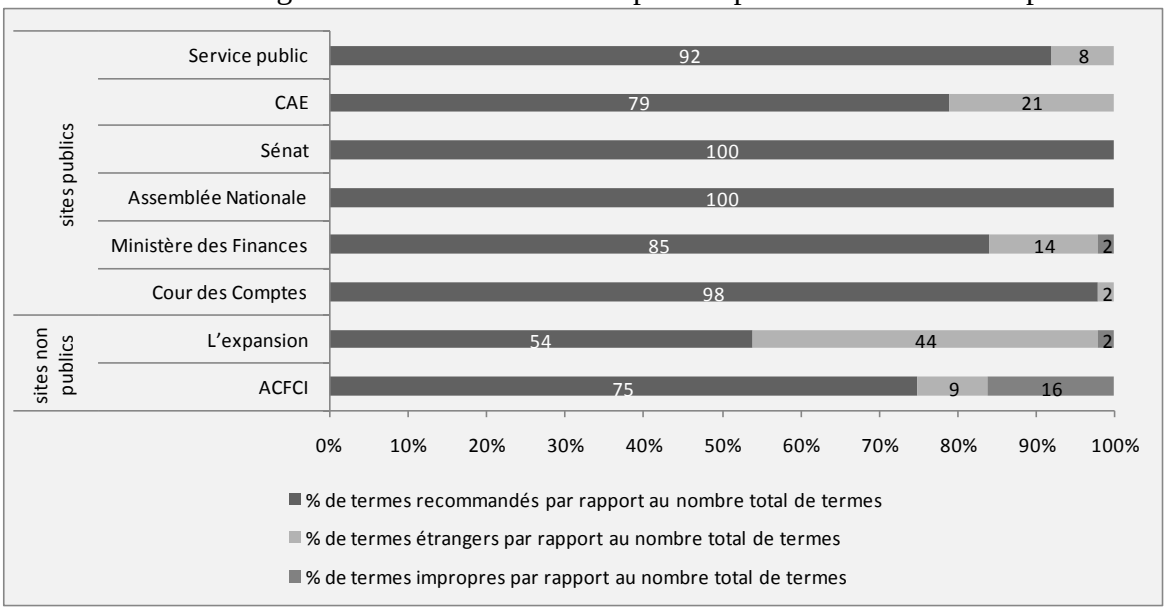

Figure 4 : distribution des trois types de formes (recommandées, étrangères et impropres) dans les documents de l'ensemble des sites sélectionnés en 2007

Si l'on considère la distribution des formes telle qu'elle apparaît dans le corpus de textes extraits de sites publics en 2007, on constate d'abord que les documents examinés contiennent toujours exclusivement ou quasi exclusivement des termes recommandés et ce quels que soient les sites du gouvernement français d'où ils sont issus.
Dans les deux sites non publics que nous avons explorés en 2007, on observe deux tendances distinctes. Dans le journal *L'expansion*, on constate que les termes recommandés sont très fortement concurrencés par les termes étrangers. En revanche, dans les documents récoltés sur le site des Chambres consulaires, la concurrence est moins forte et elle est surtout le fait des termes impropres. Contrairement à ce qu'on observait en 2001, les termes impropres sont donc moins le fait des articles de journaux que des chambres de commerce.

Finalement, comme en 2001, l'implantation des termes recommandés se fait principalement dans les documents des organismes publics. Dans les autres sites examinés, la concurrence des termes étrangers reste très forte. On peut en conclure que les administrations publiques ont réussi à chasser les termes étrangers des textes qu'elles produisent, alors qu'ailleurs leur usage reste massif.

