

## Les noms composés anglais et français du domaine d'Internet : une radiographie bilingue

Céline Ahronian et Henri Béjoint

Volume 53, numéro 3, septembre 2008

URI : <https://id.erudit.org/iderudit/019245ar>

DOI : <https://doi.org/10.7202/019245ar>

[Aller au sommaire du numéro](#)

### Éditeur(s)

Les Presses de l'Université de Montréal

### ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

### Citer cet article

Ahronian, C. & Béjoint, H. (2008). Les noms composés anglais et français du domaine d'Internet : une radiographie bilingue. *Meta*, 53(3), 648–666.  
<https://doi.org/10.7202/019245ar>

### Résumé de l'article

Cet article se propose d'étudier le lexique d'un domaine de spécialité, Internet, qui présente la particularité de n'être pas réservé à un groupe d'experts. Les termes apparaissant quasi exclusivement en langue anglaise, la question centrale est de savoir comment se traduit l'influence de l'anglais sur la terminologie du français, langue cible.

Pour apporter des éléments de réponse, nous avons examiné les noms composés, qui sont particulièrement nombreux dans ce domaine. Après avoir dressé le profil morphosyntaxique des termes collectés dans notre corpus à travers l'étude de leurs structures compositionnelles, y compris la pseudo-confixation, nous avons tracé leur portrait sémantique avec une analyse de la centricité et des relations sémantiques entre leurs éléments. Notre conclusion est que le lexique français d'Internet conserve des traces de l'origine anglaise de nombreux termes mais que le nombre d'emprunts est faible.

# ÉTUDES TERMINOLOGIQUES ET LINGUISTIQUES

---

## Les noms composés anglais et français du domaine d'Internet : une radiographie bilingue

**CÉLINE AHRONIAN**

*Université Lumière Lyon 2, Lyon, France*

*celine.ahronian@univ-lyon2.fr*

**HENRI BÉJOINT**

*Université Lumière Lyon 2, Lyon, France*

*henri.bejoint@univ-lyon2.fr*

### RÉSUMÉ

Cet article se propose d'étudier le lexique d'un domaine de spécialité, Internet, qui présente la particularité de n'être pas réservé à un groupe d'experts. Les termes apparaissant quasi exclusivement en langue anglaise, la question centrale est de savoir comment se traduit l'influence de l'anglais sur la terminologie du français, langue cible.

Pour apporter des éléments de réponse, nous avons examiné les noms composés, qui sont particulièrement nombreux dans ce domaine. Après avoir dressé le profil morphosyntaxique des termes collectés dans notre corpus à travers l'étude de leurs structures compositionnelles, y compris la pseudo-confixation, nous avons tracé leur portrait sémantique avec une analyse de la centricité et des relations sémantiques entre leurs éléments. Notre conclusion est que le lexique français d'Internet conserve des traces de l'origine anglaise de nombreux termes mais que le nombre d'emprunts est faible.

### ABSTRACT

This article examines the lexicon of a specialized field, the Internet, which is characterized by the fact that it is not restricted to a group of experts. Since the terms are coined almost exclusively in English, the main question is: is the terminology of French, the target language; influenced by the original English terms?

We collected the compound nouns in our corpus, because compound nouns are particularly numerous in the field. After drawing up the morphosyntactic profile of the collected terms through the study of their compositional structures, including pseudo-confixation, we sketched their semantic profile, with an analysis of their 'centricity' and of the semantic relations between their elements. Our conclusion was that the French lexicon of the Internet is morphologically and semantically influenced by the English terms but at the same time the number of borrowings is small.

### MOTS-CLÉS/KEYWORDS

emprunts, Internet, noms composés, terminologie

## 1. Introduction

Aujourd'hui, utiliser Internet fait partie des compétences de base dans les sphères privées et professionnelles de la plus grande majorité d'entre nous. Internet est un

des domaines de spécialité, avec la télévision ou la téléphonie, qui n'est pas réservé à un noyau d'experts. Son lexique technique, développé quasi exclusivement en langue anglaise, est très fécond. Ces particularités ne peuvent manquer d'influencer la terminologie qui y est utilisée et suscitent quelques interrogations : comment se traduit l'influence de l'anglais dans les terminologies des langues cibles ? Par des emprunts ? L'influence de la terminologie source est-elle toujours perceptible dans la terminologie cible ?

Pour répondre à ces questions, nous nous sommes intéressés à la terminologie Internet en français. Nous avons examiné un échantillon de termes, plus particulièrement de noms composés, catégorie abondante dans le domaine et pouvant donner lieu à des observations intéressantes tant sur le plan sémantique que sur le plan morphosyntaxique. Les composés étudiés ont été relevés dans un corpus comparable bilingue de revues datant de 2001<sup>1</sup>. Ces revues étant grand public, la terminologie relevée ne représente pas le lexique de l'ensemble du discours d'Internet, mais une étude des termes métaphoriques d'Internet (Brolles 2001) montre que les revues adressées à un public plus expérimenté contiennent, en général, les mêmes termes et que seules les revues « underground », qui traitent de l'actualité du piratage informatique, présentent des termes différents. Elles regorgent d'emprunts et de sigles (notamment les termes relatifs aux attaques) car elles visent un public revendiquant plus l'appartenance à une communauté.

La collecte a été effectuée manuellement à partir de la définition suivante du nom composé : lexie désignant un concept unique dont au moins un formant est autonome comme EN *Internet provider*, FR *aspirateur de site* ou EN *email*, FR *e-commerce*<sup>2</sup>. Tous les composés correspondant à notre définition et correspondant à un concept faisant partie du domaine d'Internet ont été retenus dans chaque sous-corpus. Il n'y a donc aucun effort particulier pour que les composés français soient les équivalents des composés anglais, même si nous avons relevé beaucoup de couples d'équivalents.

Le Tableau 1 montre la répartition des composés collectés (nombre d'occurrences) selon les langues et le nombre de formants.

TABLEAU 1

**Composés extraits du corpus comparable trilingue**

Composés	Nombre de formants					Total
	2	3	4	5	6	
Anglais	398	126	21	7	1	553
Français	540	98	9	2	0	649

Les composés sont plus nombreux en français qu'en anglais. Cela s'explique principalement par la concurrence terminologique plus forte en français : il existe parfois plus de termes français que de termes anglais pour désigner un seul et même concept puisqu'il peut y avoir concurremment emprunt, calque et traduction. Par exemple, trois termes français désignent le concept désigné par EN *pop-up* : *fenêtre publicitaire*, *fenêtre pop-up* et l'emprunt *pop-up*. L'instabilité désignationnelle, phénomène courant dans les domaines récents, n'épargne pas le domaine d'Internet et se manifeste de différentes manières dans les listes de termes tirés du corpus. On trouve des désignations concurrentes dans toutes les langues : EN *ad-blocker* et *ad-blocking program*, FR *adresse email*, *e-courrier* et *adresse électronique*. Elles peuvent

être des réductions de composés longs (EN *popup ad* > *popup*, FR *fournisseur d'accès Internet* > *fournisseur d'accès*), des emprunts (en français, *backdoor* est en concurrence avec *porte dérobée*, *porte de service* et *porte de derrière*), ou parfois simplement des variantes orthographiques (*dotcom*, *dot-com* et *popup*, *pop-up* ou *email* et *e-mail*).

Les composés à plus de deux formants, parfois dits « surcomposés » (Tournier 1991b : 171-172), sont moins nombreux en français qu'en anglais. À ce stade, on peut formuler l'hypothèse que ceci est dû à l'adoption par la langue cible des raccourcissements et des réductions des surcomposés anglais plutôt que des formes longues.

## 2. Profil morphosyntaxique

### 2.1. Structures compositionnelles

Nous avons calculé la fréquence « type » (Bauer 2001 : 47) des patrons syntaxiques, c'est-à-dire le nombre de composés différents qu'ils ont effectivement créés – ou leur productivité passée. Nous avons recensé le nombre de composés formés par chaque structure compositionnelle.

