

Université Lumière – Lyon 2
Faculté des Langues
École doctorale Humanités et sciences humaines
Département des Langues Étrangères Appliquées

Les noms composés anglais français et espagnols du domaine d'Internet :

par Céline AHRONIAN

Thèse de doctorat en Lexicologie et Terminologie multilingues – Traduction
sous la direction d'Henri BÉJOINT
soutenue le 9 décembre 2005

Composition du jury : Pierre ARNAUD, Professeur à l'Université Lumière Lyon 2 Henri BÉJOINT,
Professeur à l'Université Lumière Lyon 2 Xavier BLANCO, Professeur à l'Université Autonome de
Barcelone John HUMBLEY , Professeur à l'Université Paris 7

Table des matières

Remerciements ..	1
Liste des abréviations ..	3
Introduction ..	5
Première partie. Les notions de composition et de nom composé ..	11
Chapitre 1. Qu'est-ce que la composition ? Qu'est-ce qu'un nom composé ? ..	11
1. Le nom composé ..	12
2. Conclusion ..	24
Chapitre 2. Composition et	26
1. ... accentuation ..	27
2. ... graphie ..	32
3. ... distribution ..	33
4. ... Sémantique ..	39
5. ... Fréquence ..	41
6. Conclusion ..	42
Chapitre 3. Quelques typologies des noms composés ..	44
1. Typologies conçues pour la langue générale ..	44
2. Typologies conçues pour des domaines de spécialité ..	53
Deuxième partie. Les noms composés anglais, français et espagnols du lexique d'Internet ..	61
Chapitre 1. Le corpus : constitution et contenu ..	61
1. Travail terminologique et corpus ..	61
2. La constitution du corpus ..	65
3. Typologie morphosyntaxique des noms composés du corpus ..	69
Chapitre 2. Les noms composés anglais ..	81
1. Le profil morphosyntaxique des noms composés collectés ..	81
2. Le profil sémantique des noms composés collectés ..	105

3. La combinatoire intermatricielle . . .	111
Chapitre 3. Les noms composés français . . .	119
1. Le profil morphosyntaxique des noms composés collectés . . .	119
2. Le profil sémantique des noms composés collectés . . .	144
3. La Combinatoire intermatricielle . . .	148
Chapitre 4. Les noms composés espagnols . . .	152
1. Le profil morphosyntaxique des noms composés collectés . . .	152
2. Le profil sémantique des noms composés collectés . . .	175
3. la combinatoire intermatricielle . . .	179
Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol . . .	183
Chapitre 1. Pré-enquête . . .	183
1. La traduction du lexique d'Internet . . .	184
2. La circulation des termes dans le domaine d'Internet . . .	196
3. Des équivalences types anglais-français et anglais-espagnol sont-elles possibles ? . . .	198
Chapitre 2. Les noms composés anglais et leurs équivalents français et espagnols . . .	202
1. Les équivalences types anglais-français et anglais-espagnol . . .	202
2. Identification des exceptions . . .	218
3. Les traducteurs automatiques et la traduction des noms composés du domaine d'internet . . .	233
Chapitre 3. Vers l'automatisation du système d'équivalences types anglais-français, anglais-espagnol . . .	245
1. Future base du système d'équivalences types automatisé . . .	245
2. Présentation du Générateur d'équivalents . . .	257
Conclusion . . .	281
Bibliographie . . .	289
SITES INTERNET . . .	299
Glossaires Internet : . . .	299
Ressources terminologiques : . . .	299
Dictionnaires : . . .	299

Index terminologique ..	301
A ..	301
B ..	301
C ..	302
D ..	302
E ..	303
F ..	303
G ..	304
H ..	304
J ..	304
L ..	304
M ..	305
N ..	305
O ..	305
P ..	306
Q ..	306
R ..	306
S ..	306
T ..	307
U ..	308
Annexe 1. Liste des figures ..	309
Annexe 2. Liste des tableaux ..	311
Annexe 3. Application du critère dictionnairique à un échantillon de collocations ..	317
Annexe 4. Application du triple test de compositionnalité à dix candidats composés ..	321
Annexe 5. Fiches des revues anglaises, françaises et espagnoles ..	323
1. Fiches des revues anglaises ..	323
2. Fiches des revues françaises ..	324
3. Fiches des revues espagnoles ..	326
Annexe 6. Rubriques des revues anglaises, françaises et espagnoles ..	329

1. Revues anglaises .	329
2. Revues françaises .	330
3. Revues espagnoles .	331
Annexe 7. Les noms composés du corpus .	333
1. Les noms composés anglais .	333
2. Les noms composés français .	349
3. Les noms composés espagnols .	369
Annexe 8. Typologies morphosyntaxiques des noms composés du corpus .	387
1. Les noms composés anglais .	387
2. Les noms composés français .	393
3. Les noms composés espagnols .	399
Annexe 9. Typologies de la centricité des noms composés du corpus .	407
1. Les noms composés anglais .	407
2. Les noms composés français .	411
3. Les noms composés espagnols .	416
Annexe 10. Types de fenêtres . .	423
Annexe 11. Autres procédures de réalisation du générateur d'équivalents envisagées .	425
1. Une table unique .	425
1.1. Travail préliminaire .	425
1.2. Fonction principale du générateur : Recherche des équivalents .	426
1.3. Avantages .	427
1.4. Inconvénients .	427
1.5. Schémas .	427
2. Une table unique basée sur le type compositionnel des composés sources et cibles .	429
2.1. Travail préliminaire .	429
2.2. Recherche des équivalents : fonction principale du générateur .	430
2.3. Avantages .	431
2.4. Inconvénients .	431

<i>2.5. Explication des opérations sous-jacentes</i>	432
<i>2.6. Schémas</i>	433
Annexe 12. La base de données du générateur d'équivalents	437
1. Les tables	437
2. Les requêtes	438
<i>2.1. Requêtes Sélection</i>	438
<i>2.2. Requêtes Union</i>	438

Remerciements

Les résultats que je soumets dans la présente thèse ont été développés au sein du CRTT (Centre de Recherche en Terminologie et Traduction) de l'université Lyon II.

Toute ma reconnaissance s'adresse à :

- Henri Béjoint, mon directeur de recherche, qui m'a aiguillée, conseillée, lue et relue avec minutie et soutenue avec gentillesse et bonne humeur.
- Pierre Arnaud pour ses précieuses remarques sur la partie définitionnelle.
- Les membres du CRTT qui ont manifesté leur intérêt pour le projet d'équivalences types et son informatisation lors d'une présentation à l'occasion d'un séminaire. Dans les moments de doute penser à leur intérêt a ravivé mon énergie.
- Yannick Brolles pour son aide inestimable dans l'élaboration du Générateur d'équivalents. Sans ses connaissances en informatique, le générateur ne serait pas ce qu'il est.
- Le département de Formation Continue Langues de Lyon II qui a rendu la cohabitation entre la fin de mes travaux de recherche et mes débuts dans l'enseignement la plus agréable possible.
- Ma famille et mes amis pour avoir partagé les sautes d'humeur et m'avoir encouragée jour après jour jusqu'au point final de ce manuscrit

Les noms composés anglais français et espagnols du domaine d'Internet :

Liste des abréviations

CALDO	<i>Cambridge Advanced Learner's Dictionary Online</i> (2003)
DGLE	<i>Diccionario General de la Lengua Española</i> (en ligne, mise à jour : 2003)
DRAE 1996	<i>Diccionario de la Real Academia Española</i> (version CD-Rom)
DRAE 2001	<i>Diccionario de la Real Academia Española</i> (version papier)
EN	l'exemple est en langue anglaise
ES	l'exemple est en langue espagnole
FR	l'exemple est en langue française
GDT	<i>Le Grand dictionnaire terminologique</i> (en ligne, mise à jour : septembre 2005)
GR 2001	<i>Le Grand Robert</i> (version papier)
MWOD	<i>Merriam-Webster Online Dictionary</i> (version électronique du <i>Merriam-Webster's Collegiate Dictionary, Tenth Edition</i> , 2002)
NND	<i>Newton Nouveau Dictionnaire des télécommunications, de l'informatique et de l'Internet</i> (2000)
OED 1992	<i>Oxford English Dictionary</i> (version CD-Rom)
PR 1996	<i>Le Petit Robert</i> (version CD-Rom)

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Introduction

Ce projet tire son origine d'une constatation. On dit des domaines de spécialité qu'ils sont riches en néologismes morphologiques, c'est-à-dire en dérivés, mots composés, sigles, troncations et amalgames. Cette remarque s'applique plus particulièrement aux langues germaniques (Melka, 2002 : 100) dont l'anglais, langue de création d'Internet, fait partie. Notre étude des anglicismes utilisés en espagnol dans les domaines de la génétique et d'Internet (Ahronian 2001) corrobore cette observation. Dans l'échantillon de termes analysés, la présence d'anglicismes composés est particulièrement manifeste. La lecture de magazines d'Internet anglais, français et espagnols confirme que le nombre considérable de mots composés est une réalité.

Le présent travail de recherche sera consacrée à trois langues ayant un statut différent au regard du domaine d'Internet : l'anglais¹ est la langue de néologie primaire du domaine, et le français et l'espagnol sont deux langues recourant à la néologie traductive, c'est-à-dire recevant et adaptant les termes anglais.

Pourquoi avons-nous choisi d'étudier le lexique d'Internet ?

Dromard et Seret² définissent Internet comme un réseau informatique international

¹ Par langue anglaise, ou anglais, nous entendons l'anglais parlé dans tous les pays anglophones : Grande-Bretagne, Irlande, Etats-Unis, Australie, Canada, Nouvelle Zélande, etc. De même, langue française et langue espagnole renvoient au français et à l'espagnol parlés dans les pays francophones et hispanophones.

² *Encyclopædia Universalis* en ligne (intranet Lyon 2 : http://intranet.univ-lyon2.fr/wt_encyclo.html)

réalisant l'interconnexion d'un nombre croissant de réseaux de toutes dimensions. Colombain (1998 : 112) ajoute que le réseau s'est très rapidement développé à l'échelle internationale, de manière un peu anarchique au fur et à mesure de l'interconnexion des réseaux. Pour le grand public, utiliser Internet c'est envoyer un message par le biais d'un serveur de courrier électronique, dialoguer en temps réel avec un interlocuteur se trouvant au bout du monde ou à deux pas grâce à un canal IRC, faire partie d'un groupe de discussion au moyen d'un serveur Usenet, naviguer sur le « World Wide Web », la partie multimédia d'Internet accessible grâce au protocole http, à la recherche d'informations ou de divertissements, ou encore de télécharger des programmes gratuits ou payants grâce au protocole FTP.

Le nombre d'utilisateurs du réseau Internet est considérable et ne cesse d'augmenter depuis 1994, date de son ouverture au grand public :

Tableau 1 : Nombre d'internautes

Pays	En millions ³	Date
Grande-Bretagne	30,4	Décembre 2002
	28	Décembre 2001
États-Unis	174,2	Mars 2002
	166,14	Août 2001
France	23	Décembre 2002
	21,9	Décembre 2001
Canada	16,84	Mars 2003
	14,44	Août 2001
Espagne	17	Décembre 2002
	10,1	Décembre 2001
Argentine	3,88	Juillet 2001
Mexique	3,42	Juillet 2001

Le nombre d'internautes dans le monde était de 384 millions en 2000, 498,7 millions en 2001, 565,7 millions en 2002 et 633,6 millions en 2003. Il est estimé à 724,9 millions en 2004.⁴ En mai 2003, la population internaute de la Grande-Bretagne était évaluée à 47%, soit près de la moitié de la population britannique (43% en 2002). En mars 2003, l'Espagne comptait 25,5% d'internautes. Le pays accuse un retard mais fait preuve d'un réel dynamisme (augmentation de 16% en un an). À la même période, on recense 40,9% d'internautes en France.

Les vocabulaires technoscientifiques constituent la plus grande partie du lexique d'une langue mais aussi la plus productive (Candel, 2000 : 339). Ils occupent une place prépondérante dans les échanges entre spécialistes et les relations entre les spécialistes et le grand public.

³ Chiffres pour les pays anglophones, francophones et hispanophones dont les langues sont les cibles de cette étude (source : www.journaldunet.com/chiffres-cles.shtml).

⁴ Source : www.journaldunet.com/chiffres-cles.shtml

Internet est une véritable merveille technologique qui évolue rapidement, utilise de très nombreux concepts et dont aucune autorité centrale ne régit les innovations et les avancées. Par conséquent, les concepts ne sont pas définis par un seul et unique spécialiste ou un groupe de spécialistes et un flou conceptuel et/ou terminologique s'installe. Par exemple, le terme anglais *log file* a, à notre connaissance, quatre équivalents français : *fichier log*, *fichier de logs*, *inventaire des connexions* et *fichier journal*. Le lexique d'Internet est un des plus féconds des sciences et des techniques depuis la fin du XX^e siècle (nombre de dictionnaires, glossaires, lexiques monologues, bilingues ou multilingues ont vu le jour depuis quelques années). Le terminologue découvre, analyse et exploite les termes dénommant les concepts pour fournir au traducteur les données utiles à la traduction des textes de ce domaine récent. La traduction est nécessaire pour endiguer le flot de termes anglais qui inonde les vocabulaires français et espagnols d'Internet pour conserver une information scientifique et technique en français et en espagnol. La vulgarisation des connaissances, la dissémination de l'information est plus rapide dans le domaine d'Internet que dans d'autres domaines, comme la médecine ou la physique nucléaire. Les populations francophones et hispanophones sont attirées par ce moyen de communication et de documentation de plus en plus ancré au sein des foyers, des entreprises, des écoles, des institutions, etc. et sont avides d'en découvrir plus. Pour ces raisons, il est utile de faciliter la traduction des termes et de réduire le flou terminologique.

Pourquoi porter notre intérêt exclusivement sur une catégorie de composés du lexique d'Internet : les noms composés ?

Rappelons tout d'abord que le début des réflexions de l'époque contemporaine sur la composition est marqué par le *Traité de la formation des noms composés* de Darmesteter publié en 1874. De nombreux travaux ont dès lors été consacrés à la composition dans la langue générale. G. Gross (1996) et Arnaud (2003) passent en revue la littérature et citent entre autres les travaux de Jespersen (1942/1961), Marchand (1960), Martinet (1965), Benveniste (1967), Rohrer (1967), Grevisse (1969), Adams (1973), Guilbert (1975), Saussure (1982), Tournier (1985), Benson et al. (1986), Anscombe (1990), Riegel (1990), Pavel (1995), Bauer (1998).