4.3 Deux études de cas

Après un examen quantitatif des mesures effectuées, nous nous proposons maintenant d'illustrer en considérant deux cas de figure opposés. Nous examinerons d'abord un terme bien implanté et ensuite une forme qui résiste à l'implantation.
Le terme bien implanté que nous avons retenu est la forme composée *zone euro*. Après avoir rappelé que ce terme est paru au *Journal officiel* le 14 septembre 1999 pour remplacer ces nombreux concurrents étrangers : *single currency area, euro area, euro zone* ainsi que les termes impropres en usage : *euroland* et *eurolande*, nous allons observer

la distribution de ces six formes dans les deux corpus considérés, celui de 2001 et celui de 2007 :

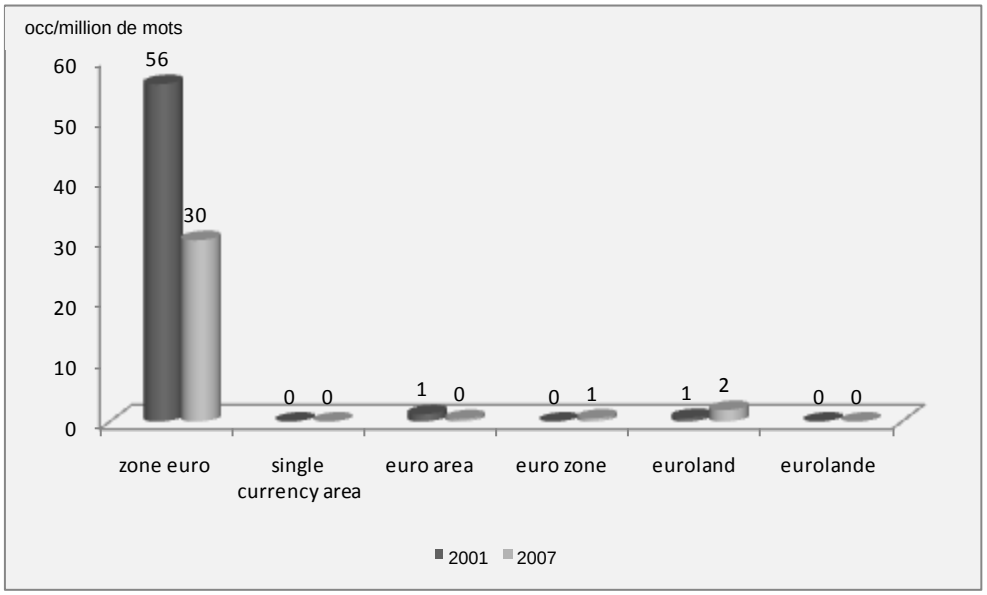

Figure 5 : distribution comparée des concurrents de la forme recommandée zone euro *dans les deux corpus étudiés*

On constate que le concurrent du terme bien implanté *zone euro* a changé : en 2001, le terme est concurrencé par la forme anglaise *euro area* ; alors qu'en 2007, c'est la forme impropre *euroland* qui lui fait de l'ombre. Il s'agit toutefois d'une concurrence faible puisque, dans les deux périodes, l'usage de la *zone euro* domine très clairement. La comparaison nous permet de rappeler que l'usage évolue et que cette évolution affecte de la même façon les termes recommandés et ceux qui sont proscrits, comme *euroland*.

Examinons à présent le cas d'une forme qui ne s'implante pas. Il s'agit de la forme *veille économique* publiée au *Journal officiel* daté du 14 aout 1998 et proposée pour remplacer les termes *business intelligence*, son équivalent anglais, et *intelligence économique*, terme français jugé impropre. La distribution de ces trois formes est fournie dans la figure suivante :