A- Corpus anglais : les pseudo-confixes fortement représentés

Ce sont les composés binaires qui sont les plus nombreux en anglais (71,97 %).

TABLEAU 2

Répartition des composés en anglais du corpus selon la nature de leurs formants<sup>3</sup>

	Structures	Exemples	Fréquence « type »	%
1	N + N	<i>snail mail</i>	201	50,50 %
2	PS + N	<i>cyberspace</i>	96	24,12 %
3	A + N	<i>intelligent agent</i>	72	18,09 %
4	13 structures mineures <sup>4</sup>	<i>business-to-business</i>	29	7,29 %
	<b>Total</b>		<b>398</b>	<b>100,00 %</b>

Les composés anglais se décomposent en seize structures selon la combinaison de leurs formants, dont trois produisent presque 93 % des composés.

Que le patron syntaxique le plus fréquent soit N + N (*snail mail*, *virus definition*) n'est pas une spécificité du lexique d'Internet. Tournier (1991b : 127) et Jouneau (1997 : 9) font la même observation dans la langue générale<sup>5</sup>. Les N + N représentent respectivement 59,80 % et 58,94 % de leur corpus. Les N + N sont également bien représentés dans d'autres études de langues de spécialité : 70,95 % de l'informatique (Béciri 1999 : 138), 50,3 % de la science-fiction (Gindre 1998 : 37) ou 20 % de la physico-chimie (Depierre 2001 : 106-108). Brocard (1998 : 21) et Paillard (2000 : 51) écrivent, sans donner de chiffre, que ce procédé est très utilisé dans le vocabulaire du cinéma et de la langue générale.

La fréquence « type » du patron PS + N (*cyberspace*, *Web access*)<sup>6</sup> (24,12 %), par contre, est inhabituelle. Elle ne semble pas avoir d'équivalent dans d'autres domaines et dans la langue générale; en tout cas, nous n'avons trouvé aucun chiffre pouvant être comparé. Les patrons compositionnels avec pseudo-confixes (PS + N, N + PS et

A + PS) représentent en tout 26,14 % des composés. La création de termes à partir de pseudo-confixes semble être caractéristique d'Internet.

Les composés A + N (*intelligent agent, secure server*) – au troisième rang du classement avec 18,09 % – sont plus fréquents que dans la langue générale. Dans les travaux de Tournier (1991b : 128) et de Jouneau (1997 : 9), respectivement 7 % et 11,19 % des composés binaires sont des A + N. Dans la terminologie de la physico-chimie, les composés A + N représentent 18 % du total des composés (27 % des composés binaires)<sup>7</sup> (Depierre 2001 : 107). En ce qui concerne l'informatique, Béciri (1999 : 138) a relevé 12,16 % de N + A (18 sur 148 composés binaires). La fréquence de ce patron serait probablement plus élevée sans la présence de pseudo-confixes qui occupent une position normalement occupée par un adjectif : dans *email, ebook, e-card, e* « remplace » *electronic*.

Les noms composés comprenant un verbe sont peu fréquents dans notre corpus (0,75 %). Ils font partie des treize patrons mineurs. La présence des verbes semble plus marquée dans la langue générale (Tournier 1985 : 114-115). En informatique, Béciri (1999 : 138) dénombre une vingtaine de formes à élément verbal (V + N, V-n + N, N + V-ing) qui représentent 11 % des composés. En physico-chimie, les composés possédant un élément verbal représentent un pourcentage similaire, 14 % des composés binaires (Depierre 2001 : 103). Quant au corpus de Jouneau (1997 : 9), il contient entre autres 5,62 % de N + V-er, 3,68 % de N + V-ing, 2,64 % de V + part. adv., ce qui représente 11,94 % des composés. Il est à noter que les éléments déverbaux, *browser, integrator, connection, indexing, hosted, etc.*, ont été considérés comme des substantifs et donc comptabilisés dans les patrons comprenant un N ou un A comme N + N, A + N, PS + N.

Les structures N + to + N (1,01 %), N + and + N (0,25 %), N + of + N (0,76 %) (*business to business, Denial of Service, click and mortar, peer-to-peer*) sont très faiblement représentées par rapport à la langue générale. Selon Tournier (1985 : 118), cette structure est généralement utilisée dans les registres très soutenus (les terminologies officielles, les vocabulaires juridique, religieux ou littéraire).

#### B- Corpus français : un faible taux d'emprunts intégraux

Le profil morphosyntaxique des composés français a été construit au moyen d'une double typologie, l'une basée sur l'origine et l'autre sur la fonction (en détaillant les patrons syntaxiques). La première, basée sur l'origine des termes (composés endogènes vs composés exogènes), distingue les emprunts à l'anglais, les composés endogènes purs, les composés endogènes hybrides et les composés exogènes. Les éléments formant les composés de la première sous-catégorie sont tous d'origine française (FR *attaque de ver, groupe de discussion*)<sup>8</sup>. Les noms composés de la deuxième sous-catégorie contiennent à la fois au moins un élément d'origine française et au moins un élément anglais (FR *canal de chat, chiffrement d'email*). Les composés exogènes sont des composés dont tous les éléments sont des emprunts morphosémantiques à l'anglais (FR *backdoor, black hat*). On parlera d'emprunts intégraux.

Un deuxième classement, reposant sur les patrons syntaxiques des composés et surcomposés endogènes uniquement (purs et hybrides), a permis de savoir si les patrons syntaxiques observés dans les lexiques français d'Internet sont ceux habituellement employés pour créer des composés dans ces langues ou si ce sont des calques des structures anglaises.

TABLEAU 3

## Répartition des composés en français selon l'origine et la nature de leurs formants

	Structures	Exemples	Fréquence « type »	%
<b>1</b>	<b>Composés endogènes</b>		<b>461</b>	<b>85,37 %</b>
<b>1.1</b>	<b>purs</b>		<b>225</b>	<b>41,11 %</b>
1.1.1	N + de + N	<i>aspirateur de site</i>	99	18,33 %
1.1.2	N + A	<i>adresse électronique</i>	85	15,74 %
1.1.3	N + N	<i>compte utilisateur</i>	24	4,44 %
1.1.4	PS + N	<i>télétravail</i>	5	0,93 %
1.1.5	N + à + N	<i>boîte aux lettres</i>	3	0,56 %
1.1.6	A + N	<i>haut débit</i>	2	0,37 %
1.1.7	N + PS	<i>espioniciel</i>	2	0,37 %
1.1.8	non + N	<i>non-connecté</i>	1	0,19 %
1.1.9	V + N	<i>pare-feu</i>	1	0,19 %
<b>1.2</b>	<b>hybrides</b>		<b>236</b>	<b>44,26 %</b>
1.2.1	N + N	<i>serveur proxy</i>	93	17,22 %
1.2.2	PS + N	<i>cybercriminalité</i>	63	11,67 %
1.2.3	N + de + N	<i>chasseur de bug</i>	32	5,93 %
1.2.4	N + PS	<i>accès web</i>	19	3,52 %
1.2.5	N + A	<i>ADSL individuel</i>	17	3,15 %
1.2.6	Flud	<i>wappanaute</i>	10	1,85 %
1.2.7	PS + A	<i>Web mobile</i>	3	0,56 %
1.2.8	N + de + PS	<i>citoyen du Net</i>	2	0,37 %
<b>2</b>	<b>Composés exogènes</b>	<b>backdoor</b>	<b>79</b>	<b>14,63 %</b>
	<b>TOTAL</b>		<b>540</b>	<b>100,00 %</b>

Les composés du lexique français d'Internet sont massivement endogènes (plus de 85 %). La proportion des termes exogènes est faible (15 %), surtout au regard de certaines particularités d'Internet : a) Internet est vraisemblablement un des domaines les plus dominés par l'anglais ; b) son développement et sa diffusion sont très rapides ; et c) il est très médiatisé. Ces faits étant généralement vecteurs d'emprunts, comment expliquer leur présence minime ici ?