Les noms composés occupent une place de choix en terminologie :

Si l'on fait l'inventaire du vocabulaire des langues de spécialité, on se rend compte que les substantifs constituent l'essentiel du vocabulaire et que parmi eux les noms composés s'y taillent la part du lion. (G. Gross, 1996 : 27)

Parmi les récents travaux de recherche en terminologie auxquels nous avons eu accès, certains, dont ceux de Depierre (2001) et Brocard (1998), sont entièrement consacrés aux noms composés d'une langue de spécialité, d'autres leur accordent une partie, un chapitre ou leur consacrent une description quantitative comme ceux de Béciri (1999) et Gindre (1999).

Les noms composés génèrent des problèmes de traduction. La plupart des travaux mentionnés ci-dessus sont des analyses descriptives. Aucun travail terminologique n'est orienté sur la manière de traduire les noms composés d'un domaine. Or, selon Meyer et al. (1998 : 637) :

[...] Pour éviter l'anglicisation tous azimuts de la terminologie, il est essentiel que les langagiers travaillant dans d'autres langues que l'anglais puissent créer des termes adéquats dans leur propre langue.

La traduction des termes du lexique d'Internet est un problème qui se pose chaque jour un peu plus en raison du développement et de la diffusion incessants et rapides de cette technologie. Le premier problème auquel le traducteur est confronté est l'identification des unités lexicales dénommant les concepts du domaine. Peu d'indices permettent d'identifier une séquence lexicale comme un nom composé, une collocation ou un syntagme de discours. Deuxième tâche difficile, la mise en relation des termes anglais avec leurs équivalents de traduction lorsqu'ils existent ou la création d'équivalents de traduction. Le troisième obstacle à franchir pour le traducteur est l'acquisition de connaissances sur les concepts désignés par des composés.

Internet a besoin de traducteurs puisque les compétences en anglais des internautes et leur capacité à traduire sont inégales, tout comme l'est leur maîtrise de l'outil Internet. L'objectif de notre recherche est de tenter de répondre à certains besoins des traducteurs. La terminologie peut se mettre au service de la traduction spécialisée en matière de composition nominale. Assister le traducteur spécialisé dans le domaine d'Internet exige de répondre à quelques interrogations :

- Quelles sont les particularités morphosyntaxiques et sémantiques des noms composés anglais, français et espagnols ?

Dans le cadre de la traduction de documents d'Internet, le traducteur travaille de l'anglais vers sa langue maternelle (le français et l'espagnol ici). Une connaissance approfondie des composés anglais lui permettra de les découper, de les comprendre et de les reformuler plus aisément. Mieux comprendre la formation des composés français et espagnols du domaine peut faciliter la création d'équivalents de traduction pour les termes anglais à venir.

- Quel est le devenir des noms composés anglais en français et en espagnol dans le domaine d'Internet ?

L'objectif que nous nous sommes assigné est de savoir comment il faut traduire les noms composés anglais en français et en espagnol. Pour le découvrir, les composés anglais existants et leurs équivalents de traduction français et espagnols seront analysés. Nous déterminerons s'ils sont rendus par des composés, des mots simples, des mots dérivés, etc. et dans quelles proportions. L'originalité de ce travail résidera dans la conception d'équivalences types anglais-français et anglais-espagnol. Nous entreprendrons ensuite l'informatisation des équivalences pour élaborer un outil d'aide à la traduction des noms composés de l'anglais vers le français et vers l'espagnol. Le système d'équivalences types sera conçu pour les traducteurs, les terminologues, les professionnels d'Internet, les internautes et les étudiants hispanophones et francophones.

Le choix de la méthodologie est dicté par les besoins du public auquel s'adresse cette étude.

Partie I. La première partie exposera le cadre théorique du présent travail. L'objectif

est de cerner l'objet d'étude, le nom composé, et de préparer le terrain pour les analyses morphosyntaxiques et sémantiques. Le **chapitre I** répondra à deux interrogations : Qu'est-ce que la composition ? Qu'est-ce qu'un nom composé ? Au **chapitre II**, les critères permettant traditionnellement de distinguer les noms composés des autres unités lexicales seront examinés. Dans le **chapitre III**, des typologies créées pour les noms composés de la langue générale et de certains domaines de spécialité seront analysées dans le but de disposer de modèles de classification.

Partie II. Cette partie dressera le profil linguistique des noms composés anglais, français et espagnols du lexique d'Internet pour fournir au traducteur les informations utiles à leur traduction. Le **chapitre I** sera consacré à la constitution du corpus qui servira à la collecte des termes, collecte dont le résultat sera présenté sous la forme d'une typologie morphosyntaxique des noms composés anglais, français et espagnols recueillis. Avec le **chapitre II** débutera l'analyse du profil linguistique des noms composés. Ce chapitre, consacré à l'examen des composés de la langue source, l'anglais, présentera une organisation tripartite : le profil morphosyntaxique des noms composés recueillis, leur profil sémantique et la combinatoire intermatricielle constatée. Dans les **chapitres III et IV**, les lexiques français et espagnol seront analysés.

Partie III. Cette partie constituera l'aboutissement du projet. Le **chapitre I** plantera le décor théorique et méthodologique essentiel à la conception des équivalences types. Le **chapitre II** développera la méthode de création des équivalences types, les résultats obtenus et l'identification des exceptions. Il reposera sur une étude des composés anglais et de leurs équivalents de traduction français et espagnols. L'intérêt que présenterait l'informatisation de notre système pour le traducteur sera ensuite considéré. Dans le **chapitre III**, l'automatisation des équivalences types sera entreprise. L'étude sera couronnée par la présentation de l'outil final.

Les noms composés anglais français et espagnols du domaine d'Internet :

Première partie. Les notions de composition et de nom composé

Chapitre 1. Qu'est-ce que la composition ? Qu'est-ce qu'un nom composé ?

La présente recherche porte sur les termes composés désignant des notions du domaine d'Internet. Le flot de littérature anglaise, française et espagnole sur la composition témoigne de l'utilité pour les lexicographes et terminographes de recourir à ce concept. Comme le souligne G. Gross (1996 : 8), le terme *composé* est polysémique et prête à confusion. C'est pourquoi, dans un premier temps, nous nous livrerons à un travail définitionnel. Seront analysées les définitions du mot composé et du nom composé mais aussi celles de certains mots ou groupes de mots pouvant être confondus avec des composés (mots simples, mots dérivés, collocations⁵ et séquences fortuites⁶). Dans un

⁵ « La distinction entre termes et collocations de la langue spécialisée reste parfois difficile à opérer [...] » (Maniez, 2002 : 359).

⁶ D'autres dénominations existent pour *séquence fortuite* : *syntagme*, *syntagme de discours*, *syntagme non lexicalisé*, *syntagme occasionnel*, *syntagme libre*.

deuxième temps, une étude des propriétés accentuelles, graphiques, distributionnelles, sémantiques et fréquentielles des noms composés aura pour but de tenter de les distinguer de tout autre groupe de mots présentant des similitudes.

1. Le nom composé

Notre analyse définitionnelle sera basée sur des définitions de Tournier (1985, 1991a, 1991b) – élaborées pour l'anglais, mais appropriées aux travaux lexicologiques portant sur le français et l'espagnol. Deux raisons majeures sont à l'origine de cette décision. Premièrement, à l'instar de Gross (1996 : 3), nous avons constaté que les définitions du nom composé sont contradictoires. L'élaboration d'une définition ‘unique’ du nom composé à partir de traits conceptuels tirés de différentes définitions serait vouée à l'échec. Les approches des auteurs étant divergentes, leurs définitions ne sont pas superposables. L'assemblage de traits conceptuels d'origine diverse nous plongerait dans une grande confusion.⁷ Les travaux de Marchand (1969), Catach (1981), Darmesteter (1872) et Kocourek (1980) sont une bonne illustration de la diversité des approches :

- A l'aide de trois exemples, EN *steamboat*, *colorblind*, *rewrite*, Marchand (1969)⁸ explique que les composés sont des « expansions », c'est-à-dire des combinaisons AB dans lesquelles B (*boat*, *blind*, *write*) est un morphème libre. L'élément A peut être employé seul (*steam*, *color*) ou non (*re-*). Dans ce dernier cas, Marchand parle de lexie⁹ préfixée. Composés et préfixés sont des sous-ensembles de la classe des « expansions ».
- La définition de Catach (1981)¹⁰ aborde simultanément les aspects formels, sémantiques, grammaticaux et la perception du locuteur. La définition de l'auteur est ambiguë : les affixes qui forment les dérivés ne sont-ils pas des « éléments morphologiques » ? Les dérivés sont-ils un type de composés ? D'autre part, si les éléments de FR *vinaigre* ou de *gendarme* ne sont plus sentis par les locuteurs, ce n'est pas le cas des formants *d'acides gras* ou *d'eau-de-vie*.

⁷ Pour Darmesteter (1967, 1^{ère} édition 1872)¹¹ et Guilbert (1971b, 1975), la composition appartient à la syntaxe et non à la formation lexicale. Un mot composé est une « proposition en raccourci » (Darmesteter, 1967 : 5). Par exemple, Guilbert (1975), Marchand (1967-1969). Quirk et al. (1972).

⁸ Darmesteter distingue trois types de compositions : a) la juxtaposition qui est une « When two or more words are combined into a morphological unit [...] we speak of a compound » (Marchand, 1969 : 11). « réunion de termes » (FR *coffre-fort*, *platond*, *blanc-bec*), b) la composition elliptique

⁹ qui est une « union intime de mots » (FR *timbre-poste*, *oiseau-mouche*) et c) la Voir Part. I, Ch. 1, Sec. 1.2. pour la définition du terme lexie.

¹⁰ Pour Catach, un composé est un « mot formé d'éléments morphologiques ou lexicaux, à l'origine distincts, tendant vers l'unité sémantique et grammaticale, unité qu'ils finissent en général par atteindre au point que les composants ne sont plus sentis » (1981 : 16).

¹¹ « Le français, en effet, comme les autres langues romanes, possède un certain nombre de procédés qui lui permettent d'unir entre eux deux ou plusieurs termes, et de les faire servir sous cette forme nouvelle à l'expression de faits nouveaux » (Darmesteter, 1967 : 1).

composition par particules, c'est-à-dire au moyen d'adverbes ou de prépositions (FR *contredire*, *confrère*). Ce dernier type de formation est généralement appelé dérivation et est différencié de la composition.

Kocourek (1991b) place la composition¹² en rapport avec la dérivation et la confixation¹³ mais, pour lui, elle est surtout liée à la lexicalisation :¹⁴ « il n'y a parfois que la différence graphique entre un mot composé et un syntagme lexical » (1991b : 131). Par exemple, FR *eau-de-vie* et *lave-vitre* sont des composés, *niveau de vie* et *lavage de cerveau* des syntagmes lexicalisés. Kocourek reconnaît que la graphie fluctue (présence/absence du trait d'union) mais le trait d'union est, selon lui, basé sur une particularité sémantique du composé (caractère figuré, exocentricité, etc.). L'auteur va jusqu'à établir une taxinomie : « termes syntagmatiques » (ou « termes-syntagmes », « syntagmes terminologiques », c'est-à-dire des syntagmes lexicalisés ayant une définition spécialisée) et « termes lexématiques » (ou « termes-mots »). Tous deux forment la terminologie. Les syntagmes libres sont des assemblages « non-lexicalisés, contingents, fortuits, accidentels » (1991b : 136).

Le caractère flou et même contradictoire des définitions du nom composé apparaît encore plus clairement lorsque des termes de construction similaire ont selon les auteurs un statut de composés ou de dérivés. À titre d'exemples, EN *autograph* est un composé néoclassique pour Adams (1973), FR *autocuiseur* est un composé pour Catach (1981) alors que EN *nanometre* est un mot affixé pour Tournier (1991b). Or, ces trois unités lexicales sont formées d'un morphème lié d'origine classique et d'un morphème libre.

Deuxièmement, en recourant à la terminologie d'un seul auteur, nous éviterons le flottement terminologique. Les auteurs emploient généralement une terminologie qui leur est propre, aussi plusieurs dénominations¹⁵ sont-elles parfois utilisées pour un même concept :

Kocourek (1991b : 136-137) donne un aperçu des différentes appellations

¹² « Il y a ~~synonymiques~~ des syntagmes lexicalisés dans les syntagmes selon leur continuité (longueur et complexité) (Rey), synapsie (Benveniste), paratème (Debove) et [...] » (1991a : 129) mot complexe (Picoche), syntagme codé (Rey), groupe lexical (Vachek), lexie complexe (Pottier), lexème complexe (Goffin), groupe lexicalisé (Rey), locution composée (Bally), syntagme lexical (Bonnard), syntagme autonome (Guibert), syntagme lexicalisé (Rey-Debove), syntagme dénominatif (Quemada), lexème

¹³ « Si deux mots (Dufy) rassemblent toutes leurs parties, syntagme (dans ce cas il s'agit d'un mot confié, ou syntagme (1991c : 12) syntagmatique (Lyons), composé syntagmatique (Auger et al.), dénomination

¹⁴ Pour Kocourek, la lexicalisation (ou formation syntagmatique) est la formation des unités lexicales à plusieurs mots (des syntagmes lexicaux) ». (1991b : 135) Personne ne nous emploiera la lexicalisation comme pour désigner la stabilisation d'une unité lexicale au sein d'une communauté comme dénomination d'un processus. L'intégration d'un lexème dans un dictionnaire est l'aspect le plus évident de la lexicalisation. Elle dépend de la fréquence d'emploi du néologisme mais aussi de l'importance dans la communauté de langage qui l'emploie pour la première fois et de l'influence des médias. On utilisera *lexicalisation* pour désigner le processus et *lexicalité* pour dénommer le résultat : « La lexicalité est le fait, pour un mot ou suite de mots, de Gross (1996 : 5) cite deux exemples de flou terminologique : a) la synapsie de

constituer un signe de la langue, c'est-à-dire l'association stable d'un signifiant et d'un signifié [...] ». (Arnaud, 2003 : 20)

¹⁵ La dénomination est « [...] l'inscription d'un signe dans le lexique d'une langue » (Arnaud 1997 : 308). Le terme *dénomination* renvoie également à l'acte de dénomination. Les termes *dénomination*, *désignation* et *appellation* sont synonymes dans notre étude.