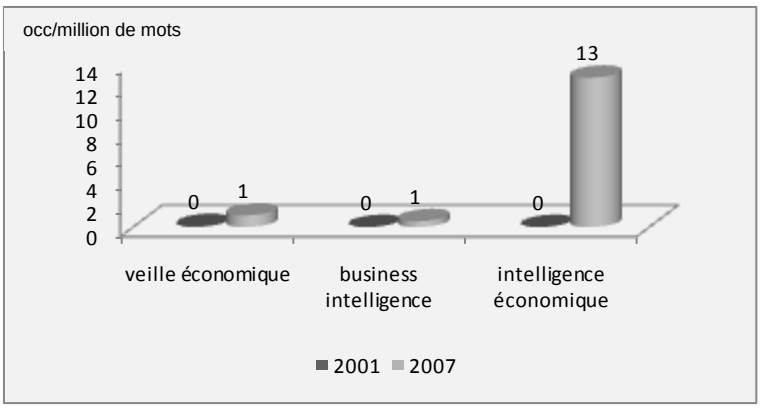

Figure 6 : distribution comparée des concurrents de la forme recommandée veille économique *dans les deux corpus étudiés*

Depuis 2001, la situation a très sensiblement évolué puisque d'un emploi quasi nul des termes utilisés pour désigner le concept de "veille économique", on est passé, en 2007, à un usage effectif. Cependant, le concept n'est pas dénommé par le terme préconisé par la

commission spécialisée dans le domaine de l'économie et des finances, c'est même au contraire la forme dont l'usage est reprouvé qui s'impose.
Quand on examine plus précisément la distribution des trois formes dans les sites du corpus, on constate que, dans les sites où elles sont employées, les usages varient considérablement. Seul le site de la Cours des comptes emploie uniquement le terme recommandé. L'autre site public où l'on trouve cette forme est le site du Ministère des finances, mais là, elle est très fortement concurrencée par l'anglicisme réprouvé par les instances officielles. Dans les deux sites non publics, c'est également cette forme qui domine. Dans le journal *L'expansion*, elle cohabite toujours avec la forme anglaise, mais on peut penser que celle-ci va la remplacer, comme c'est déjà le cas dans le site de l'ACFCI.

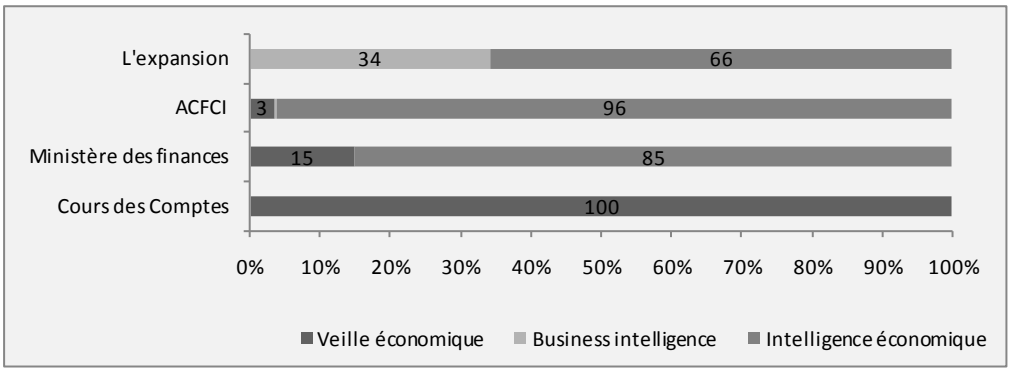

Figure 7 : distribution des formes veille économique, business intelligence, intelligence économique *dans les sites où elles sont employées*

Finalement, le terme qui s'est imposé dans l'usage pour remplacer *business intelligence* est donc bien la forme calquée sur l'anglais.
L'étude détaillée de ces termes concurrents est cependant à prendre avec précautions, puisque les effectifs sont très faibles (quelques dizaines d'occurrences). Par contre, ces tendances peuvent aisément être confirmées par d'autres moyens. On observera notamment que, contrairement à *veille économique*, le terme *intelligence économique* possède une entrée dans l'encyclopédie wikipedia. La plupart des formations supérieures en économie ont dans leurs intitulés de masters le terme *intelligence économique*.