Tout d'abord, la longueur des unités lexicales étudiées (entre deux et cinq éléments) explique sans doute le faible taux d'emprunts intégraux. Les lexies empruntées sont souvent brèves, que ce soit dans la langue générale ou dans certaines langues de spécialité, et le domaine d'Internet ne fait pas exception : *bug*, *chat*, *spyware*, *adware* ou *popup*. Autre fait non négligeable, les composés ont été recueillis dans un corpus écrit, et les emprunts sont plus utilisés dans la communication orale que dans la communication écrite, notamment en informatique, entre spécialistes initiés (Humbley 1988 : 99).

Ensuite, seuls les emprunts morphosémantiques et les sigles (FR *adware*, *pop under*, *bookmark*, *accès FTP*, *ADSL individuel*) sont comptabilisés comme emprunts à l'anglais. Les emprunts sémantiques (FR *fichier* < EN *file* dans *fichier MP3*, FR *moteur* < EN *engine* dans *moteur de recherche*) et les calques (FR *renifleur de navigateur* < EN *browser sniffer*, FR *agent intelligent* > EN *clever agent*) respectent la syntaxe de la langue française et entrent dans la catégorie des composés endogènes.

Enfin, la nature du corpus peut être à l'origine de la faible proportion d'emprunts. Les revues françaises dans lesquelles les termes ont été recueillis sont des revues traitant de l'actualité d'Internet qui intéresse tous les internautes. Certaines sont destinées à des profanes (*Web Magazine*) et d'autres à un public un peu plus expérimenté (*Netsurf*, *Officiel du Net*, *.Net*) mais toutes présentent les mêmes termes. Le taux d'emprunts est évidemment plus élevé dans les publications spécialisées que dans les ouvrages de vulgarisation.

Si les emprunts intégraux sont peu fréquents, les emprunts partiels représentent à eux seuls 44,26 % des noms composés du corpus. Pourquoi une telle supériorité numérique des composés hybrides ?

D'une part, les moyens de communication actuels font que les néologismes anglais sont rapidement introduits et diffusés en français. Dans une situation de néologie d'urgence, les spécialistes français d'Internet n'ont peut-être pas toujours le temps de créer un terme totalement endogène, calqué ou non (FR *accès à distance* < EN *remote access*, *clé de chiffrement* < EN *encryption key*) ou ne cherchent pas à le faire (FR *bloqueur de popup*, *filtrage d'email*, *chasseur de bug*). Mais surtout, les composés hybrides constituent une position d'équilibre entre le souci de clarté, qui conduit à ne pas emprunter, et le souci de l'authenticité, qui conduit au contraire à emprunter. Le souci de l'authenticité peut être particulièrement fort dans ce domaine et lorsque les spécialistes – souvent commerciaux – veulent impressionner. Cet équilibre donne lieu à des emprunts à chaque fois que ça ne compromet pas (trop) la clarté, c'est-à-dire lorsque l'emprunt est court et enchâssé dans une structure française qui augmente la clarté.

D'autre part, l'élément emprunté est : a) un sigle (FR *ADSL individuel*, *protocole FTP*, *langage HTML*) ; b) un terme anglais court implanté en français et donc réutilisé dans la formation de composés désignant de nouveaux concepts (FR *chasseur de bugs*, *session de chat*, *fichier log*, *virus de boot*) ; c) un pseudo-confixe (FR *cybercriminel*, *hyperlien*, *métamoteur*), ces derniers ne posant guère de problèmes de compréhension. Les termes pseudo-confixés, fréquents en français, sont considérés ici comme des créations hybrides puisque les pseudo-confixes ont d'abord été utilisés en anglais.

TABLEAU 4

**Composés français endogènes selon l'origine et la nature de leurs formants**

	Structures	Exemples	Fréquence « type »	%
1	N + de + N	<i>carnet d'adresses</i>	131	28,42 %
2	N + N	<i>protocole Internet</i>	117	25,38 %
3	N + A	<i>bannière publicitaire</i>	102	22,13 %
4	PS + N	<i>cyberterrorisme</i>	68	14,75 %
5	8 Structures mineures <sup>9</sup>	<i>site hacké</i>	43	9,32 %
	<b>Total</b>		<b>461</b>	<b>100,00 %</b>

Les composés endogènes utilisent douze procédés, dont quatre représentent plus de 90 % des créations. Les composés les plus représentés sont : N + de + N (*carnet d'adresses*, *notificateur de mail*), N + N (*code source*, *adresse IP*, *accès Internet*), N + A (*adresse électronique*) et PS + N (*cybercafé*, *télétravail*, *nétiquette*).

Comme l'illustre le Tableau 4, le français a souvent recours au patron N + de + N (28,42 %). Ceci n'est pas une spécificité d'Internet puisque, selon Kocourek (1991 : 129-135), cette composition est la plus fréquente dans les langues de spécialité. En informatique, les composés N + prép. + N concernent 47,75 % des composés contre 16 % dans la langue générale (Béciri 1999 : 137). L'explication de leur présence bien plus significative que dans le lexique d'Internet semble être le calque. L'anglais crée beaucoup par juxtaposition (50,50 % de N + N, 18,09 % de A + N) et le calque permet de traduire ces créations anglaises au moyen d'équivalents indigènes (FR *outil de recherche* < EN *search tool*, FR *groupe de discussion* < EN *chat group*). C'est un procédé de traduction assez rapide et facile à utiliser. D'après Candel (2000 : 358), 70 % des N + prép. + N contiennent la préposition *de* dans la langue générale.

Le deuxième patron le plus fréquent, N + N (25,38 %), est lui aussi habituel dans la langue générale et les langues de spécialité. Les dénominations par juxtaposition nominale sont observées dans la langue générale au moins depuis la période 1880-1914 (Candel 2000 : 345) et ont tendance à augmenter (Humbley 2000 : 80). La terminologie d'Internet semble aller dans ce sens. En informatique (Béciri 1999 : 138), avec 58 %, les juxtaposés N + N sont dominants parmi les composés binaires, pourcentage bien supérieur au nôtre. Selon Kocourek (1991 : 133), c'est une marque des « langues fonctionnelles contemporaines non esthétiques », comme les langues technoscientifiques.

Encore une fois, il n'est pas étonnant qu'avec environ 22 % N + A soit un patron fréquent. Il s'agit d'une construction courante du français, tout comme N + N (Candel 2000 : 345). Toutefois, la terminologie d'Internet semble moins apprécier ces formations que d'autres domaines : par exemple les constructions N + A sont couramment utilisées pour dénommer les disciplines sportives comme *danse acrobatique*, *danse gymnique*, *danse rythmique*, *danse sportive* (Tetet 2000 : 552). Béciri (1999 : 138) comptabilise 28 % de N + A dans son corpus spécialisé mais ne parle pas de la langue générale. Dans le lexique Internet du français, comme pour l'anglais, les composés formés d'un adjectif et d'un substantif sont délaissés au profit des PS + N. Par exemple, les termes construits avec *e-* (*e-courrier*, *e-tourisme*, *e-métier*) « remplacent » des N + A (*courrier électronique*, *tourisme électronique*, *métier électronique*), ce qui peut expliquer la présence plus faible des N+A dans le domaine qui nous intéresse qu'en informatique.