Benveniste qui est une « unité de signification composée de plusieurs morphèmes lexicaux » comme FR *machine à coudre*. Cette définition, correspondant à la définition traditionnelle du nom composé, est opposée au mot composé (*timbre-poste*) et au dérivé (*ferblanterie*), b) le synthème de Martinet qui est une « séquence formée de plusieurs monèmes lexicaux fonctionnant comme une unité syntaxique minimale ». Martinet considère comme des synthèmes les dérivés tels FR *désirable*, *refaire*.

1.1. Composition, mot composé et nom composé

C'est donc dans un souci de clarté que cette partie réunit les définitions de Tournier de toutes les unités présentant des points communs avec les noms composés. La connaissance des propriétés caractéristiques de chaque type d'unités devrait permettre l'identification des composés, au moins aux fins de ce travail.

Tournier définit la composition, le mot composé et ses dérivés (noms, adjectifs, verbes composés) dans plusieurs ouvrages : *Introduction descriptive à la lexicogénique de l'anglais* (1985), *Précis de lexicologie anglaise* (1991a), *Structures lexicales de l'anglais* (1991b). La composition, l'une des douze matrices lexicogéniques¹⁶ de formation et de développement du lexique, est le processus de formation des mots composés. Un mot composé est :

un mot formé de plusieurs éléments, dont au moins deux peuvent, par ailleurs, être utilisés de façon autonome [...]. Le mot composé doit être distingué de la séquence fortuite [...]. (1991b : 38)

Cette définition est enrichie de quelques exemples révélant la variété des mots composés anglais (*house-dog*, *doghouse*, *cat's eyes*, *centre of gravity*, *red-hot*, *hand-made*, *made by hand*, *stage-manage*, *look down*). Étudier ces exemples permettra de comprendre la signification de « élément » et « autonome ». L'appellation élément renvoie à des formants de nature variée : des noms (*dog*, *house*, *centre*, *hand*, etc.), des adjectifs (*red*, *hot*, *absent*), des prépositions (*of*), des adverbes (*down*) et des verbes (*look*). Tous ont pour point commun d'être autonomes.

Il apparaît, à la lumière des exemples, que trois graphies sont possibles pour écrire les composés. Leurs éléments constitutifs sont accolés (*doghouse*), reliés par un trait d'union (*red-hot*, *house-dog*) ou séparés par un blanc (*centre of gravity*, *look down*). La nature grammaticale des composés est, elle aussi, variée. Peuvent être composés des adjectifs, des adverbes, des verbes et des noms. Cette diversité se retrouve en français et en espagnol. Les adjectifs composéssont formés de diverses manières. EN *red-hot*, FR *aigre-doux* et ES *socioeconómico*, azul marino sont des assemblages de deux adjectifs ; EN *hand-made* FR *fait main*, ES *bien educado* sont les combinaisons d'un nom et d'un participe passé et EN *made by hand*, FR *commis par erreur*, ES *hecho a mano* sont les associations d'un participe passé et d'un nom relié par une préposition. EN *stage-manage*, FR *chamleverb* sont des verbes composés construits avec un verbe et un

¹⁶ « On appelle matrices les douze processus fondamentaux de formation et développement du lexique : préfixation, suffixation, dérivation inverse, composition (par juxtaposition), amalgame, onomatopée, conversion, métaphore, métonymie, troncation, siglaison, emprunt » (Tournier, 1991b : 115-116).

nom ES *maldecir* comprend un verbe et un adverbe. Les adverbes EN *in fact*, FR *en fin de compte* et ES *al fin y al cabo* sont formés de noms, de prépositions et d'articles.

En ce qui concerne le nom composé, Tournier cite des exemples formés par quatre constructions : dans EN *house-dog* et *doghouse* deux noms sont reliés par un trait d'union ou accolés (FR *avion-citerne*, ES *baloncesto*), EN *cat's eyes* est formé de deux noms reliés par le génitif et EN *centre of gravity* est une combinaison de deux noms reliés par une préposition (FR *centre de gravité*, ES *centro de gravedad*). Tournier parle de modèle germanique lorsque le déterminant d'un composé anglais précède le déterminé et que la relation n'est pas exprimée par une préposition (EN *access control*) et de modèle roman lorsque le déterminant suit le déterminé et qu'ils sont reliés par une préposition (EN *denial of service*) (1991b : 84).

Les composés néo-classiques se démarquent des composés cités jusqu'à présent, c'est-à-dire de ceux dont les éléments peuvent être utilisés de façon autonome. La principale caractéristique des composés néo-classiques réside dans la nature de leurs éléments que Tournier appelle *quasi-morphèmes*. Ces composés comportent « au moins deux quasi-morphèmes (ou *combining forms*) » (1991b : 127). Les quasi-morphèmes (q.m.) sont « des éléments du lexique qui ne peuvent apparaître que combinés avec d'autres. [...] Les q.m. se combinent de préférence entre eux [...] et ont la possibilité d'apparaître aussi bien à gauche qu'à droite [...] » (1991b : 151-152). Leur contenu sémantique est semblable à celui des lexies simples. Tournier cite l'exemple de EN *rhododendron* dans lequel l'élément grec *rhodo* a le même sens que le mot anglais *pink*, et l'élément *dendron* a la même signification que l'anglais *tree*. *Rhododendron* équivaut donc à une formation composée non réalisée¹⁷ de la langue anglaise, *pink tree*. Les quasi-morphèmes, qui sont dépourvus d'autonomie, peuvent se situer à gauche ou à droite, comme les éléments autonomes des composés : EN *psycho path*, *path ology*, *phono gram* et *gramo phone*. Le français et l'espagnol partagent les propriétés énoncées pour l'anglais. Si l'on reprend l'exemple de l'auteur, *rhodon* peut être remplacé par *rose* en français et par *roso* en espagnol et *dendron* par *arbre* et par *árbol*. FR *arbre rose* et ES *árbol roso* font partie du lexique potentiel du français et de l'espagnol. Les quasi-morphèmes se combinent aussi entre eux et peuvent avoir une position préfixale ou suffixale : FR *psycho pathe / path ologie*, ES *psicó pata / pato logía*, et FR *phono gramme / gramo phone*, ES *fonó grafo / gramó fono*.

Les quasi-morphèmes doivent être différenciés de ce que Kocourek (1991 : 129-133) appelle les *pseudo-confixes*. Ces derniers étant très présents dans la terminologie d'Internet, il est utile d'introduire cette notion dès à présent, même si elle n'est pas abordée par Tournier. Les lexies pseudo-confixées sont formées de mots abrégés, formellement semblables à des confixes – c'est-à-dire des racines savantes liées –, et de lexies autonomes. Ce 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 ni des confixes (ce sont des troncations et non des racines grecques ou latines), ni des quasi-morphèmes (ils sont associés à une

¹⁷ L'ensemble des lexies possibles mais non réalisées d'une langue constitue le lexique potentiel. Les lexies réalisées d'une langue constituent le lexique réel. (Tournier 1991b : 111).

base et non combinés entre eux). Dans la présente étude, ces unités lexicales proches des amalgames (voir *infra* p. 13) seront appelées *composés pseudo-confixés*. Voici quelques exemples de pseudo-confixes présents dans le lexique d'Internet : EN *cyber-* (troncation de *cyberspace* > *cybercrime*), *web-* (troncation de *World Wide Web* > *website*), FR *hyper-* (troncation de *hypertexte* > *hyperlien*), *-nauta* (troncation de *astronaute* > *entreprenaute*), ES *e-* (troncation de *electrónico* > *e-compra*), *-nauta* (troncation de *astronauta* > *jubilnauta*). Les composés pseudo-confixés contiennent un seul élément tronqué (EN *cybercafé* < *cyber* (*cyberspace*) + *café*) et les amalgames en possèdent deux (EN *modem* < *modulator* + *demodulator*).¹⁸

Lorsqu'une combinaison rassemble entre trois et cinq éléments autonomes – ou non autonomes dans le cas des quasi-morphèmes et des pseudo-confixes –, Tournier parle de surcomposition¹⁹ (*ibid.* : 38) : EN *electrocardiography*, *biozoology*. L'auteur classe les surcomposés à trois éléments selon l'ordre déterminant-déterminé de leurs formants. Le déterminé est l'élément sémantiquement fondamental du composé et le déterminant est l'élément secondaire qui précise le sens du déterminé :²⁰

- AB + C (C + AB en français et espagnol). Le composé AB est le déterminant de l'élément C comme dans EN [[*ad-blocking*] *program*], FR [*fournisseur d'*[*applications hébergées*]], ES [*bloqueador de [pop-up]*]],
- A + BC (BC + A en français et espagnol). L'élément A est le déterminant du composé BC comme dans EN [[*electronic [newsletter]*]], FR [[*newsletter*] *électronique*], ES [[*dirección IP*] *pública*],
- A + B + C. A et B déterminent C comme dans EN [[*round*] [*unvarnished*] *tale*], FR [*copie [carbone] [invisible]*] et ES [*red [privada] [virtual]*]

Autre type de composition étudié par Tournier, l'amalgame.²¹ Dans un amalgame, « les signifiants des éléments du composé ne sont pas juxtaposés, mais emboîtés l'un dans l'autre » (1991b :17) comme dans EN *Internet* (*Inter national* + *net work*).

Une autre propriété des composés doit être abordée, la centricité. Un nom composé est appelé endocentrique si « la classe d'objets à laquelle il correspond est la même que celle à laquelle correspond l'élément déterminé » (1991b : 68). Dans le cas contraire, lorsque « la classe d'objets à laquelle il correspond n'est pas la même que celle à laquelle correspond le déterminé (dans la mesure où on peut le distinguer d'un déterminant) » (1991b : 72), le composé est dit exocentrique.²² Une EN *mailbox* est une *box*. *Box* est le

¹⁸ Les composés pseudo-confixés anglais, français et espagnols font l'objet d'une étude détaillée (Part. II, Ch. 2, 3 et 4).

¹⁹ D'autres dénominations existent. Clas utilise la désignation *composés lourds* pour parler des composés longs (1994 : 114).

²⁰ Par exemple, dans EN *mailbox* le déterminé est *box* et le déterminant *mail*.

²¹ Cette notion est ultérieurement analysée en détail (Part. II, Ch. 2, Sec. 3.).

²² Certains auteurs dont Boisson (1980) utilisent la terminologie sanscrite et parlent de *dvandva* (*endocentrique*) et de *bahuvrihi* (*exocentrique*).

« substantif-tête »,²³ *mailbox* est endocentrique. Un EN *black hat* n'est pas un *hat* dans le domaine d'Internet. Autrement dit, le sens du composé est différent de celui de la somme de ses formants : il est exocentrique.

Pour conclure provisoirement, un composé, d'après Tournier,

- comprend au minimum deux éléments. Dans les composés juxtaposés (EN *blind copy*) et syntagmatiques, ou N + prép + N, (FR *moteur de recherche*, ES *gestor de correo*), les éléments sont autonomes. Dans les amalgames et composés pseudo-confixés, il s'agit de réductions de lexies autonomes (EN *motel* = **mo** tor + **tel**). Dans les composés néo-classiques, sont présents des éléments non autonomes, les quasi-morphèmes (EN *phonograph*).
- est formé de noms, d'adjectifs, de verbes, de prépositions, d'adverbes qui peuvent être des dérivés (EN *comparison site*, FR *gestionnaire de fichiers*, ES *administrador web*), des sigles (EN *ADSL connection*, FR *fichier MP3*, ES *protocolo IP*), des emprunts (EN *browser cache*, FR *site underground*, ES *banner publicitario*), des quasi-morphèmes (EN *photography*, *chromolithograph*, FR *psychologie*, ES *democracia*), des pseudo-confixes (EN *web designer*, FR *e-business*, ES *metabuscador*) et des amalgames (EN *ADSL modem*, FR *accès Internet*, ES *módem externo*).
- possède un ordre de détermination. Certains sont endocentriques (EN *chat server*, FR *adresse électronique*, ES *grupo de discusión*) et d'autres exocentriques (EN *black hat*, FR *agent intelligent*, ES *caballo de Troya*). Comme le montrent les exemples, la métaphore (ou la métonymie) interfère avec la centricité.²⁴

Un composé peut être confondu avec une séquence fortuite, c'est-à-dire un groupe de mots occasionnel, ou formé pour les besoins du discours. Des critères ont été mis en place pour tenter de différencier ces deux sortes d'unités :

Il existe des critères d'ordre graphique, sémantique, distributionnel, accentuel et fréquentiel, qui permettent le plus souvent au lexicographe de faire la distinction [entre le mot composé et la séquence fortuite] (Tournier, 1991b. : 38)

Il est maintenant nécessaire de lever les ambiguïtés qui existent entre le mot composé et d'autres unités lexicales ou non qui partagent avec lui certains traits conceptuels (la distribution syntaxique dans un énoncé, par exemple) : le mot simple, le mot dérivé, la séquence fortuite, la collocation et l'expression idiomatique.

1.2. Lexie composée vs. lexie primaire, lexie dérivée, lexie simple, lexie prépositionnelle, lexie complexe, collocation et séquence fortuite

Tournier préfère parler de *lexie* plutôt que de *mot* car « le mot n'est pas une unité suffisamment solide épistémologiquement pour constituer la base d'une description scientifique du lexique » (1991a : 10). Un seul exemple suffit pour souligner la faiblesse

²³ Termen de définition traditionnelle selon laquelle un blanc en anglais (Bn) est « ce qu'il y a entre deux blancs typographiques » (ibid.). L'unité EN *waste paper* contient un blanc typographique, elle est

²⁴ Des critères devront être mis en place pour définir jusqu'à quel degré de ressemblance entre deux concepts un composé sera endocentrique (Part. II, Ch. 2, Sec. 2.1.).

donc formée de deux mots. En revanche, dans EN *wastepaper basket*, les éléments *waste* et *paper* constituent un mot puisqu'ils sont accolés. Des exemples similaires existent en français et en espagnol. FR *eurovision*, ES *telenovela* ne sont ni des mots simples, ni des mots dérivés mais des composés formés de deux mots dont un au moins est tronqué (FR *euro* < *Europe*, ES *tele* < *televisión*). Il y a parfois hésitation entre soudure et trait d'union (FR *portefeuille* et *porte-monnaie*) ou entre trait d'union et absence de lien (FR *petit-four* et *petit four*).