CONCLUSION

Si l'on synthétise les observations quantitatives que nous venons de faire, en 2001 puis en 2007, on retiendra trois éléments principaux. Premièrement, les sites publics et parapublics emploient beaucoup plus massivement les termes recommandés que ne le font les sites qui ne relèvent pas de tels organismes. Pour ce qui concerne les termes étrangers, les sites non publics se distinguent très nettement par le fort, voire très fort, pourcentage d'emploi de ces formes. Quant aux formes considérées comme impropres, elles sont d'une manière générale peu employées. Cependant ce qu'il faut noter c'est que, quand elles sont utilisées, elles le sont à peu près de la même manière quel que soit le site, public ou non. Ces résultats montrent donc que les prescriptions des instances de normalisation semblent avoir un effet sur les pratiques langagières des services de l'état mais qu'elles parviennent moins à influencer, directement ou indirectement, les usages linguistiques de ceux qui n'y sont pas contraints.

À terme, un autre de nos objectifs est de mettre au point une procédure automatique de veille terminologique visant à repérer sur la Toile, et plus précisément sur des sites considérés comme pertinents, des formes françaises ou étrangères utilisées par les locuteurs à la place des formes recommandées et non encore repérées par les instances officielles, autrement dit des formes qui ne sont pas signalées comme étant "impropres". Le repérage de ce type de formes devrait pouvoir bénéficier du repérage des structures syntaxiques partagées (cf. Bourigault, 2002) par les termes recommandés et leurs concurrents (français ou étrangers). Ce repérage repose sur l'hypothèse que les formes qui partagent un même contexte distributionnel sont des concurrents potentiels. De cette manière, il serait envisageable de mettre au jour de nouveaux usages.

REFERENCES

Bourigault D. (2002). Upery : un outil d'analyse distributionnelle étendue pour la construction d'ontologies à partir de corpus, *Actes de la 9ème conférence annuelle sur le Traitement Automatique des Langues* (TALN 2002), Nancy, 2002, pp. 75-84.

Bourigault D. et Tanguy, L. (2001). Etude de l'implantation sur le Web des termes recommandés par la DGLF – Domaine de l'économie et des finances. Rapport d'étude.

Bowker, L. et Pearson, J. (2002). *Working with specialized language. A practical guide to using corpora*. Routledge.

Candel, D. (2005). La néologie d'aménagement: l'expérience française, *Les néologies contemporaines*, (ed.) L. Depecker, Paris, Société Française de Terminologie, coll. Le savoir des mots, p. 55-77.

Depecker, L. et Mamavi, G. (eds). (1997). *La mesure des mots. Cinq études d'implantation terminologique*. Rouen : Université de Rouen.

Hundt, M., Nesselhauf, N and Biewer, C. (eds) (2007). *Corpus Linguistics and the Web*. Rodopi, Amsterdam.

Quirion, J. (2003a). *La mesure de l'implantation terminologique : proposition d'un protocole. Étude terminométrique du domaine des transports au Québec*, Montréal, Office québécois de la langue française, coll. "Langues et sociétés", n°40.

Quirion, J. (2003b). Methodology for the Design of a Standard Research Protocol for Measuring Terminology Usage, *Terminology*, 9, 1, 29-49.

Quirion, J. et Lanthier, J. (2006). Intrinsic qualities favouring term implantation: verifying the axioms. Bowker, L. (ed.). Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer. Ottawa: University of Ottawa Press, pp.107-118.

Quirion, J. et Lanthier, J. (2007). Étude contrastive des principes et des méthodes de la lexicographie et de la terminométrie. M.-C. L'Homme et S. Vandaele (eds). Lexicographie et terminologie : compatibilité des modèles et des méthodes. Presses de l'université d'Ottawa, 219-245.

Sablayrolles, J.-F. (ed.) (2003). *L'innovation lexicale*. Paris : Champion.

Tanguy, L. et Hathout, N. (2003). *Le Français sur Internet : recherche automatique de néologismes*, Multimédia, Internet et Études Françaises 2, Vancouver.

(1994). Implantation des termes officiels. Actes du séminaire (Rouen, décembre 1993). Terminologies nouvelles, n°12.