La formation de composés à partir de pseudo-confixes n'est pas propre à la langue anglaise. Cela semble être une spécificité de la terminologie d'Internet. En français, les PS + N occupent le quatrième rang des patrons les plus fréquents avec 14,75 %. Les pseudo-confixés représentent au total 20,38 % des composés endogènes. Cette construction apparaît moins fréquente en français qu'en anglais en raison de la structure de nos typologies. Des composés qui appartiennent à une classe de PS en anglais sont des emprunts intégraux en français (*e-mail*, *webmaster*, *e-business*, *freeware*, etc.) puisque les constructions ont été adoptées telles quelles.

En somme, l'influence de l'anglais se caractérise, au moins dans notre corpus, beaucoup par le calque et très peu par l'emprunt (respectivement moins de 15 % d'emprunts en français). Le français semble vouloir imiter l'anglais mais en créant des termes plus ou moins indigènes plutôt qu'en empruntant les termes anglais. L'étude de formants nouveaux, les pseudo-confixes, confirmera ou infirmera-t-elle cette tendance ?

## 2.2. Les pseudo-confixes

Un pseudo-confixe est un mot abrégé formellement semblable à un confixe – qui est une racine savante liée ou racine gréco-latine – entrant dans la formation de composés (Kocourek 1991 : 129-133). Les lexies pseudo-confixées sont des composés de type N abrégé + N, comme FR *autoroute* dans lequel *auto-* est la réduction de *automobile* et *téléspectateur* dans lequel *télé-* est la troncation de *télévision*. *Auto-* et *télé-* ne sont pas des confixes puisque ce sont des troncations. Il faut également les distinguer de ce que Tournier appelle des quasi-morphèmes (1991b : 127), autrement dit des éléments du lexique liés se combinant de préférence entre eux et pouvant apparaître à gauche comme à droite. La nouveauté et la complexité de la notion de pseudo-confixe sont bien illustrées par leur traitement lexicographique. Dans le *Merriam-Webster Online Dictionary (MWOD)*, *cyber-* et *e-* sont présentés comme des « *combining forms* », *hyper-* et *meta-* comme des préfixes classiques (sans aucune mention de leur utilisation dans le domaine d'Internet et de leur nouvelle signification), *net-* n'a pas de statut (seul figure un renvoi à l'article sur Internet), *web-* et *-ware* sont des noms (dans le cas de *web-*, aucune mention n'est faite de sa faculté de former des composés). Dans le *Cambridge Advanced Learner's Dictionary (CALD)*, *cyber-*, *hyper-*, *meta-*, *e-* et *-ware* sont des confixes, *net-* est un nom et *web-* n'a pas de statut (renvoi à « *World Wide Web* »).

Pour le français, hormis le *Grand Robert (GR 2001)* et le *Grand dictionnaire terminologique (GDT)*, les encyclopédies et dictionnaires récents consultés (en ligne ou version papier) répertorient très rarement les pseudo-confixes et très peu de termes formés à partir de pseudo-confixes. Pourtant, certains sont fermement implantés<sup>10</sup>. Le « retard » de la lexicographie française s'explique peut-être par le fait que les termes sont trop « anglais », ce qui constitue un frein à leur introduction. C'est le cas de la plupart des emprunts morphosémantiques mais aussi de termes comme *net surfeur*, *adresse e-mail*, *e-tourisme*. Mais la non-inclusion des pseudo-confixes est surtout due à leurs particularités ; ils ne sont pas des mots dans le sens traditionnel du terme et les dictionnaires répugnent donc à les cataloguer.

Dans la terminologie française d'Internet, Mopoho (1998 : 553-554) recense trente-deux pseudo-confixes qui lui semblent nouveaux et entrent dans la composition de nombreux néologismes : 1) *e-*, *é-*, *É*, *cyb-*, *cyber-*, *info-*, *M-*, *n-*, *net-*, *nul-*, *P-*, *publi-*, *web-* sont des pseudo-confixes antérieurs (apocopes) ; 2) *-el*, *-iciel*, *-icien*, *-iel*, *-ique*, *-naute*, *-net*, *-zine* sont des pseudo-confixes postérieurs (aphérèses) ; 3) *anté-*, *anti-*, *hyper-*, *inter-*, *macro-*, *micro-*, *nano-*, *rétro-*, *super-*, *téra-*, *trans-* sont des confixes sémantiques, c'est-à-dire des homographes de confixes classiques par glissement sémantique (*nano-*) ou par troncation de termes autonomes (*macro-*). Certains des éléments de cette troisième catégorie semblent contestables. L'auteur démontre que FR *micro-* est l'apocope de *micro-ordinateur* ou *micro-informatique* dans *micro-édition*, que *macro-* est l'apocope de *macrocommande* dans *macro-instruction*. Mais en quoi *anti-* et *nano-* sont-ils des homographes de confixes classiques ? Dans FR *antivirus*, *anti-spam*, *anti-* n'a-t-il pas le sens du confixe classique « qui s'oppose à »<sup>11</sup> ? Nous incluons dans la classe des pseudo-confixes les éléments résultant de la simplification de termes (EN *web* < *World Wide Web*, *-ware* < *software*) et ceux dont le changement sémantique est clair (*hyper-*).

Les pseudo-confixes de notre corpus trilingue peuvent être regroupés dans trois catégories. EN, FR *Cyber-* (*cyberspace*), EN, FR *meta-* (EN *meta tag*, FR *méta balise*),

EN, FR *hyper-* (EN *hypertext*, FR *hyperlien*), FR *télé-* (FR *télétravail*), FR *-naute* (FR *internaute*) sont des **pseudo-confixes sémantiques**. L'utilisation dans le domaine d'Internet de troncations de lexies qui sont des homographes de confixes classiques confère, en quelque sorte, une nouvelle vie à d'anciens confixes. EN, FR *Web-* (EN *website*), *net-* (FR *net-économie*), *e-* (FR *e-achat*), *-ware* (EN *freeware*) sont des **pseudo-confixes antérieurs** (apocopes) et des **pseudo-confixes postérieurs** (aphérèses). *Net-* et *e-* entraînent dans la composition de termes avant l'avènement d'Internet. *Net-* était alors la troncation de EN *network* et non de *Internet*. *E-* était déjà la troncation de EN *electronic* utilisée dans le cadre d'Arpanet, l'ancêtre d'Internet.

Les différents pseudo-confixes sont-ils aussi fréquemment employés en anglais et en langue cible ?

TABLEAU 7

## Fréquence des pseudo-confixes en anglais et français

Pseudo-confixes	Fréquence « type »	
	Corpus anglais	Corpus français
<i>cyber-</i> ; <i>ciber-</i>	44	49
<i>e-</i>	40	32
<i>web-</i>	37	46
<i>-ware</i>	13	7
<i>net-</i>	9	7
<i>hyper-</i> ; <i>hiper-</i>	6	3
<i>meta-</i>	3	5
<i>télé-</i> ; <i>tele-</i>	0	10
<i>-naute</i> ; <i>-nauta</i>	0	3

Dans les deux langues, *cyber-* est le pseudo-confixe le plus fréquent. Il est suivi de *e-* et de *web-*. Un relevé du nombre d'occurrences à l'aide de google.com, google.fr montre que la fréquence des lexies produites avec un même pseudo-confixe est extrêmement variable et quelques formations sont peu usitées. Par exemple, EN *cyberspace*, *e-business* et *weblog* comptent plus de 5 000 000 occurrences, *cyberstress*, *hypertext*, *e-learning* et *web address* sont répertoriés plus de 1 000 000 de fois alors que *cyber-directory*, *e-tourism* ou *web attack* comptent moins de 10 000 occurrences. Aucun terme avec *meta-* ou *net-* n'atteint une fréquence dépassant 999 999, et même 9 999 pour *-ware*. *Cyberspace* est très sollicité car il désigne un concept beaucoup plus employé que *cyber-recruitment*, par exemple, et il est souvent utilisé comme synonyme d'*Internet*. Les termes avec *cyber-* ou *e-* sont généralement bien plus fréquents que ceux contenant *meta-* ou *hyper-*. La fréquence des pseudo-confixes provient de l'importance des domaines conceptuels auxquels ils appartiennent. Le domaine conceptuel de *cyber-* est vaste – il désigne des activités du monde réel possibles grâce à Internet – alors que *meta-* – qui dénomme un principe de fonctionnement – intervient dans un domaine conceptuel plus restreint.