Tournier parle de *lexie* car c'est une unité « souple » (1991a : 11) rendant compte de la nature et de la longueur variées des éléments assimilés au cours de l'apprentissage. Les usagers d'une langue apprennent des éléments plus ou moins longs sans les fractionner (FR *chat*, FR *tirer par les cheveux*, etc.). Une lexie est une « unité lexicale mémorisée » (*ibid.*) dont Tournier propose une classification qui peut nous aider à délimiter le cadre de la création lexicale par composition. *Lexie* est l'hyperonyme dont les co-hyponymes sont *lexie primaire*, *lexie simple*, *lexie dérivée*, *lexie complexe*, *lexie composée*, *lexie prépositionnelle*. En somme, la lexie est une unité lexicale mémorisée allant du mot, pris au sens traditionnel (EN *dog*, FR *chien*, ES *perro*), à la phrase lexicalisée²⁵ (EN *give a dog a bad name and hang him*, FR *celui qui veut noyer son chien l'accuse de la rage*, ES *en dando en que el perro ha de rabiar, rabia*).

1.2.1. La lexie simple

Pour Tournier, la dénomination *lexie simple* regroupe deux sortes de lexies : les lexies primaires et les lexies dérivées. Elles sont toutes deux formées d'un seul élément lexical autonome. Une lexie primaire (EN *boy*, *hole*, FR *nid*, *trou*, ES *ojos*, *vaso*) ne résulte d'aucune des douze matrices lexicogéniques. Une lexie dérivée peut être affixée ou non. Une lexie dérivée affixée comporte un élément autonome, appelé *base*, et au moins un affixe (suffixe ou préfixe). Le processus de la suffixation ou de la préfixation est à l'origine de la formation (EN *social ist*, *un happy*, FR *plâtr ier*, *pré colombien*, ES *protestad or*, *protecc ión*). Deux points distinguent les affixes des quasi-morphèmes. Les affixes, contrairement aux quasi-morphèmes, ne se combinent pas entre eux et ils ont une contrainte d'ordre ; ils se situent à droite ou à gauche alors que les quasi-morphèmes peuvent occuper les deux positions. Une lexie dérivée non affixée ne compte qu'un élément autonome ayant subi un changement de fonction ou de sens.²⁶ Le verbe EN *pocket* est un emploi dérivé du nom *pocket* et EN *bear* dans le sens de 'peu social' dérive de *bear* signifiant 'ours'. Le français et l'espagnol connaissent eux aussi des changements sémantiques : FR *rat* (rongeur / personne avare), ES *pesado* (lourd / ennuyeux, assomant). Une lexie simple, primaire ou dérivée, peut entrer dans la formation de composés (EN *service provider*, FR *fournisseur de services*, ES *proveedor de servicios*).

Aucune confusion n'est possible entre une lexie simple et une lexie composée. Dans

²⁵ « Une lexie lexicalisée est reconnue comme faisant partie du code, c'est-à-dire du système de la langue » (Tournier, 1991b : 105).

²⁶ Tournier explique que *dérivation* comporte deux acceptations. Au sens restreint, *dérivation* et *affixation* sont synonymes. Au sens large, *dérivation* désigne le processus expliqué ici.

la plupart des cas, le nombre d'éléments autonomes diffère : un seul pour la première, au moins deux pour la seconde (composés juxtaposés et syntagmatiques) ou aucun dans le cas des composés néoclassiques²⁷ (ES *cardiólogo*) et des amalgames (EN *brunch* = *breakfast* + *lunch*). Seuls les composés pseudo-confixés contiennent un élément autonome mais ils sont facilement repérables car le nombre de pseudo-confixes est limité dans le lexique d'Internet (moins d'une dizaine dans notre corpus).

1.2.2. La lexie prépositionnelle

Une lexie prépositionnelle est un « syntagme prépositionnel lexicalisé, c'est-à-dire un groupe introduit par une préposition » (1991b : 110), comme EN *with all due respect, with bell, book and candle*, FR *par tous les moyens, sans compter*, ES *al fin y al cabo, con cuenta y razón*. Une lexie composée ou surcomposée peut commencer par une préposition mais elle ne comporte pas de syntagme lexicalisé (ou groupe de mots). Elle n'est pas formée d'une préposition et d'un groupe répertorié mais d'une préposition et d'une lexie simple répertoriée (EN *upgrade, down payment*, FR *sous-main, avant-garde*, ES *no conectado*). Certaines lexies sont formées sur le modèle N + prép + N (EN *peer-to-peer*, FR *serveur de courrier*, ES *gestor de sitio*). La préposition n'étant pas placée au début de la combinaison, ces unités sont des lexies composées et non prépositionnelles. Ces dernières comportent parfois un article, ce qui ne semble pas être le cas des composés. Ces deux types de lexies – même si elles sont formées d'au moins deux éléments autonomes et contiennent une préposition –, sont différenciables par d'autres caractéristiques qui leur sont propres : place de la préposition, présence/absence de syntagme lexicalisé, présence/absence d'articles.

1.2.3. La lexie complexe

Pour Tournier, la lexie étant une unité lexicale mémorisée sans limite de longueur, les unités complexes que sont les citations, les expressions idiomatiques et les phrases complexes peuvent être des lexies. Cette définition de la lexie est donc plus large que ce dont nous avons besoin pour définir le nom composé dans la présente étude.

Une lexie complexe est une lexie dont la « structure syntaxique complexe ne lui permet pas de correspondre à la définition des autres types des lexies » (1991b : 110). Selon Tournier (1991a : 13), sont complexes : certaines expressions idiomatiques (EN *as snug as a bug in a rug*, FR *être bien au chaud*, ES *andar de boca en boca*), certaines phrases complexes (EN *a woman, a dog, and a walnut-tree, the more you beat them the better they be*, FR *quand il n'y en a plus, il y en a encore*, ES *e s una locura amar, a menos de que se ame con locura*) et certaines citations (EN *there is no such thing as public opinion. There is only published opinion*, Winston Churchill, FR *il faut laisser du temps au temps*, Mitterrand, ES *e l arte es una mentira que nos acerca a la verdad*, Pablo Picasso).

Certaines lexies complexes sont modifiées lorsqu'elles sont utilisées en discours (FR

²⁷ Les composés néoclassiques sont surtout présents dans les terminologies médicales (EN *chromolithograph, hysterectomy, osteopath*). Ils ne seront donc pas pris en compte dans la présente étude.

he caught a cold last week, FR il est bien au chaud dans son lit). Ce n'est pas le cas des lexies composées²⁸ (EN *John's Internet provider is expensive*, ES *el buzón de correo de Javier está lleno*, FR *Google est un moteur de recherche performant*). Les autres ne subissent pas de transformation mais sont reconnaissables car introduites par des expressions comme FR *comme l'a dit X..., il est connu que..., EN X once said..., it is well known that..., ES según X..., como lo decía X... : FR comme l'a dit Mitterrand, « il faut laisser du temps au temps », EN “There is no such thing as public opinion. There is only published opinion” Winston Churchill once said, ES como lo decía Pablo Picasso,« e l arte es una mentira que nos acerca a la verdad ».*

La structure interne des lexies complexes citées ci-dessus est très éloignée de celles des lexies composées. Ces dernières sont des assemblages dépassant rarement cinq à six éléments (EN *HyperText Transfer Protocol, Transmission Control Protocol/Internet Protocol*) contrairement aux lexies complexes qui sont des phrases (EN *a woman, a dog, and a walnut-tree, the more you beat them the better they be*). Une même lexie complexe comporte parfois des pronoms personnels, plusieurs verbes et plusieurs articles et/ou prépositions (ES *e l arte es una mentira que nos acerca a la verdad*) alors qu'un composé ne comporte que plusieurs prépositions (FR *fichier de signature de virus*, ES *proveedor de acceso a Internet*).

Tournier (1991b : 72) ajoute que les dénominations *lexie pr épositionnelle* ou *lexie complexe*, selon le cas, doivent remplacer *expression figée*. Une expression n'est pas figée et parler d'*expression* est maladroit. En discours, ces deux types de lexies sont modifiables par un locuteur (EN *as snug as a bug in a rug* peut être détourné en EN *as snug as a dog on a rug*).

Tournier classe les lexies décrites en 1.2.1., 1.2.2. et 1.2.3. dans deux catégories :

- le mot :²⁹ lexies simples – lexies dérivées et lexies primaires – et lexies composées,
- la phraséologie : lexies prépositionnelles et lexies complexes.

Pour conclure, à l'instar de l'auteur, nous utiliserons *mot* comme équivalent de *lexie simple* ou de *lexie composée*. Dans notre travail, *mot* n'aura jamais le sens de la définition traditionnelle imprécise vue précédemment. Connaître la définition de chacune des unités lexicales suffisait pour les distinguer des composés. Avec les notions de collocation et de séquence fortuite la différenciation s'avère quelquefois plus délicate.

1.2.4. La séquence fortuite

Tournier présente la séquence fortuite comme un groupe de mots assemblés en discours (EN *blue towel*, FR *chaise confortable*, ES *libro apasionante*). Nous compléterons nos données avec les travaux de Lehmann et Martin-Berthet (1997) qui définissent le composé par rapport au syntagme. *Syntagme* est hyperonyme de *composé*. Un syntagme est un groupe de mots et un composé est un groupe de mots possédant des

²⁸ Les lexies composées présentent toujours les particularités des marques graphiques (soudure, trait d'union) et morphosyntaxiques.

Comme le soulignent Lehmann et Martin-Berthet, la soudure concerne les composés

²⁹ « Rien ne s'oppose à ce qu'on les appelle toutes [les lexies] des mots, dès l'instant où l'emploi de ce terme est ainsi précisé à partir de la notion de lexie » (Tournier, 1991a : 13).

savants (FR *biographe*, *philanthrope*), des composés anciens démotivés (FR *bonhomme*, *embonpoint*, *vinaigre*) et des composés récents ayant une lexie tronquée³⁰ et présentant l'ordre déterminant-déterminé (FR *reprographie* = *repro* < *reproduction* + *graphie*). Quant au trait d'union, il n'est pas systématique et semble plus fréquent dans la structure V + N (FR *abat-jour*). L'orthographe de composés construits avec une même lexie est fluctuante : FR *enjeu* et *en-cas*, *eau-de-vie* et *eau de rose*, *télé-achat* et *téléachat*. Comparé à une séquence fortuite, le composé a des spécificités :

- morphosyntaxiques internes. Lehmann et Martin-Berthet citent les exemples de FR *rouge-gorge* dans lequel l'adjectif est antéposé et le composé comporte le genre masculin (le genre du nom du générique *oiseau*) et de *chaise longue* dans lequel la postposition de l'adjectif *longue* n'est pas conforme pas à la syntaxe de l'adjectif *long* (*c'est un long chemin* et non **c'est un chemin long*).
- distributionnelles. Dans la phrase FR *Jeanne est très fleur bleue*, c'est la nominalisation du syntagme *fleur bleue* qui dénote la composition. La lexie composée a la même forme que la séquence fortuite *fleur bleue* dans *regarde cette magnifique fleur bleue*.
- morphosyntaxiques internes et distributionnelles. FR *essuie-glace*, de type V + N, a une fonction nominale et non verbale au sein de la phrase, ne comporte pas de déterminant devant le nom *glace* (*un *essuie la glace*), et *essuie* est invariable (*des essuie-glaces*).

Certains composés ne sont pas marqués par la graphie et/ou la morphosyntaxe. FR *grand magasin* et *pomme de terre* sont analogues à des syntagmes non lexicalisés (structure et distribution identiques). Des critères linguistiques tentent d'évaluer le figement qui en fait des « signes compacts ». Toutefois, une étude de Bauer (1998) diminue grandement leur portée (Part. I, Ch. 2).³¹ Sur le plan conceptuel, les composés se distinguent des séquences fortuites. Seuls les composés renvoient à des concepts précis. FR *grand magasin* renvoie à un concept (/établissement de commerce rassemblant les produits de différents commerces, présentés dans des rayons spécialisés/) alors que FR *petit magasin* renvoie à deux concepts (/taille ou volume inférieur à la moyenne/ et /établissement de commerce/).

1.2.5. La collocation³²

« La collocation d'un mot est son contexte immédiat » (Tournier, 1991b : 34). L'analyse d'une lexie, dans une série d'énoncés, révèle que

³⁰ Dans le cadre du lexique d'Internet, il faut ajouter les amalgames et les composés pseudo-confixés.

³¹ Les critères linguistiques étant également appliqués aux collocations, leur étude est reportée au Ch. II.

³² Nous nous intéresserons uniquement aux collocations A + N et N + A dans la mesure où les autres types de collocations ne présentent pas d'ambiguités avec les noms composés (N + V, FR *un doute subsiste* ; V + prép + N, FR *être rongé de doutes* ; ou V + N, FR *éprouver un doute*, par exemple).

« *tel mot n'apparaît jamais à proximité immédiate de tel autre (collocation non attestée), ou bien qu'il apparaît quelquefois (collocation fortuite), souvent (collocation fréquente) voire obligatoirement (collocation nécessaire)* » (*ibid.*).

Cette définition est inopérante pour établir une distinction entre composés et collocations. Avec un tel continuum, comment différencier une collocation fréquente et nécessaire d'un composé ? Tournier cite l'exemple EN *fast bowler*. Dans les commentaires de matches de cricket l'adjectif *fast* précède souvent le nom *bowler*. Le groupement *fast bowler* n'est pas répertorié dans les dictionnaires comme nom composé. Selon l'auteur, étant donné la fréquence du groupement, ce serait un composé en voie d'intégration. Néanmoins, en 2005 *fast bowler* n'est toujours pas répertorié dans le GDT, ni dans le MWOD qui répertorie *pace bowler*. Le CALDO cite l'exemple *fast / pace bowler*. Les travaux de Hausmann (1979), basés sur ceux de Bally (1909),³³ semblent plus appropriés à l'établissement de la distinction composé/collocation :

« *Il y a série ou groupement usuel lorsque les éléments du groupe conservent leur autonomie, tout en laissant voir une affinité évidente qui les rapproche, de sorte que l'ensemble présente des contours arrêtés et donne l'impression de déjà vu* ». (Bally in Hausmann 1979 : 189)

Deux exemples illustrent cette définition : FR *gravement malade* et *grièvement blessé*. Malgré leur signification proche, les adverbes *grièvement* et *gravement* ne sont pas interchangeables (**grièvement malade* et ?*gravement blessé*). Pour répondre à son objectif – la création d'un dictionnaire des collocations –, Hausmann doit affiner cette définition. Il faut différencier les combinaisons usuelles de celles qui ne le sont pas. Le nombre de combinaisons usuelles étant considérable, cela suppose que le corpus à constituer sera gigantesque. L'auteur cite l'exemple de FR *arbre*. *Abattre* ou *couper* sont des cooccurrences usuels mais qu'en est-il de *regarder* et *mesurer* ? S'ils ne sont pas moins usuels, il est peu probable qu'un locuteur cherche *regarder* ou *mesurer un arbre* dans un dictionnaire.