Nous constatons que, en matière de pseudo-confixation, la terminologie française d'Internet est passive. Les formants anglais sont réemployés dans les terminologies cibles et il n'y a pas de nouveaux pseudo-confixes. Le seul pseudo-confixe français recensé, *-iciel*, connaît des difficultés d'implantation. Il est supplanté par *-ware* pour

lequel il aurait dû être une alternative. La brièveté des pseudo-confixes et le fait qu'ils ne semblent pas employés dans d'autres domaines font que les internautes francophones et hispanophones les conservent et donnent ainsi, en quelque sorte, une touche d'originalité à leur lexique.

### 3. Profil sémantique

Les critères d'appartenance des composés à des catégories sémantiques sont fragiles. Lors d'un récent inventaire de la littérature sur les relations de détermination entre les éléments des noms composés (relations D1-D2), Arnaud (2003: 37 *seq.*) conclut que: a) les méthodes de classification sont subjectives et donc discutables; b) il est probablement impossible d'aboutir à un inventaire clos; c) certains composés ne peuvent être classés de façon univoque.

#### 3.1. La centricité: l'endocentricité en tête

Les composés sont souvent classés en fonction de leur degré de transparence par rapport à la classe d'objets qu'ils désignent. Un composé correspondant à la même classe que son déterminé est endocentrique alors qu'un composé correspondant à une classe différente de celle de son déterminé est exocentrique. Ce classement est difficile à réaliser car les composés dont au moins un des éléments est métaphorique sont problématiques. Par exemple, EN *address book* a pour déterminé (ou tête) *book*. Peut-on considérer que la classe d'objets à laquelle il renvoie (/fonction d'un logiciel de courrier électronique/) est la même que celle à laquelle correspond le déterminé *book* (/assemblage de feuilles, destinées à être lues, inclus entre deux couvertures/)? Il faudrait décider jusqu'à quel degré de ressemblance entre deux concepts un composé métaphorique peut être considéré comme endocentrique. Vaste problème, qui n'est pas sans rappeler les interminables débats sur la polysémie. Les composés qui sont métaphoriques et qui associent les deux conditions suivantes seront considérés comme endocentriques<sup>12</sup>:

- Ils ont une tête identifiable.
- L'analogie à l'origine de la métaphore reste perceptible: analogie de fonction (*mail-box*), analogie d'une propriété perceptive (comme le côté peu attirant et inélégant véhiculé par *nerd* dans *techno-nerd*), analogie de forme (*Web* dans *World Wide Web*), etc.

Par exemple, le composé *address book* a une lexie tête (*book*) et présente une analogie avec son équivalent papier (fonction: répertorié des adresses). *Address book* réunit les deux critères d'endocentricité. Le composé *firewall* est exocentrique dans le domaine d'Internet (mais endocentrique dans sa première acception, celle de /mur construit pour empêcher la propagation d'un feu/) puisqu'il n'a pas de tête sémantique: un *firewall* n'est pas un type de *wall*. Une analogie de fonction (protection) est sentie mais ce seul critère est insuffisant puisqu'un composé correspondant à une classe différente de celle de son déterminé est nécessairement exocentrique.

Un composé exocentrique entrant dans la formation d'un surcomposé peut donner un surcomposé endocentrique. Selon les deux conditions préalables, *proxy firewall*, *network firewall*, *hardware firewall* et *software firewall* sont endocentriques. Ces composés ont tous *firewall* pour tête sémantique, ce sont tous des hyponymes de

*firewall*. Une analogie de fonction (protection) est identifiable. Un *software firewall* est un dispositif de protection comme l'est un *firewall* dans le domaine de la protection incendie.

Pour pouvoir classer les composés du corpus en fonction de leur centricité, il est nécessaire de doter chacun d'une paraphrase.

- Par exemple, *Internet protocol* est un type de *protocol* mais *black hat* n'est pas un type de *hat*. Pour uniformiser les classes, nous formulons des paraphrases types dans lesquelles D1 représente le déterminé et D2 le déterminant.

- Les composés sont ensuite répartis en quatre catégories selon les paraphrases, où D1 désigne le déterminé et D2 le déterminant : D2D1 est un D1, D1D2 est un D1, D2D1 est un D1 et un D2, \*D2D1 est D1 et un D2.

TABLEAU 8

## Centricité des composés anglais du corpus

Catégorie	Ordre de détermination	Fréquence « type »	
		Corpus anglais	Corpus français
<i>Composés endocentriques</i>			
1	D2D1 est un D1	510	
	D1D2 est un D1		473
2	D1D2 est un D1	8	
	D2D1 est un D1		132
3	D2D1 est un D1 et un D2	5	
	D1D2 est un D1 et un D2		5
<i>Composés exocentriques</i>			
4	*D2D1 est un D1 et/ou un D2	30	
	*D1D2 est un D1 et/ou un D2		39

D'après le Tableau 8, les composés endocentriques dominent très largement, ce qui n'est pas surprenant. Les catégories 1, 2 et 3 représentent presque 95 % du total des termes anglais et 94 % des termes français. Pourquoi une telle prépondérance de l'endocentricité ?

L'endocentricité facilite la compréhension, surtout pour les non-spécialistes. Les combinaisons endocentriques ont évidemment une motivation plus claire que les exocentriques : elles « exposent dans leur signifiant un rapport non arbitraire, mais quoique non iconique (du genre de l'onomatopée), avec leur signifié » (Arnaud 2003 : 35).

En anglais, les composés de la catégorie 'D2D1 est un D1' sont les plus nombreux (92,22 % du corpus), à cause de l'ordre déterminant - déterminé de cette langue. Selon Tournier (1991a : 71), cette composition, dite de type germanique, est aussi la plus fréquente dans la langue générale. Les composés endocentriques dans lesquels le déterminant suit le déterminé, 'D1D2 est un D1' (*Denial of Service*), et ceux dont les deux éléments sont sur le même plan, 'D2D1 est un D1 et un D2' (*modem*), sont peu nombreux (respectivement 1,45 % et 0,90 %). Les composés de la catégorie 2 sont en majorité de type roman, c'est-à-dire reliés par une préposition, comme *network of networks*. En français, les composés endocentriques les plus nombreux sont de type 'D1D2 est un D1' (72,89 %). De fait, le nombre de composés endocentriques 'D2D1

est un D1' paraît élevé (20,34 %), surtout par rapport à d'autres terminologies, à cause des emprunts à l'anglais. Les composés pseudo-confixés amplifient le phénomène car la plupart des pseudo-confixes ont une position préfixale (FR *e-courrier*, *net-économie*, *cyber-conférence*, etc.).

Les composés endocentriques 'D1D2 est un D1 et un D2' dont les éléments ont un poids identique (ou composés additifs; EN *hacktivist*, FR *entreprenante*) sont aussi peu représentés dans les trois langues (environ 1 %).

La catégorie 4, celle des composés exocentriques, occupe la deuxième place en anglais et la troisième en français. Ses pourcentages (5,43 % en anglais et 6 % en français) sont bien inférieurs à ceux des endocentriques. Dans notre classification, la catégorie 4 est principalement constituée de composés empruntés à d'autres domaines: FR *porte de derrière* (construction), EN *bottleneck* (gestion/circulation routière), EN *Trojan Horse*, FR *cheval de Troie* (mythologie), EN *thumbnail* (anatomie, publicité), EN *guestbook* (hôtellerie), EN, FR *cookie* dans *Internet cookie* (cuisine). Les trois termes exocentriques qui ne sont pas des emprunts sémantiques ont été créés par la matrice de la métonymie: *click and mortar* (les matériaux désignent la nature de l'entreprise), *dot-com* et *dot-org* (le nom de l'extension désigne l'activité).