Pour Hausmann (1979 : 191), les collocations relèvent de la langue en tant que norme dans la mesure où « il y a une contrainte plus ou moins arbitraire des mots sur les mots ». Des comparaisons entre les langues permettent de révéler cette contrainte. En français, on parle de *luxe princier* et, en anglais, de *dazzling luxury* qui est équivalent³⁴ à FR **luxue éblouissant/aveuglant* (FR *démenti formel* et EN *flat denial*). Une collocation est l'assemblage d'une base (le nom) et d'un collocatif (l'adjectif) qui est sous contrainte et orienté. Le signifié de la base est autonome. Il n'a pas besoin du collocatif pour être défini. En revanche, le collocatif ne se réalise entièrement que combiné à une base. Dans la collocation FR *célibataire endurci*, la base *célibataire* est sémantiquement autonome alors que le collocatif *endurci* a besoin d'une base – *célibataire, pécheur* – pour être complètement défini :³⁵

³³ Bally établit une distinction, au niveau syntagmatique, entre les « associations libres et occasionnelles, les séries phraséologiques ou groupements usuels et les unités indissolubles » (Hausmann, 1979 : 189).

³⁴ Des termes équivalents désignent des concepts homologues dans plusieurs langues (Thoiron et al., 1996 : 512). Voir Part. III, Ch. 1, Sec. 1.1.2. pour la notion d'équivalence en terminologie.

célibataire adj. Fl n

• 1711; de *célibat*

1 • Qui vit dans le célibat. *H* est *célibataire* — Propre au célibataire « les habitudes *célibataires* » (Queneau)

2 • N Personne qui vit dans le célibat → garçon, fille. *Une célibataire* « l'équivalence réjouie d'un vieux célibataire » (France). C'est un *célibataire endurci*. *Célibataire en concubinage*³⁵

notoire.

○ Appos. *Mère célibataire* (remplace *fillette-mère*).

□ Fam. *Je voyage en célibataire*, sans conjoint. (PR 1996)

endurci, ic adj.

• XIII⁸; de *endurer*

1 • Qui est devenu dur, insensible. ⇒ fam. blindé. *Cœurs endurcis*. « Je n'ai jamais vu d'âme aussi endurcie que la vôtre » (Camus)

• Qui avec le temps s'est fortifié, figé dans son opinion, son occupation. → invétéré. « Ces francs pecheurs, pecheurs endurcis, pleins et achavés » (Pascal). Un *célibataire endurci*. « Ses instincts de rond de cuir endurci » (Courceline). (PR 1996)

Comme le montrent les deux définitions, le collocatif *endurci* ajoute une qualité à la base *célibataire*. Les collocations permettent d'acquérir une bonne maîtrise de la langue, d'où l'importance pour les usagers non natifs d'une langue d'assimiler les collocations consacrées par l'usage. L'adjectif *endurci* dans *célibataire endurci* ne signifie pas « devenu dur, insensible » mais « figé dans son opinion ». Heid et Freibott, qui voient la collocation comme une « combinaison polaire non arbitraire de deux lexèmes qui a un caractère conventionnel à l'intérieur d'un groupe linguistique » (1991 : 78), partagent l'avis d'Hausmann : les collocations doivent être apprises. Les locuteurs considèrent que les collocations sont des « produits semi-finis »³⁶ puisqu'elles ont un caractère conventionnel et sont souvent répertoriées.

En conclusion, composés et collocations se ressemblent dans la mesure où ce sont des groupes de lexies simples qui doivent être appris et répertoriés dans les dictionnaires. En revanche, ces deux types de groupements se différencient dans le sens où les éléments d'un composé, le déterminé et le déterminant, sont définissables isolément, contrairement au collocatif d'une collocation qui est lié à la base. D'autre part, dans les dictionnaires la collocation est fréquemment présentée dans l'article consacré au collocatif et non à la base (Hausmann 1979). Le composé est ordinairement répertorié dans l'article du déterminé. Par exemple, dans le PR 1996 la collocation FR *ignorance crasse* figure dans l'article du collocatif *crasse* alors que le composé *courrier électronique* est défini à l'adresse du déterminé *courrier*.

En somme, un double test à partir du PR 1996 pourrait peut-être déterminer si un groupe de mots est un composé ou une collocation. Si le groupe est répertorié à l'adresse

³⁵ Dans les deux définitions, les acceptations correspondant à la base et au collocatif sont en caractères gras.

³⁶ La dénomination *produit semi-fini* est utilisée par opposition aux combinaisons de lexies qui relèvent du « hasard du texte », c'est-à-dire non lexicalisées.

du substantif et si la définition de l'adjectif ne dépend pas de celle du substantif, alors le groupe est un composé et non une collocation. Pour connaître la fiabilité de ce test, nous allons l'appliquer aux vingt-et-un groupements qui sont des collocations pour Haussmann (1979 : 189) et voir s'ils sont effectivement répertoriés à l'adresse de l'adjectif et si la définition de ce dernier dépend de celle du nom : *démenti formel*, *ferme résolution*, *vif intérêt*, *activité fiévreuse*, *soin minutieux*, *attention soutenue*, *agitation fébrile*, *preuve irréfutable*, *refus catégorique*, *liens indissolubles*, *sanglant outrage*, *profonde reconnaissance*, *envie folle*, *rapidité vertigineuse*, *prix exorbitant*, *victoire décisive*, *ignorance crasse*, *laideur repoussante*, *avarice sordide*, *luxe princier*, *répugnance insurmontable*. Le test fonctionne plutôt bien : treize collocations le passent avec succès, dans sept cas la réussite est partielle et il n'échoue que pour une seule unité (voir Annexe 3). Des nuances doivent cependant être apportées. D'abord, le critère dictionnaire est fragile car les dictionnaires ne répertorient pas toujours les groupements de la même manière. Ensuite, ce critère n'est pas toujours suffisant : les termes ne figurent pas dans les dictionnaires de langue générale et de langue de spécialité dès leur création et certains ne feront jamais leur apparition dans les ouvrages lexicographiques (Gile 1995 : 137, Béjoint 2000).

Reprendons le critère conceptuel évoqué lors de la distinction composé-séquence fortuite. Comme on l'a dit, un nom composé correspond à un concept. Une collocation devrait correspondre à l'addition des concepts de la base et du collocatif. Cette règle n'est pas infaillible puisque si les collocations FR *prix exorbitant* et *activité fiévreuse* renvoient à deux concepts (/valeur d'un bien/ + /excessif/ et /occupation/ + /passionné, intense/), on peut hésiter à propos d'*ignorance crasse* qui semble renvoyer à un seul concept (/ignorance dans laquelle on se complaît/). L'hésitation entre un concept et deux concepts pour *ignorance crasse* naît de la multiplicité des utilisations sémantiques du collocatif (crasse peut signifier grossière). Le sens d'une même lexie diffère en fonction du contexte dans lequel elle apparaît. De plus, il existe plusieurs types de variations sémantiques (Cruse, 1986 : 50 et seq.). Dans certains cas, comme EN *Sue is visiting her cousin*, la communication est claire. La lexie EN *cousin* est générale, elle renvoie à FR *cousin* et *cousine*. Dans d'autre cas, comme EN *we finally reached the bank*, la communication est ambiguë. EN *bank* renvoie à la fois à FR *établissement bancaire* et à *berge* (*d'un cours d'eau*). Aucun sens général ne regroupe les deux possibilités. Il est difficile d'établir le sens des lexies polysémiques.

2. Conclusion

Ce premier chapitre ne prétend pas lever les ambiguïtés contre lesquelles les chercheurs se battent depuis des décennies. L'objectif est d'exposer quelles combinaisons seront considérées comme des noms composés au sein de la présente étude. Il met l'accent sur les aspects caractéristiques des noms composés. Confronter les définitions de tous les types de lexies – composées, primaires, dérivées, simples, prépositionnelles, complexes – et celle de la collocations et de la séquence fortuite fait ressortir les particularités des lexies composées. Nos conclusions seront complétées par quelques données nécessaires à l'étude des noms composés du lexique d'Internet.

- La nature
 - Les noms composés sont des lexies formées par la matrice lexicogénique de la composition. Ils se situent dans le cadre de la néologie³⁷ et, plus exactement, de la néologie morphosémantique. En tant que lexies, les noms composés sont des signes linguistiques qui réfèrent à des êtres, objets ou notions du monde extra-linguistique – Internet dans ce travail.
 - Une lexie composée, comme toute autre unité de la langue, entre dans une catégorie grammaticale. Un nom composé a une fonction nominale dans la phrase.
 - Les noms composés sont des noms construits motivés³⁸ (EN *click and mortar*).
 - Les noms composés sont créés par différents types de composition : (a) la juxtaposition (EN *access control*, FR *connexion réseau*, ES *comunidad hacker*), (b) l'amalgame (EN *modem*) et (c) la composition néoclassique (EN *osteopath*). *Composition syntagmatique* ou *composition de type roman* sont les termes employés pour les composés, majoritairement français et espagnols, de type N + prép + N. À ces types, nous ajoutons la composition par pseudo-confixation fréquente dans le lexique d'Internet (EN *e-mail*, FR *entrepreneute*, ES *cibercafé*).
 - Les éléments des composés anglais présentent en général l'ordre déterminant-déterminé (EN *virus attack*). Pour les composés français et espagnols, l'ordre des éléments est habituellement déterminé-déterminant (FR *fournisseur d'accès*, ES *sitio comercial*).

- La forme
 - Les éléments entrant dans la formation des composés sont les lexies simples, c'est-à-dire primaires ou affixées (EN *dot-com industry*, FR *e-commerce*, ES *emot icono*), les lexies composées, c'est-à-dire formées d'au moins deux morphèmes ou lexies (EN *high-speed connection*, FR *serveur de newsgroup*, ES *alojamiento de sitio web*), les sigles, les amalgames et les troncations (EN *IP protocol*, FR *modem externe*, ES *servidor FTP*). Des morphèmes grammaticaux autonomes, des conjonctions et des prépositions entrent parfois dans la formation de composés (EN *down load*, FR *en-tête*, ES *denegación de servicio*).
- ³⁷ « La néologie est l'introduction dans un langage de nouveaux éléments sans rapport avec la tradition de nouveauté lexicales, quelle qu'en soit la nature » (J. Tournier, 1991, p. 125).
- ³⁸ Le nombre d'éléments autonomes offre selon le type de composés. Les composés juxtaposés et syntagmatiques possèdent au moins deux éléments pouvant être autonomes (EN *ad blocker*, FR *logiciel de chat*, ES *intercambio de bienes*). Les composés pseudo-combinés n'en possèdent qu'un seul, l'autre étant la troncation d'un élément autonome (EN *openware*, FR *cyberespace*, ES *e-compra*). Enfin, les composés néo-classiques et les amalgames ne présentent aucun élément autonome (EN *osteopath*, *modem*).
- La motivation est la conscience que peut avoir personnellement un usager de la raison pour laquelle tel signifié correspond à tel signifiant (ou inversement) (J. Tournier, 1991b, p. 122). La motivation de type indirect en matière de composition, peut être morphologique, sémantique ou morphosémantique. Par exemple, il n'y a aucune raison pour que une boîte soit dénommée *box* et le courrier *par mail*. En revanche, une raison de second degré fait que le dispositif destiné à recevoir le courrier à acheminer par le service postal soit désigné par la lexie *mailbox*. Dans la terminologie d'Internet, une analogie de fonction – Les composés de plus de deux éléments, autonomes ou liés, sont des (métaphore) entre la *mailbox* traditionnelle et l'application informatique gérant le courrier électronique fait que le terme *mailbox* surcomposés (EN *Bulletin Board System*, FR *connexion Internet sans fil*, ES est sémantiquement motivé.

cuenta de correo electrónico).

- La graphie des composés anglais, français et espagnols est variable y compris pour un même composé (Part. I, Ch. 1, Sec. 2.). Il existe trois graphies possibles (EN *download*, FR *webzine*, ES *hipervínculo* ; EN *cyber-delinquent*, FR *start-upeur*, ES *e-empleo* ; EN *electronic hoax*, FR *serveur mail*, ES *correo basura*).
- Le sens
 - La présence d'une métaphore modifie quelquefois le sens des lexies constituantes des composés par rapport à leur emploi autonome (EN *cybercrime*, FR *navigation anonyme*, ES *banda ancha*) et interfère avec la centricité.
 - On parle de composés exocentriques lorsque la classe d'objet désignée par le composé est différente de celle désignée par la lexie tête (EN *thumbnail*, FR *agent intelligent*, ES *ciberpunk*) et de composés endocentriques dans le cas contraire (EN *general portal*, FR *fichier journal*, ES *página personal*).

Cette analyse de la composition et du nom composé est essentielle car la composition nominale est un « phénomène complexe » (Gross, 1996 : 29). Savoir quels éléments entrent dans la formation des composés est nécessaire pour identifier les noms composés du lexique d'Internet. Cependant, les propriétés étudiées ne sont pas toujours suffisantes pour distinguer les lexies composées N + prép + N, N + A et A + N des séquences fortuites et de certaines collocations de structure identique. Boisson (1980 : 244) va jusqu'à dire qu'aucune définition rigoureuse ne peut être donnée et qu'il n'existe probablement pas de définition opératoire permettant de savoir si une séquence est un mot composé ou une séquence fortuite. C'est pourquoi, dans un premier temps, nous n'écartons aucun candidat terme³⁹ et plus précisément aucun candidat composé. Tant que tous les types de groupements constitués d'unités lexicales ne seront pas clairement différenciés, seront extraits de nos corpus les assemblages sentis comme des composés (par rapport aux données dont nous disposons). Ceci ne signifie évidemment pas que nous renonçons à tout critère de différenciation. Une batterie de tests sera appliquée aux candidats composés.

Chapitre 2. Composition et ...