L'influence de l'anglais est également sentie sur le plan sémantique. Les constructions françaises étant des calques (parfois des emprunts), les proportions de composés endocentriques et exocentriques sont proches dans les deux langues. La langue cible adapte les constructions sources en inversant le plus souvent l'ordre de détermination pour que les équivalents respectent leur système. Internet ne se différencie donc pas d'autres terminologies ou de la langue générale.

### 3.2. Les relations sémantiques: similitude des relations sémantiques en anglais et en français

La méthode de classification des composés en fonction des relations sémantiques de leurs formants est semblable à celle utilisée par Arnaud (2003: 61 *sq.*):

- Chaque composé est doté d'une paraphrase: par exemple, EN *domain name* se paraphrase en 'D1 est attribué à D2' et *web address* en 'D1 permet de situer quelque chose à l'intérieur de D2'.

- Les paraphrases sont ensuite « uniformisées ». Par exemple, les paraphrases 'D1 est attribué à D2' et 'D1 permet de situer quelque chose à l'intérieur de D2' sont réunies sous l'étiquette 'D1 identifie, situe D2'.

- Dans le cas d'ambiguïtés catégorielles, le composé est classé dans une nouvelle catégorie regroupant les deux catégories possibles.

TABLEAU 9

#### Synthèse des relations sémantiques des composés anglais

Code	Description	Exemples	Fréquence « type »
4	D1 a un fonctionnement dont le procédé, le mode est D2	<i>ADSL modem, ebook</i>	104
13b	D1 est un instrument, mode, moyen utilisé pour D2	<i>access control, Internet protocol</i>	92

3a	D1 (logiciel, matériel, technique) a pour activité, fonction D2	<i>archive manager, download site</i>	82
6	D1 se distingue des autres D1 car il est D2 (nature)	<i>freeware, general portal</i>	78
7	D1 a D2 comme localisation, environnement	<i>email virus, cyber-shop</i>	74
3b	D1 (humain, entreprise) a pour objet d'activité, fonction D2	<i>space provider, bug hunter</i>	32
5	*D1D2 est un D1 et/ou un D2	<i>firewall, backdoor</i>	24
9	D1 est engendré, provoqué par D2	<i>browsing habit, virus alert</i>	16
3c	D1 (virus, attaque, pirate) a pour fonction D2	<i>DoS attack, buffer overflow attack</i>	15
11	D1 évalue, mesure D2 (durée, vitesse, nombre)	<i>chat session, connection time</i>	8
10	D1 est constitué, formé par D2	<i>hacker community, mail bombing</i>	6
1	D1 est désigné par analogie à une propriété de D2	<i>snail mail, mirror site</i>	5
8	D1 est combiné à D2	<i>hactivist, modem</i>	5
13a	D1 est un instrument, mode, moyen utilisé par D2	<i>client interface, user account</i>	5
2	D1 s'adresse à D2 (e-commerce, échange)	<i>peer-to-peer, business to consumer</i>	4
12	D1 est transféré en direction D2	<i>download, upload</i>	3
	<b>TOTAL</b>		<b>553</b>

TABLEAU 10

**Synthèse des relations sémantiques des composés français**

Code	Description	Exemples	Fréquence « type »
4	D1 a un fonctionnement dont le procédé, le mode est D2	<i>connexion dédiée, accès haut-débit</i>	118
3a	D1 (logiciel, matériel, technique) a pour activité, fonction D2	<i>annuaire de sites, générateur de virus</i>	110
13b	D1 est un instrument, mode, moyen utilisé pour D2	<i>clé de chiffrement, barre de navigation</i>	109
6	D1 se distingue des autres D1 car il est D2 (nature)	<i>lien consulté, e-mail encrypté</i>	99
7	D1 a D2 comme localisation, environnement	<i>cyberattaque, cyberépiciier</i>	77
3b	D1 (humain, entreprise) a pour objet d'activité, fonction D2	<i>concepteur web, fournisseur de courrier</i>	39
5	*D1D2 est un D1 et/ou un D2	<i>porte de service, jeune pousse</i>	31
3c	D1 (virus, attaque, pirate) a pour fonction D2	<i>pirate informatique, attaque DoS</i>	14
9	D1 est engendré, provoqué par D2	<i>habitude de navigation, infection virale</i>	13
11	D1 évalue, mesure D2 (durée, vitesse, nombre)	<i>session de chat, temps de téléchargement</i>	8

1	D1 est désigné par analogie à une propriété de D2	<i>courrier escargot, serveur mandataire</i>	7
10	D1 est constitué, formé par D2	<i>communauté hacker, réseau des réseaux</i>	7
8	D1 est combiné à D2	<i>entrepreneute, apprenti pirate</i>	6
2	D1 s'adresse à D2 (e-commerce, échange)	<i>peer-to-peer, business-to-business</i>	5
13a	D1 est un instrument, mode, moyen utilisé par D2	<i>compte utilisateur, interface client</i>	3
12	D1 est transféré en direction D2	<i>download, upload</i>	2
14	omission	<i>non-connecté</i>	1
	<b>Total</b>		<b>649</b>

Deux faits sont particulièrement marquants. Premièrement, le nombre de relations sémantiques étant très supérieur à celui des relations syntaxiques, un même patron syntaxique pourra correspondre à plusieurs relations sémantiques. Deuxièmement, on note une parfaite similitude des relations sémantiques dans les composés anglais et français. Une seule relation concernant un composé (FR *non-connecté*) est nouvelle: 14 'D1 est omis, D2 est devenu autonome'. Autrement dit, la tête sémantique FR *utilisateur* ou *ordinateur* est omise et le déterminant *non-connecté* est employé seul. Deux raisons peuvent expliquer cette similitude:

- Parmi les 649 noms composés français, 79 sont des composés exogènes (soit 14,63% du corpus). Leur relation identique se retrouve donc dans les analyses anglaises et françaises. Pour les termes exocentriques français, les relations sont identiques à celles observées dans le corpus anglais. Par exemple, pour EN *backdoor* et FR *backdoor, porte de derrière*, la paraphrase est \*D1D2 est un D1 et/ou un D2 (et l'interprétation est 'D1D2 est un accès secret'). Ces composés entrent dans la catégorie 5.

- Nombreux sont les noms composés français qui sont des calques des composés anglais. La relation sémantique entre le déterminé et le déterminant de EN *high-speed access* et FR *accès haut-débit* est la même: 4 'D1 a un fonctionnement dont le procédé, le mode est D2'. Celle entre les formants de EN *discussion group* et FR *groupe de discussion* est 13b 'D1 est un instrument, mode, moyen utilisé pour D2'.

Aucune catégorie ne regroupe plusieurs relations sémantiques mais cela ne signifie pas que la relation de certains noms composés n'a pas été difficile à paraphraser. Comme le constate Jouneau (1998: 59):

Lorsqu'au moins l'un des éléments d'un nom composé implique un changement de sens (métaphore ou métonymie), l'analyse des relations syntactico-sémantiques entre les éléments devient automatiquement plus complexe.

Par exemple, la relation qui unit EN *address* (D2) et *book* (D1) est-elle 10 'D1 est constitué, formé par D2' ou 13b 'D1 est l'instrument, mode, moyen utilisé pour D2'? De prime abord, les deux semblent envisageables. Mais les traits conceptuels n'ont pas tous le même poids et, dans le cadre d'Internet, le trait 'fonction' est plus important que le trait 'composition'. La fonction du 'carnet' est de permettre à l'internaute de stocker ses adresses de courrier électronique. Dire que le 'carnet' est constitué d'adresses est imprécis. Un *address book* est une fonction dans un logiciel et non un véritable carnet.

La classification des relations sémantiques appelle quelques remarques supplémentaires :

- Les catégories 3 et 13 ont des sous-catégories. Les paraphrases 3a / 3b / 3c et 13a / 13b sont trop proches pour appartenir à des catégories distinctes. Les descriptions 3a, 3b et 3c indiquent toutes une fonction. Seule la nature de D1 est variable (logiciel, matériel, technique, site / humain / virus, attaque, pirate). Les descriptions 13a et 13b indiquent aussi une fonction mais il y a une différence par rapport aux paraphrases 3 qui rend un regroupement impossible. Dans les paraphrases 3, D1 est actif et concret, il est un objet, une personne ou une entreprise dont la fonction principale est d'accomplir D2 (EN *ad blocker*, *ADSL provider*). Dans les paraphrases 13, D1 est passif et immatériel, il est un moyen, une technique, utilisé pour agir sur D2 ou qui est utilisé par D2 (EN *distribution list*, *encryption key*).

- Pour éviter de multiplier les catégories hapax, la catégorie 5 '\*D1D2 est un D1 et/ou un D2' réunit les composés pour lesquels aucune paraphrase analytique type n'a pu être formulée. Elle contient majoritairement des exocentriques dont les interprétations sont radicalement différentes les unes des autres : EN *firewall* > 'D2D1 est un dispositif de protection' ou *mailto* > 'D1D2 est un mot clé', etc. Cependant, sur les 30 composés anglais exocentriques, seuls 24 figurent dans la classe 5. Que sont devenus les six autres ? Les termes *cookie ID*, *Internet cookie*, *temporay cookie*, *permanent cookie*, *search engine spider* et *shopbot* présentent une relation simple malgré leur exocentricité. Un de nos critères d'identification – l'analogie doit être perceptible – est à l'origine du classement de ces termes en composés exocentriques. Ces composés ont une tête sémantique mais l'analogie est, d'après nous, imperceptible (les termes sont opaques).

- Les relations sémantiques les plus fréquentes occupent des rangs similaires dans les deux langues et leurs proportions diffèrent très peu.

TABLEAU 11

#### Bilan bilingue des relations sémantiques

Relations sémantiques	Anglais	Français
3a 'D1 (logiciel, matériel, technique) a pour activité, fonction D2', 3b 'D1 (humain, entreprise) a pour objet d'activité, fonction D2', 3c 'D1 (virus, attaque, pirate) a pour fonction D2'	23,33 %	25,12 %
4 'D1 a un fonctionnement dont le procédé, le mode est D2'	18,81 %	18,18 %
6 'D1 se distingue des autres D1 car il est D2 (nature)'	14,10 %	15,25 %
7 'D1 a D2 comme localisation, environnement'	13,38 %	11,86 %

La prépondérance des calques et la présence d'emprunts en français expliquent que les chiffres ci-dessus sont proches. Les quatre relations sémantiques les plus fréquentes sont identiques dans les deux langues et leur poids est constant (moins de 4 % d'écart entre le pourcentage le plus faible et le plus fort).

#### 4. Conclusion

Nous avons observé que, sur le plan morphosyntaxique, les composés du domaine d'Internet se caractérisent par :

- La fréquence d'un nombre réduit de structures compositionnelles : 82,10 % des composés du corpus anglais sont formés d'après six structures (N+N, PS+N, A+N, N+N+N, A+N+N, PS+N+N) alors que 58 patrons ont été relevés. En français, les 45 patrons endogènes ont généré 86,29 % des composés collectés. Cinq structures vives (N+de+N, N+N, N+A, PS+N et Emp2) ont produit 76,57 % des termes. Les composés exogènes représentent 13,71 % des termes français.

- Le faible pourcentage de composés français exogènes (ou emprunts intégraux). Cela est aussi surprenant qu'intéressant étant donné la dépendance technologique des pays francophones en matière d'Internet. La brièveté généralement constatée des emprunts – alors que les unités considérées dans cette étude sont longues –, le fait que le corpus exploité soit un corpus écrit écartant les revues « underground » – les emprunts fourmillent dans la communication orale et underground – et que seuls les emprunts morphosémantiques et les sigles sont comptabilisés ont été les raisons évoquées pour expliquer ce pourcentage. Paradoxalement, de nombreux composés endogènes hybrides, qui sont un signe de l'équilibre entre le souci de clarté et le souci d'authenticité, ont été relevés (44,26 % des composés français). Ils semblent être une particularité du lexique d'Internet qui le différencie des autres terminologies plus soutenues.

- La formation de composés à partir de pseudo-confixes. Ceci semble être une nouveauté. Les premiers termes pseudo-confixés du domaine ont été créés en anglais mais la pseudo-confixation s'est étendue à la langue cible.

Sur le plan sémantique, nos observations ont dévoilé les spécificités suivantes des noms composés :

- Dans les deux langues, les composés endocentriques sont très fortement majoritaires : 94,21 % du corpus anglais et 93,68 % du corpus français. Cette prépondérance s'explique par le fait que l'endocentricité facilite la compréhension du nom composé.

- En ce qui concerne les relations sémantiques entre les formants, treize relations principales ont été identifiées en anglais et quatorze en français. Pour les deux langues, les relations les plus fréquentes sont (ordre décroissant) : la fonction (D1 actif), le principe de fonctionnement, la nature et la fonction (D1 passif). En français, la relation de localisation est fortement représentée.

Cette analyse morphosyntaxique et sémantique des noms composés anglais et français permet de tirer quelques conclusions :

- Cette radiographie des composés est révélatrice de l'utilisation du calque dans la langue cible et de l'influence de l'anglais sur cette dernière. Le calque est le procédé de traduction le plus employé pour rendre les composés anglais. L'usage de l'emprunt est limité, contrairement à ce que l'on aurait pu penser. Les emprunts sont plus fréquents dans les conversations entre spécialistes et dans les milieux plus spécialisés que dans des textes destinés au grand public. « Les calques sont une [...] forme de trace d'une langue dans une autre langue » (Arnaud 2004 : 346). Comme ils respectent la formation des composés en langue cible, ils passent inaperçus mais ils augmentent peut-être le nombre de composés. La langue cible aurait pu utiliser d'autres matrices lexicogéniques.

- Les journalistes recourent au calque car ils sont sous l'influence de l'anglais. En témoignent, premièrement, l'augmentation des constructions avec deux substantifs non reliés comme FR *code source*, *accès Internet*, *adresse IP*, *fichier log*, *site*

*marchand* et, deuxièmement, l'ordre parfois déterminant - déterminé des constructions au lieu de déterminé - déterminant (*cyber-*, *e-*, *meta-*, *net-*, *télé-*, + emprunts). Des facteurs extralinguistiques poussent le français vers ces créations plus synthétiques. En effet, les techniques ont une préférence pour les termes sémantiquement complexes et concis puisqu'il faut suggérer le plus d'informations possible au moyen d'un terme le plus compact possible.

- Les particularités des composés anglais se retrouvent dans les composés français: a) les patrons compositionnels sont identiques; b) les pseudo-confixes anglais ont des équivalents français eux aussi fréquents; c) les proportions d'endocentricité et d'exocentricité sont similaires; et d) les relations déterminant - déterminé sont les mêmes. Cette dépendance terminologique est sans doute le reflet de la dépendance technologique.

- La communauté internaute n'est pas limitée à un petit groupe d'initiés, d'où la simplicité des termes qui se caractérise par une endocentricité massive, l'utilisation de la métaphore et la brièveté des composés.