Sur le plan de la construction, EN *blacklist*, *black hacker* et *black server* sont identiques : un adjectif précède un substantif. La première séquence est lexicalisée et répertoriée dans les dictionnaires, la deuxième est un terme pas toujours répertorié, la troisième est une séquence fortuite que nous venons de créer et qui n'est donc ni lexicalisée, ni

³⁹ Les candidats termes sont des « mots ou séquences de mots susceptibles d'être retenus comme termes par le terminologue ». (Bourigault et Jacquemin, 2000 : chapitre 9) Source : <http://www.univ-tlse2.fr/erss/membres/bourigault/bourigault-jacquemin.rtf>

répertoriée. Comme nous l'avons précisé, pour recenser des composés il faut autant que possible les distinguer de toutes les autres unités lexicales. Or, l'existence de deux continuums (collocation-lexie composée et séquence fortuite-lexie composée) complique l'identification des composés. Une séquence fortuite (Part. I, Ch. 1, Sec. 1.2.4.) ou une collocation (Part. I, Ch. 1, Sec. 1.2.5.) peut voir sa fréquence d'emploi augmenter et être lexicalisée puis répertoriée (Tournier, 1991a : 66). Cinq critères de distinction sont traditionnellement utilisés pour lever les ambiguïtés : l'accentuation, la graphie, la distribution, la sémantique et la fréquence. Seront ajoutés quelques autres critères utilisés par Arnaud (2003), Boisson (1980), Lehmann et Martin-Berthet (1997). Appliquer ces critères à des unités du lexique d'Internet qui sont de bons candidats composés, des séquences fortuites et des collocations⁴⁰ montre leur manque de fiabilité. Nous considérerons Bauer (1998) pour qui ces critères sont insuffisants pour décider si une séquence anglaise N + N est une construction morphologique (un composé) ou une construction syntaxique (une séquence fortuite). Bauer doute même de l'existence de deux types de constructions séparées.

1. accentuation

L'accentuation est une proéminence d'énergie articulatoire qui se manifeste par une augmentation physique de longueur, d'intensité et éventuellement un changement de fréquences en passant de syllabe inaccentuée à accentuée et/ou au cours de l'évolution de la syllabe accentuée (Léon, 2000 : 107).

L'accent tonique n'intervient pas de la même façon et n'a pas le même rôle en anglais, en espagnol et en français. L'accentuation des composés de chaque langue sera donc étudiée dans une section différente. Le système de marquage de Tournier (1991a) sera utilisé pour interpréter les exemples :

- 0 indique que la syllabe est inaccentuée (ou atone),
- 1 signifie que la syllabe porte l'accent principal,
- 2 signifie que la syllabe porte l'accent secondaire.

1.1. L'accentuation des composés anglais

La règle générale pour l'accentuation des lexies anglaises est : les lexies simples de plus d'une syllabe comportent un accent tonique principal sur une syllabe (1), les autres syllabes sont inaccentuées (0) ou comportent un accent secondaire sur une autre syllabe (2) : *browser* (10), *favorite* (100), *server* (10), *authentication* (02010).

D'après Tournier (1991a : 65), l'assemblage de lexies habituellement indépendantes en séquences fortuites n'entraîne pas de modifications accentuelles. L'auteur cite l'exemple EN a *German diplomat* (10100), pour expliquer que chaque lexie constitutive garde son accent principal. En revanche, combiner des lexies pour former un composé

⁴⁰ Les groupements A + N et N + A testés dans ce chapitre sont considérés comme des collocations par Bally (in Hausmann, 1979).

influe sur l'accentuation des formants qui est, en général, semblable à celle des lexies simples : un seul accent principal, normalement situé sur le premier élément. Selon Adams (1973) et Bauer (1998 : 70), l'accent de nombreux composés porte sur le premier élément, contrairement aux séquences fortuites dans lesquelles il porte sur le deuxième élément. Adams cite les exemples *bo ttleneck* (100) et *bottle's neck* (001), auxquels on peut ajouter *high chair*, *high jump* (12) et *high noon* (21), *high pressure* (210). Mais cette règle a des limites. Elle n'est pas infaillible notamment pour les composés en *self-* qui ne sont pas mono-accentués mais portent un double accent principal (*self-preservation* (12010)) (Tournier, 1991a : 66). Arnaud souligne que « les suites occasionnelles NN, très fréquentes, peuvent être accentuées sur le composant gauche » (2004 : 340).

Pour Bauer (1998), faire une distinction à partir de l'accentuation soulève plusieurs problèmes. Des expériences montrent qu'interroger des locuteurs sur des unités N + N n'est pas concluant puisqu'ils sont en désaccord sur l'accent des unités testées et qu'ils ne sont pas constants dans leurs propres choix. Une alternative serait d'enregistrer des conversations ou des émissions radiophoniques/télévisées. Cette solution présente un inconvénient majeur. Il faudrait une masse considérable d'enregistrements pour obtenir plusieurs occurrences d'un même composé et surtout d'une même séquence libre. Les temps d'enregistrement et d'analyse seraient très longs.

Boisson (1980), dont nous ne donnerons qu'une vue d'ensemble, cherche à déterminer si l'accent des composés anglais est sur l'élément de gauche ou sur celui de droite.⁴¹ Sur les 9 450 composés de son corpus (toutes catégories grammaticales confondues), 64,50% sont des composés gauches et 35,50% sont des composés droits. En considérant uniquement les 7 786 noms composés,⁴² le pourcentage de composés gauches atteint 72%. La gaucherie est normale chez les noms composés. Boisson en déduit que la « gaucherie est une propriété discriminante des catégories grammaticales » (ibid. : 384).

L'auteur poursuit avec l'étude des types compositionnels des noms composés dont les deux plus productifs sont :

- N + N. Avec 4 643 composés, il représente 60% des noms composés. Le pourcentage des composés gauches (*boyfriend*, *frogman*, *homeland*) atteint 88%, soit 4 081 composés, et celui des composés droits (*garden city*, *salmon trout*, *club-foot*) atteint 12%, soit 562 combinaisons. L'accentuation des N + N anglais est constante étant donné qu'une partie des composés droits (148 sur 562) qui sont en outre minoritaires est peu stable.⁴³ En somme, près des 9/10 des composés N + N sont gauches.
- A + N / N + A. Il représente 20% des noms composés et les composés A + N sont les plus fréquents. Ces juxtapositions ressemblent aux séquences fortuites nominales (même accentuation) et les frontières entre ces deux types de regroupement sont moins précises que pour le type N + N. Sur les 1 546 composés, 1 046 (68%, *abstract*

⁴¹ Les composés sont alors appelés « composés gauches » ou « composés droits » et Boisson parle de gaucherie ou de droiterie. moins précises que pour le type N + N. Sur les 1 546 composés, 1 046 (68%, *abstract*

⁴² Les composés sont tirés du *Oxford Advanced Learner's Dictionary of Current English* (Hornby : 1974).

⁴³ Les trois dictionnaires britanniques *Longman Dictionary of Contemporary English* (1978), *Penguin English Dictionary* (1969) et *English Pronouncing Dictionary* (Jones, 1977) n'accentuent pas tous les composés sur les mêmes syllabes.

noun, comic opera, black coffee) sont droits et 500 (32%, *freelance, redskin*) sont gauches (1980 : 480). Environ les 2/3 des A + N sont des composés droits. Boisson constate que les adjectifs polysyllabiques favorisent la droiterie alors que la spécialisation du sens et la non-compositionnalité favorisent la gaucherie (1980 : 497), comme dans *hardware* (10), *software* (10). Si le déterminé a un sens général, le composé a tendance à être gauche (*yellowbird, redfish, deadwood*) et si le déterminé est un terme scientifique, le composé est généralement droit (*carbonic acid, atomic energy*) (1980 : 496).

Boisson fait également le lien entre le plan sémantique et le plan phonétique en énonçant le « principe de Forrest » qui est valable pour le mot, le groupe phonologique mais pas pour la séquence fortuite, une entité syntaxique sans intérêt phonologique :

Les termes sémantiquement plus spécifiques tendent à attirer l'accent plus facilement que les termes sémantiquement moins spécifiques [...]. Dans le cas des composés, cela signifie que l'accentuation dépend de l'ordre de modification. Dans les langues à ordre modifiant-modifié, l'accent tombe sur l'élément de gauche. Dans les langues à ordre modifié-modifiant, l'accent tombe sur l'élément de droite (1980 : 253).

En conclusion, le critère de l'accentuation est peu satisfaisant et les règles d'accentuation appartiennent au domaine de la probabilité. L'analyse de l'accentuation doit tenir compte de la graphie, de la fréquence, de la longueur, de la date d'apparition des composés, de l'hétérogénéité de la communauté linguistique et de l'irrégularité du comportement linguistique des locuteurs (Boisson, 1980). Ce critère n'est ni nécessaire, ni suffisant. Seule une tendance peut être dégagée à cause des variations individuelles. Par exemple, dans le domaine d'Internet, *hypertext* comporte un accent principal sur l'élément de gauche et un accent secondaire sur celui de droite (102) dans le MWOD. En revanche, il comporte un seul accent principal à gauche (100) en anglais britannique et en anglais américain selon le CALDO.⁴⁴

1.2. L'accentuation des composés français

L'accent tonique français est un accent de groupe de mots et non un accent de mot. Cet accent oxytonique n'est pas un accent distinctif. Il ne sert pas à différencier des lexies ou des syntagmes comme en anglais, ni à marquer les exceptions comme en espagnol. L'accent français a une valeur démarcative ; il sépare des unités syntaxico-sémantiques. De fait, les règles de séparation des lexies ne sont pas les mêmes à l'écrit et à l'oral. Lorsqu'on écrit, on sépare les lexies par des blancs et lorsqu'on parle les lexies s'enchaînent (Marchello-Nizia, 1995 : 184).

L'accent étant oxytonique, le premier élément d'un composé est le plus affecté par l'assemblage surtout si c'est un adjectif monosyllabique (Catach, 1981 : 36). Il devient une particule du second élément. L'auteur cite comme exemples :

- je vais dans la maison (02) / des voisins (01)

⁴⁴ Sources : www.m-w.com www.dictionary.cambridge.org

- je vais en maison (02) d'étudiants (01)
- je vais en maison (02) d(e) repos (01)

On remarque, à travers ces exemples, qu'un transfert d'accent s'opère de *maison* vers la syllabe finale. La lexie *maison* ne porte plus l'accent principal comme lorsqu'elle est employée seule mais un accent secondaire. Les exemples de Catach montrent que, dans les séquences fortuites (*maison des voisins*) et les composés (*maison de repos*), c'est la dernière syllabe qui est accentuée. Catach ajoute que l'intégration du premier élément s'accompagne d'autres modifications comme la perte de la nasalité et la disparition de l'e caduc.

Léon (2000) fournit l'exemple suivant : *la petite* (01), *la jolie* (01), *la jolie petite* (0001), *la jolie petite maison* (000001). Dans cette séquence fortuite l'accent se déplace sur la syllabe finale. Les composés suivent-ils la même règle ? Il ne doit pas y avoir de différence accentuelle entre une séquence fortuite, une collocation et un composé puisque l'accent français est un accent de groupe. Si l'on reprend l'exemple de Catach, le dernier élément porte bien l'accent principal mais on constate la présence d'un accent secondaire (*maison de repos*, 02/01).

Pour conclure, en français, le critère de l'accentuation ne permet pas de déterminer la nature d'un groupe de lexies. Le caractère oxytonique de l'accent français le rend inutilisable pour la différenciation entre une lexie composée, une séquence fortuite et une collocation.

1.3. L'accentuation des composés espagnols

Notre analyse de l'accentuation des composés espagnols sera basée sur la thèse de De Bustos Gisbert (1986) dans laquelle l'accentuation de chaque type morphologique est détaillée (N + de + N, A + N ou N + A, N + N, V + complément, V + V, N + i + N, A + A, Adv + A). Ne seront considérés que les types très fréquents dans la terminologie d'Internet.

Les composés N + de + N se caractérisent par une pluralité accentuelle due à la présence de la préposition *de* entre les deux lexies : *caña* (10) *de azúcar* (010), *lengua* (10) *de gato* (10). Ces composés comportent deux accents principaux. Dans certains cas (moins de 3% du corpus de l'auteur), ils s'écrivent en un bloc (*estrella de mar* (010, 01) devient *estrellamar*(0001)) et/ou la préposition *de* est supprimée (*azul* (01) *de mar* (1) devient *azul már* (011)). Comme le montrent ces deux exemples, l'accentuation des composés « amalgamas » n'est pas homogène : ils comportent un ou deux accents. L'unité accentuelle semble plus fréquente pour les composés dont les deux éléments sont dissyllabiques (*telaraña*, 0010) que pour ceux dont le deuxième élément est trisyllabique (*palo áloe*, 10100).

Les composés A + N ou N + A oscillent aussi entre unité et pluralité accentuelle : 68,068% ont deux accents, 25,33% ont un accent, et un doute subsiste pour 6,602%. Comme l'illustre le tableau 2, une raison sémantique liée à l'endocentrisme ou à l'exocentrisme des composés explique cette fluctuation.

Tableau 2 : Accentuation des A + N espagnols⁴⁵

Type	Accents	A	P	Total	%
Endocentrique (615)	1 accent	24	24	48	7,8
	2 accents	17	516	533	86,5
	Incertitudes	9	25	34	5,5
Exocentrique (218)	1 accent	84	79	163	75
	2 accents	4	30	34	15,5
	Incertitudes	6	15	21	9,5

Ce tableau indique que les composés endocentriques ont majoritairement deux accents et les exocentriques un seul. Plusieurs facteurs expliquent les exceptions : a) la présence du – i – de liaison ou de suffixes (*caldibaldo*, *cuentacorrentista*), b) l'accentuation de l'adjectif peut s'affaiblir s'il est antéposé indépendamment de la nature sémantique, c) l'historique du composé, d) la possible lexicalisation ou la démotivation du composé s'accompagnent parfois de la perte d'un accent (*Nochebuena*, *Nochevieja*), e) l'unité accentuelle semble favorisée par l'écriture en un bloc.

Tableau 3 : Accentuation des N + N espagnols

Accent	Nombre	%
Un accent	175	73,83
Plus d'un accent	31	13,08
Incertitudes	31	13,08

Pour le type N + N, la tendance est l'unité accentuelle. Cependant, De Bustos Gisbert rappelle que son corpus est basé sur des dictionnaires, des glossaires et que, par conséquent, les composés sont lexicalisés, figés dans la langue. Les travaux d'autres chercheurs font ressortir un nombre supérieur de composés pluri-accentués.