- Pour nommer une même réalité, le webmaster et l'internaute néophyte utilisent le même terme. Dans certains domaines, comme la cardiologie, les spécialistes utilisent des termes (FR *dyspnée*) et les non-spécialistes des synonymes (FR *essoufflement*) (De Oliveira 2005). Si le lexique d'Internet est loin des terminologies spécialisées généralement soutenues, c'est parce que ce domaine fait partie du quotidien des locuteurs anglophones et francophones. Internet leur est familier et occupe une grande place dans leur vie privée et professionnelle et dans les médias. La fréquence d'utilisation de la métaphore due à l'emprunt de termes à la langue générale et à d'autres lexiques (navigation et médecine, par exemple) est peut-être également liée à l'utilisation d'Internet par le grand public.

## NOTES

1. Pour l'anglais, *Internet Magazine* (3 numéros), *Network magazine* (3 numéros), *Wired* (4 numéros) et *PC World* (12 numéros); pour le français, *Web Magazine* (3 numéros), *Netsurf* (3 numéros), *.Net* (3 numéros) et *l'Officiel du Net* (3 numéros).
2. Cette définition est peu restrictive pour ne pas nuire à l'analyse en écartant les catégories mineures. Aucune définition totalement satisfaisante n'étant disponible dans la littérature, nous avons complété l'opération de sélection par un filtrage (inclusion dans des ouvrages de référence et indexation par des moteurs de recherche sous certaines conditions), ce qui permet de réunir un faisceau de présomptions.
3. Les pourcentages qui suivent sont calculés par rapport au nombre total de composés relevés pour chaque langue.
4. Il s'agit de: N + PS (1,01 %), A + PS (1,01 %), N + to + N (1,01 %), Flud (1,01 %), N + of + N (0,75 %), N + up (0,50 %), down + V (0,50 %), in + N (0,25 %), up + N (0,25 %), N + to (0,25 %), N + over (0,25 %), up + V (0,25 %), N + and + N (0,25 %).
5. Toute référence à une source assez ancienne pose évidemment le problème de la pérennité des conclusions.
6. Voir section 2.2.
7. Ces chiffres correspondent au corpus dictionnaire de l'auteur. Si on regroupe les catégories contenant des dérivés, comme Depierre l'a fait avec les N+N, le pourcentage atteint 31 % des binaires: A+N (27 %), V-n + N (2 %), A + V-ing (1 %) et A + V-er (1 %).
8. C'est l'origine immédiate du terme qui nous intéresse et non son origine « ultime ». À ce stade, nous ne prenons pas en compte l'ajout d'un nouveau signifié à un signifiant existant sous une influence étrangère.
9. N+PS (4,55 %), N+à+N (0,65 %), N+de+PS (0,43 %), A+N (0,43 %), PS+A (0,65 %), V+N (0,22 %), Adv+N (0,22 %), Flud (2,17 %).

10. Pour FR *cyber-*, *hyper-*, *-naute* et *web-* et leurs dérivés, l'*Encyclopédie Larousse* ([www.encyclopedie-larousse.fr/](http://www.encyclopedie-larousse.fr/)) a comme entrées: *cyber-*, *cybercafé*, *cybermonde*, *cyberespace*, *cybernaute*, *hyperlien*, *hypermédia*, *internaute* et *web*. *E-* est absent.
11. Pour nous *anti-* est bien un confixe et non un pseudo-confixe car, dans les termes du domaine d'Internet (EN, FR *antivirus*, *antispam*, *antihacker*), il a bien le sens 'lutte contre' comme dans les termes médicaux (FR *antifèvre* ou *antibactérien*). Dans ces deux domaines, *anti-* exprime une action bienfaisante et sa fréquence est élevée.
12. Le premier critère suffira pour déterminer la classe des composés non métaphoriques.

## RÉFÉRENCES

- AHRONIAN, C. (2001): *Les anglicismes dans la langue espagnole, Etude de deux domaines: l'informatique et la génétique*, Mémoire de DEA, Université Lumière Lyon II.
- AHRONIAN, C. (2005): *Les noms composés anglais, français et espagnols du domaine d'Internet: Traduction des composés anglais en français et en espagnol*, Thèse de doctorat, Université Lumière Lyon II.
- ARNAUD, P. (2003): *Les Composés timbre-poste*, Lyon, PUL.
- ARNAUD, P. (2004): « Problématique du nom composé », in ARNAUD, P. (dir.), *Le Nom composé, Données sur 16 langues*, Lyon, PUL, pp. 329-353.
- BAUER, L. (2001): *Morphological Productivity*, Cambridge, CUP.
- BÉCIRI, H. (1999): *La néologie dans le domaine microinformatique professionnel en français et en anglais: création lexicale et explication en contexte*, Thèse de doctorat, Université Paris VIII.
- BROCARD, Valérie (1998): « Étude lexicologique des termes techniques du cinéma américain », in TOURNIER, J. (dir.), *Lexicologie anglaise et comparée*, Paris, Annales littéraires de l'Université de Franche Comté, Belles Lettres, pp. 15-30.
- CANDEL, D. (2000): « Français scientifique et technique », in ANTOINE, G. et B. CERQUIGLINI (dir.), *Histoire de la langue française (1945-2000)*, Paris, CNRS éd., pp. 339-369.
- CRUSE, D. A. (1986): *Lexical Semantics*, Cambridge, CUP.
- DE OLIVEIRA, I. (2005): *Nature et fonctions de la métaphore dans la terminologie médicale: étude comparée du français et du portugais*, Thèse de doctorat, Université Lumière Lyon II.
- DEPIERRE, A. (2001): *La composition en anglais scientifique: application au domaine de la physico-chimie*, Thèse de doctorat, Université du Littoral Côte d'Opale.
- GINDRE, P. (1998a): « Le Lexique de la science-fiction », in TOURNIER, J. (dir.), *Lexicologie anglaise et comparée*, Paris, Annales littéraires de l'Université de Franche Comté, Belles Lettres, pp. 31-58.
- HUMBLEY, J. (1988): « Comment le français et l'allemand aménagent la terminologie de l'informatique », *La Banque des mots*, numéro spécial, pp. 85-148.
- HUMBLEY, J. (1994): « Quelques aspects de la datation des termes techniques: le cas de l'enregistrement et de la reproduction sonores », *Meta* 39-4, pp. 701-715.
- HUMBLEY, J. (2000): « Évolution du lexique », in ANTOINE, G. et B. CERQUIGLINI (dir.), *Histoire de la langue française (1945-2000)*, Paris, CNRS éd., pp. 71-106.
- JOUNEAU, I. (1997): *La composition nominale par juxtaposition en anglais contemporain*, Thèse de doctorat, Université de Franche-Comté.
- JOUNEAU, I. (1998): « La métasémie dans la composition nominale par juxtaposition en anglais contemporain », in TOURNIER, J. (dir.), *Lexicologie anglaise et comparée*, Annales littéraires de l'Université de Franche Comté 652, Belles Lettres, pp. 59-71.
- KOCOUREK, R. (1991): *La langue française de la technique et de la science*, Wiesbaden, Brandstetter Verlag.
- МОРОХО, R. (1997): « D'hier à aujourd'hui: évolution et tendances de la néologie morphologique dans le langage de l'Internet », *La mémoire des mots*, Actualités scientifiques, Actes du colloque de Tunis 25-26, 27 septembre 1997, Tunis, AUELF-UREF, pp. 551-561.
- PAILLARD, M. (2000): *Lexicologie contrastive anglais-français, formation des mots et construction du sens*, Paris, Ophrys.

TETET, C. (1994): «La terminologie de l'alpinisme dans les dictionnaires», *Meta* 39-4, pp. 651-661.

TOURNIER, J. (1985): *Introduction à la lexicogénétique de l'anglais contemporain*, Genève, Champion-Slatkine.

TOURNIER, J. (1991a): *Précis de lexicologie anglaise*, Paris, Nathan.

TOURNIER, J. (1991b): *Structures lexicales de l'anglais*, Paris, Nathan.