Selon De Bustos Gisbert (1986 : 184), cinq facteurs expliquent la pluralité ou l'unité accentuelle des N + N :

- L'agglutination. L'agglutination partielle des deux composants se manifeste par la perte d'une syllabe (*cartapel* = *carta* + *papel*).
- La subordination des éléments du composé à une autre unité (*cuentacorrentista*). Le déterminé est un suffixe dérivatif (-ista) et le composé en est le déterminant (*cuenta corriente*).
- La relation syntaxico-sémantique entre les lexies formant le composé. Les composés sont coordinatifs (*ajíaceite*), subordinatifs (*bocacalle*, suppression de la préposition *de*) ou attributifs (*buque escuela*).
- l'endocentrisme ou l'exocentrisme du composé,
- le nombre de syllabes du composé et la position de l'accent du deuxième élément.

⁴⁵ A indique que l'adjectif est antéposé et P qu'il est postposé.

Compte tenu des divergences entre les ouvrages lexicographiques et entre les locuteurs, De Bustos Gisbert, comme Boisson et Bauer, avertit le lecteur de la valeur relative de ses résultats. De plus, étant donné la pluralité accentuelle observée pour les types N + de + N, A + N, N + A et N + N, l'accent n'est pas un critère fiable en espagnol, pas plus qu'en anglais et en français. Par conséquent, nous jugeons inutile de déterminer la nature de groupements de lexies au moyen de tests.

2. ... graphie

Utiliser un critère graphique pour identifier les composés serait inefficace puisque :

- Nombreux sont les composés dont les éléments sont séparés par un blanc. Aucune distinction entre composés et séquences fortuites ne peut donc être opérée. Dans le lexique d'Internet, des tendances se dégagent selon les langues et/ou les formants des composés. Le français et l'espagnol privilégient l'orthographe en plusieurs blocs alors que l'anglais utilise les trois orthographies (accolée, séparée, reliée). Dans notre corpus, peu de termes français et espagnols sont reliés par un trait d'union. Les termes concernés sont des surcomposés (FR *courrier non-sollicité*, *architecture client-serveur*, *pare-feu*, ES *cliente-servidor arquitectura*, *programa anti pop-up*), des composés pseudo-confixés (FR *cyber-délinquence*, *e-tourisme*, ES *e-compra*). Rares sont aussi les composés français et espagnols dont les éléments sont accolés. Seuls les amalgames et les composés contenant un pseudo-confixe ont leurs éléments accolés (seule exception *cortafuegos* en espagnol).
- En anglais, en français et en espagnol, deux ou trois graphies existent parfois pour un même composé (*popup*, *pop up*, *pop-up* ; *email*, *e-mail* ; *dotcom*, *dot com*, *dot.com*). Selon Bauer, l'orthographe des N + N anglais de la langue générale est irrégulière (1998 : 69) car il existe un processus diachronique de changement orthographique basé sur le degré de « item-familiarity » (c'est-à-dire de connaissance des unités lexicales). Mais la variation orthographique n'est pas seulement liée à ce degré de 'familiarité' avec l'unité lexicale. Par exemple, Bauer relève les trois graphies EN *daisy wheel*, *daisy-wheel* et *daisywheel* dans le *Oxford Dictionary of New Words* (Tulloch, 1991) et il n'y a aucune preuve d'une progression diachronique ; EN *airside* s'écrit en un bloc, sans trace d'une écriture séparée ou reliée et EN *college degree*, qui est une unité lexicalisée, est constituée de deux mots graphiques. Il existe tout de même des tendances : les éléments des lexies longues sont séparés par des blancs alors que ceux des lexies courtes sont plus souvent accolés. Selon Boisson, l'utilisation du trait d'union varie selon les époques et les cultures. D'après ses observations du *Webster's Third*, le trait d'union semble disparaître en anglais américain (1980 : 271) et l'écriture en bloc est moins fréquente en anglais britannique qu'en anglais américain. Pour savoir si le trait d'union est fréquemment utilisé dans le lexique d'Internet, un test a été réalisé à partir de Google.com. Il s'avère que l'hésitation entre les deux orthographies ne semble toujours pas levée en 2003.⁴⁶
- La graphie influence la place de l'accent mais elle ne permet pas de savoir si une unité est un composé ou une séquence libre (Boisson, 1980 : 271). En effet, les 3/4

des composés N + N gauches ne sont pas séparés par un blanc, contrairement aux 3/4 des composés droits. La présence d'un blanc montre que le statut d'une combinaison entre séquence fortuite et composé n'est pas parfaitement défini. D'après Boisson, un composé perçu comme tel est généralement écrit sans blanc (1980 : 400). En ce qui concerne les composés A + N, 81% des composés écrits en un bloc ou avec un trait d'union sont gauches et 82% des composés écrits avec un blanc sont droits. Toutes les combinaisons de lexies autonomes séparées par des blancs ne sont pas des syntagmes de discours (EN *black hat, address book*). La graphie est un critère peu fiable du statut du composé (Adams, 1973 : 59). Les éléments peuvent former un bloc, être séparés par un trait d'union ou par un blanc et cela indépendamment du degré d'unité unissant les formants.

3. distribution

Dans une langue, la distribution lexicale est la position qu'occupent les lexies et les morphèmes les uns par rapport aux autres. Pour savoir si une séquence est un composé, une collocation ou une séquence fortuite des tests de détermination, de modification, de changement de la place de l'adjectif, de substitution, de séparabilité, de commutation ont été proposés par différents auteurs.

3.1. Critère de détermination

Si un adjectif détermine un composé, il devrait avoir une influence sur la combinaison entière et pas uniquement sur un des ses éléments :

- EN *an [expensive [antivirus provider]]*,
- FR *un [[fournisseur d'antivirus] cher]*,
- ES *un [[proveedor de antivirus] caro]*.

Si une séquence fortuite ou une collocation subissent ce test, l'adjectif devrait déterminer l'un ou l'autre des éléments, et non l'ensemble :

- Séquences fortuites
 - EN *a [black cat], a nice black cat,*
 - FR *un [stylo neuf], un [[beau stylo] neuf],*
 - ES *un [grupo de personas], un [grupo de [personas amables]], un [[pequeño*

⁴⁶ Google (www.google.com) liste 206 000 000 occurrences pour *e-mail*, 301 000 000 pour *email*, 176 000 pour *cyber-café*, 222 000 pour *cybercafé*, 10 500 000 pour *e-business*, 3 880 000 pour *ebusiness*, 4 370 000 pour *dot(-)com*, 532 000 pour *dotcom*, 120 000 pour *dial(-)up modem*, 28 600 pour *dialup modem*, 4 950 000 pour *pop(-)up*, 3 150 000 pour *popup*. Les 9/10 des occurrences de *peer(-)to(-)peer* comportent un trait d'union et la majorité des occurrences de *file(-)sharing* n'en ont pas. Ce test se révèle assez aléatoire puisque, dans certains cas, le moteur de recherche ne fait pas la différence entre la présence et l'absence du trait d'union.

grupo] de personas].

- Collocations

- EN a [crushing victory], ?a [decisive [crushing victory],
- FR un [démenti formel], ?un [excellent [démenti formel]],
- ES un [gran interés], ?un [asombroso [gran interés]].

Dans le cas des collocations, il est difficile de trancher. Le problème de l'évaluation de l'acceptabilité en matière de test d'unités lexicales a été soulevé par Cruse (1986). FR *excellent* détermine-t-il *démenti* ou *démenti formel* ? FR *Cher* détermine *fournisseur d'antivirus* mais il n'est pas exclu qu'il détermine *fournisseur*. Dans nos exemples, les composés et les collocations ont un comportement assez similaire. La composition n'a pas de détermination interne libre à l'inverse des séquences fortuites et l'affinité entre la base et le collocatif d'une collocation fait que la détermination ne semble pas totale. Certains assemblages de lexies sont plus figés que d'autres car il existe des degrés de figement. Il y a un continuum entre les séquences fortuites et les séquences totalement contraintes (Gross, 1996 : 17).

Tournier (1991a : 65) mentionne des cas ambigus. Dans EN *traditional jazz-band*, *traditional* détermine *jazz band* mais on pourrait penser qu'il qualifie *jazz*. Pour Bauer (1998 : 72), le critère de modification adjectivale est inapproprié pour distinguer un composé d'un syntagme. Comme vont l'illustrer quelques exemples, cet exercice d'isolation syntaxique mesurant le degré de disponibilité syntaxique des éléments d'un assemblage ne permet pas de différencier ce qui est lexical de ce qui est syntaxique. Il est normalement impossible de qualifier uniquement le premier élément d'un composé N + N avec un adjectif. Par exemple, la combinaison EN [*broad [river-bed]*] est interprété comme un 'lit large pour une rivière / un fleuve' plutôt que comme un 'lit pour une rivière ou un fleuve large'. Si l'adjectif utilisé concerne *river* uniquement, comme dans **fast-flowing river-bed*, la construction semble inacceptable. Les composés EN **unsteady walking stick* et **slow walking stick* semblent inacceptables alors que *public lending right* l'est. Dans ce dernier exemple, l'interprétation dans laquelle l'adjectif modifie le premier élément du composé est tout à fait envisageable. À l'inverse, la détermination du premier élément d'un syntagme par un adjectif est normalement possible. Mais il existe des cas dans lesquels elle est inenvisageable, notamment dans les constructions matière + objet. Par exemple, *Chinese jade figure* et *Swedish steel blade* sont interprétés : [[*Chinese jade figure*]] et [[*Swedish steel*] *blade*]. Mais les interprétations [*Chinese [jade figure]*] et [*Swedish [steel blade]*] sont plus probables. En résumé, le critère de détermination ne fonctionne pas puisque le premier élément d'un composé peut être déterminé par un adjectif alors que cela devrait être impossible. Les assemblages non figés, comme les séquences fortuites, devraient être modifiables alors que les assemblages figés, comme les composés, ne devraient pas l'être.

3.2. Critère de modification de l'adjectif

Cet autre test de manipulation syntaxique, utilisé par Adams (1973) puis par Tournier (1991a), consiste à mettre au comparatif ou superlatif l'adjectif d'un composé A + N ou N + A. Ce critère n'est valable que si l'adjectif est gradable, ce qui est rarement le cas dans nos candidats composés.

- EN *an intelligent agent does its work independently, based on the user's preferences / *a more intelligent agent does its work independently, based on the user's preferences,*
- FR *un code malveillant a détruit une partie du système informatique de l'Elysée / *un code plus malveillant a détruit une partie du système informatique de l'Elysée,*
- ES *el sitio de la Casa Blanca a sufrido un ataque masivo / ?el sitio de la Casa Blanca a sufrido un ataque más masivo.*

Avec leur adjectif au comparatif, les composés forment des phrases étranges. Comme le montrent quelques exemples, ce n'est pas le cas des séquences fortuites et des collocations :

- Séquences fortuites
 - EN It is a beautiful statue It is a more beautiful statue
 - FR J'ai mis un pull chaud J'ai mis un pull plus chaud
 - ES Ésta es una película apasionante Ésta es una película más apasionante.
- Collocations
 - EN He gave a conclusive proof He gave a more conclusive proof
 - FR Cet homme est d'une laideur repoussante Cet homme est d'une laideur plus repoussante
 - ES Tiene un gran interés por el arte Tiene un interés más grande por el arte

Ce deuxième critère distributionnel semble permettre d'isoler les composés parmi les unités testées (séquences fortuites et collocations N + A ou A + N). Les formants d'un composé ne peuvent pas être déterminés indépendamment alors qu'un adjectif détermine parfois uniquement le déterminant. Mais il concerne très peu de candidats composés de notre corpus, uniquement les A^{gradables} + N et les N + A^{gradables}.

3.3. Critère de modification de la place de l'adjectif

Tournier (1991a : 65) a recours à une troisième manipulation. L'adjectif d'un composé A + N ou N + A peut rarement être en position d'attribut :

- EN *a massive attack, the attack is massive,*
- FR *le surf anonyme, le surf est anonyme / un opérateur booléen, l'opérateur est booléen,*

- ES *un servidor central, el servidor es central / un sitio comercial, el sitio es comercial.*

Ce test semble être valable uniquement dans le cas des composés exocentriques du domaine d'Internet (**the list is black, *the hat is black*). Il est inefficace lorsque l'adjectif apporte une qualification au déterminé sans idiomatique particulière, ni trope (EN *intelligent agent, encrypted virus, external modem*, FR *agent intelligent, virus crypté, modem externe* et ES *agente inteligente, virus cifrado, módem externo*). Le critère de modification de la place de l'adjectif teste l'idiomaticité et non le caractère lexical ou syntaxique d'une unité. Dans ces cas, seule la notion délicate de signifié stable permet de différencier un composé d'une collocation mais, actuellement, il semble impossible de tracer une limite (Arnaud, 2004 : 340). Par conséquent, ce critère applicable uniquement aux composés A + N et N + A n'est pas satisfaisant.

3.4. Critère de séparabilité des constituants

On ne devrait pas parvenir à intercaler librement un adjectif dans un composé :

- EN *an encrypted virus, a powerful [encrypted virus]* et **an encrypted powerful virus / an external modem, a damaged [external modem]* et **an external damaged modem / a discussion group, *a discussion big group, a cybercafé / a well-known cybercafé / *a cyber well-known café,*
- FR *un virus crypté, un puissant [virus crypté]* et **un virus puissant crypté / un modem externe, un [modem externe] défectueux* et **un modem défectueux externe / un groupe de discussion, *un groupe de bonne discussion,*
- ES *un virus cifrado, un poderoso [virus cifrado]* et **un virus poderoso cifrado / un módem externo, un [módem externo] defectuoso* et **un módem defectuoso externo / un grupo de discusión, *un grupo genial de discusión.*

Dans ces exemples, le déterminé et le déterminant des composés anglais, français et espagnols A + N, N + A, PS + N et N + N ne peuvent pas être séparés par un adjectif. Il faut préciser qu'en anglais, le placement des adjectifs n'est pas libre. Il répond à un ordre défini : les adjectifs exprimant une opinion puis les adjectifs descriptifs (a) couleur, b) origine/lieu, c) matière, d) fonction).⁴⁷ Comme nous n'avons pas prêté attention au choix des adjectifs intercalés, cela explique que **encrypted powerful virus*, entre autres, est impossible. Les éléments des séquences fortuites et des collocations peuvent-ils être séparés ?

- Séquences fortuites
 - EN *a nice boy, ?a nice obliging boy*
 - FR *un stylo neuf, *un stylo cher neuf,*
 - ES *cielo nublado, *un cielo negro nublado,*

⁴⁷ Source : Swan Michael, Walter Catherine, 2004 (1997), *How English Works, A Grammar Practice Book*, Oxford, OUP, p.73.

- Collocations
 - EN *a dazzling luxury*, **a dazzling showy luxury*,
 - FR *une agitation fébrile*, **une agitation exceptionnelle fébrile*,
 - ES *un montón de gente*, **un montón increíble de gente*.

Si les éléments des séquences libres et des collocations ci-dessus ne peuvent être interrompus, c'est parce que l'ordre des modificateurs est important. Les modificateurs proches du déterminé sont plus pertinents que ceux qui sont éloignés et les modificateurs les plus éloignés ont un impact sur ceux qui sont proches (Bauer, 1998 : 82). Sachant cela, lors d'un test, le choix et l'ordre des modificateurs peuvent être manipulés en fonction des résultats espérés (FR *un stylo neuf*, *un stylo bleu neuf* mais **un stylo cher neuf*). Par ailleurs, l'impossibilité d'interrompre les composés EN *black hat* et *firewall* est due à leur exocentricité. L'opacité sémantique (ou non-compositionnalité) et les restrictions syntaxiques vont de pair (Gross, 1996 : 12).

Enfin, si l'adjectif d'une séquence fortuite N + A est un adjectif relationnel, les éléments sont alors inséparables (Arnaud, 2003 : 26) :

- FR *un problème pulmonaire*, **un problème grave pulmonaire* mais *un grave problème pulmonaire*,
- *un événement sismique*, **un événement mémorable sismique* mais *un événement sismique mémorable*.

Ce critère, applicable à tous les types compositionnels, n'est donc pas non plus un bon moyen de distinguer les composés par rapport aux séquences fortuites et aux collocations puisque, sous certaines conditions, les constituants de séquences fortuites et de collocations sont inséparables, tout comme peuvent l'être le déterminant et le déterminé de lexies composées.

3.5. Critère de commutation

Ce critère est utilisé par Lehmann et Martin-Berthet (1997 : 172) et Paillard (2000 : 46) qui prennent comme exemple la lexie composée FR *pomme de terre* substituable dans une phrase par la lexie simple *poireau* ou *carotte*. Dans le domaine d'Internet :

- EN *I received a bulk email from Bill Gates* peut être remplacé par *I received a spam from Bill Gates*,
- FR *j'ai changé de fournisseur d'accès Internet* par *j'ai changé d'opérateur*,
- ES *tengo otra cuenta de correo* par *tengo otra dirección*.

Mais les séquences fortuites et les collocations peuvent, elles aussi, être remplacées par des lexies simples :

- Séquences fortuites

- EN my right leg hurts, my foot hurts,
 - FR je fais chauffer une casserole d'eau, je fais chauffer une marmite,
 - je construis un château de sable, je construis une forteresse,
 - ES me gustán las flores rojas, me gustán las rosas.
- Collocations
- FR il a répondu par un refus catégorique, il a répondu négativement, il a payé le prix fort, il a payé cher.
 - ES tiene una envidia loca de chocolate, tiene ganas de chocolate.

Nous remarquons que, dans ces exemples, la perte d'information est plus importante lorsqu'une lexie simple remplace une séquence fortuite plutôt qu'un composé ou une collocation. Qu'y a-t-il dans la *marmite* ? En quoi est la *forteresse* ? Ceci est logique : les composés sont des produits finis (des unités de sens), les collocations sont des produits semi-finis (des combinaisons reconnues par les locuteurs) alors que les séquences fortuites sont des combinaisons arbitraires, non idiomatiques. Ce critère semble assez efficace pour différencier les composés des séquences libres mais pas des collocations, probablement parce que ces dernières sont consacrées par l'usage. Il est assez rare que le concept désigné par un composé soit aussi désigné par une lexie simple (FR *avion suicide* et *kamikaze*, ES *paving stone* et *sett*) ou ait pour équivalent une lexie simple dans d'autres langues⁴⁸ (FR *soutien-gorge* et EN *bra* EN *paving stone* et FR *pavé*)(Arnaud, 2004 : 334).

3.6. Critère de non-coordination du déterminant avec un adjectif

Le critère de non-coordination du déterminant avec un adjectif est repris par Boisson (1980 : 256).

- Composés
 - EN an encrypted e-mail *a encrypted and secure email
 - FR un accès permanent *un accès permanent et peu rapide
 - ES la navegación anónima *la navegación anónima y ocasional
- Séquences fortuites
 - EN a comfortable house a comfortable and beautiful house
 - FR une télécommande grise une télécommande grise et carrée
 - ES un bote para lápices un bote para lápices y clipes

⁴⁸ Arnaud a procédé à des comparaisons inter-langues avec l'anglais, l'allemand, l'arménien, le basque, le bininj gun-wok, le cri, le français, le gallois, le hongrois, le khmer, le tamazight et le turc (2004 : 348-349).

- Collocations
 - EN a dazzling luxury a dazzling and showy luxury
 - FR une preuve irréfutable une preuve irréfutable et inattendue
 - ES una fealdad repulsiva una fealdad repulsiva y repugnante

D'après les exemples ci-dessus, le critère de non-coordination du déterminant est un bon moyen de distinction des lexies composées parmi les séquences fortuites et les collocations. Les composés sont le seul type de combinaisons dont le déterminant ne peut pas être coordonné. Mais attention, d'après Bauer (1998 : 74-76), deux contraintes déterminent la coordination. Premièrement, les noms coordonnés doivent appartenir au même domaine (EN *cat and dog shows* mais **antique and dog shows*). EN *buttercup* ne peut pas être coordonné (**bread and buttercup*, **buttercup and saucer*) car aucune autre fleur ayant *-cup* dans sa dénomination n'existe. Le sens de *buttercup* et le fait qu'aucune fleur n'ait un nom similaire empêchent la coordination. Deuxièmement, la coordination n'est possible que si la relation sémantique entre les éléments coordonnés est identique (EN *steel bars and weights* mais **steel bar and works* et EN *wind-and water-mills* mais **wind- and flour-mills*). Ces contraintes, valables pour les composés, les séquences libres et les collocations, font qu'il est difficile de trouver des unités pouvant être coordonnées avec un des éléments constitutifs. Par ailleurs, plus la construction est lexicalisée, plus il est difficile de trouver un adjectif pour coordonner le déterminant.

4. Sémantique

À l'aide de paires d'exemples,⁴⁹ Kocourek (1991 : 135) montre que les composés ont une spécificité perceptible mais difficilement saisissable. Les FR *acides gras* sont des monoacides organiques spéciaux et les *aliments gras* sont des aliments contenant des graisses. La première lexie est un composé et la seconde une séquence fortuite. Une lexie composée désigne une « notion différente de la somme des notions correspondant à ses éléments » (Tournier, 1985 : 64). Arnaud (2003 : 3) parle d'idiomaticité lorsque le sens du composé n'est pas réductible à celui de ses formants.⁵⁰ L'affirmation de Tournier ne s'applique pas à tous les composés et, notamment, pas aux composés endocentriques : FR un *courrier électronique* est un courrier sur un support électronique. Cependant, selon Lehmann et Martin-Berthet (1997 : 173), si chacun des formants du composé garde son sens habituel, des éléments spécifiques s'insèrent, ce qui particularise la nouvelle unité lexicale et l'éloigne du sens de ses formants. Un *courrier électronique* est bien un courrier qui fonctionne selon les lois de l'électronique mais cette description est insuffisante pour décrire le concept dans son ensemble car elle est basée sur les éléments de nomination. Or, tous les traits conceptuels ne figurent pas dans la

⁴⁹ *acides gras* et *aliments gras*, *centre de gravité* et *centre d'une région*, *terres rares* et *sentiments rares*, etc.

⁵⁰ D'autres auteurs dont Béjoint (1994 : 213) parlent de non-compositionnalité (et de compositionnalité dans le cas contraire).

dénomination (Thoiron et al., 1996). Il s'agit essentiellement d'une poste virtuelle sans papier ni enveloppes qui fonctionne via un ordinateur. Une *bannière publicitaire* n'est pas un « Mod. Étendard que l'on porte aux processions et qui sert à distinguer une paroisse ou une confrérie » (PR 1996). Il s'agit d'une zone d'un site Internet contenant une publicité. La lexie *bannière* acquiert un nouveau sens lors de association avec *publicitaire* et son utilisation dans la terminologie d'Internet.

En anglais, français et espagnol, on l'a vu, deux types de noms composés existent : les endocentriques et les exocentriques. Les composés exocentriques ont la particularité d'être figés sémantiquement et syntaxiquement. Gross les qualifie de « suites 'gelées' qui ne diffèrent d'un mot simple que par leur polylexicalité et les marques morphologiques spécifiques aux composés » (1996 : 36). Le référent correspondant au composé ne peut être associé à aucun des éléments constitutifs du composé. Ces composés sont formés par des tropes et surtout par la métaphore (EN *backdoor*, *black hat*, FR *cheval de Troie*, ES *cortafuegos*, etc.). Les composés endocentriques, plus proches des séquences fortuites et des collocations, possèdent un substantif-tête (EN *comparison site*, *cookie manager*, *electronic message*). Le substantif-tête peut être employé seul :

- EN this application is a cookie manager this application is a manager *a cookie manager is a manager*

Il n'en va pas de même pour les exocentriques :

- EN the process was slowed down by a bottleneck * a neck slowed the process down
**a bottleneck is a neck*

Bauer (1983 : 30) et la RAE (1999) distinguent plusieurs types de composés endocentriques :

- Les coordinatifs (ou copulatifs). Les deux éléments du composé sont des têtes. Le concept correspond à chacun des constituants : dans ES *café teatro*, EN *maid-servant* le concept est l'addition des deux éléments ; dans ES *sureste* le concept est un point intermédiaire des concepts des constituants *sur* et *este*.
- Les subordinatifs. La tête peut être le premier élément, ES *bocamina*, ou le deuxième, ES *cineclub*. Certains comme ES *telaraña* sont parfaits, c'est-à-dire que leurs éléments sont les éléments de composés syntagmatiques qui ont fusionné (ES *boca de mina*, *tela de araña*). D'autres, comme ES *camión cisterna*, sont imparfaits ; il y a une apposition (*un camión con una cisterna*).

Les composés exocentriques sont aisément différenciables des collocations puisque ces dernières sont toujours transparentes : « le sens global est déductible des unités de composition » (Clas, 1994 : 577). Les collocations partagent cette caractéristique avec les séquences fortuites. En revanche, les composés endocentriques sont plus proches des séquences fortuites et des collocations. Dans ces trois types d'assemblages, un des constitutifs caractérise l'autre.

Pour conclure, les lexies composées endocentriques ou exocentriques résultent d'un assemblage formel et d'une fusion conceptuelle. Comme on l'a vu, un FR *courrier*

électronique est plus que ce que les éléments de dénomination laissent voir. La lexicalisation suppose la constitution d'un signifié stable et l'accès en compréhension ou production est direct, sans calcul du sens du composé (Arnaud, 2004 : 335).

5. ... Fréquence

La définition de la collocation de Tournier présente un continuum de la combinaison non attestée à la combinaison nécessaire en passant par la combinaison fortuite et la combinaison fréquente.

Dans les revues d'Internet anglaises, *fat* ne précède jamais *mail*, *free* détermine quelquefois *mail*, *snail* précède fréquemment *mail*, et *back* est obligatoirement accolé à *door* :

[...] il est clair que dès l'instant où on parle de fréquence, on ne peut plus prétendre à une opposition tranchée entre le syntagme et le mot composé : c'est encore une fois une question de degré (Boisson, 1980 : 259).

Étant donné que, dans certains cas, il n'existe aucune différence nette entre un composé et une séquence fortuite fréquente, Tournier parle de « degré d'intégration dans le lexique » (1991a : 66). Un composé a été une combinaison fortuite avant d'être intégré dans le lexique. Sa fréquence d'emploi lui a ensuite permis d'être répertorié. Une séquence avec une fréquence telle que celle de *snail mail*,⁵¹ généralement non répertoriée, n'est vraisemblablement pas une séquence fortuite mais un composé en voie d'intégration. La fréquence d'emploi des combinaisons détermine leur présence dans les dictionnaires mais d'autres facteurs interviennent aussi : l'ancienneté, la longueur, l'étrangeté des combinaisons, le domaine couvert par le dictionnaire. Le groupement *snail mail* conjugue deux aspects : nouveauté et étrangeté. Quel rapport y a-t-il entre un courrier et un escargot ? La cooccurrence des deux éléments est rendue possible par la métaphore et l'allitération.

Pour approfondir le critère de fréquence, Boisson confronte des répertoires de mots. Après avoir sélectionné cinquante expressions anglaises ayant une entrée dans le *Dictionary of Clichés* (Partridge, 1969), il vérifie si elles figurent comme composés ou comme entrées séparées dans dix-huit autres répertoires. Il constate que les assemblages se répartissent dans deux catégories :

- les associations lexicalisées (*carte blanche*, *good for nothing*). Elles possèdent une entrée spéciale dans la plupart des dictionnaires, même peu riches.
- les combinaisons qui ont un statut confus. Elles forment un continuum de la collocation ou cliché (*diabolical skill*) à l'expression toute faite (*castle in Spain*, *burden of proof*) pour arriver enfin à des assemblages en passe de figurer comme entrée lexicales dans les dictionnaires, et donc susceptibles de se lexicaliser (*dead letter*, *acid test*) (1980 : 265).

⁵¹ Dans le jargon des internautes *snail mail* désigne le courrier postal. Il est assimilé à un escargot car son acheminement est beaucoup plus long que celui de son équivalent électronique.

Boisson établit également une gradation dans la fréquence d'apparition de ces combinaisons dans les ouvrages lexicographiques (ordre décroissant) : mot simple (*man*), mot historiquement composé devenu complètement opaque (*without*), mot historiquement composé devenu presque opaque (*housewife*), mot composé transparent avec altérations sonores minimes (-*man*), mot composé habituel sans altération sonore, assemblage courant avec hésitation (*publicity man*), assemblage-phrasé souvent absent des dictionnaires malgré sa forte fréquence (*see you later*).

Le critère de fréquence a des limites. Comme toutes les unités lexicales, un composé peut :

- ne plus être employé peu de temps après sa création (FR *babillard électronique*).
- voir sa fréquence d'emploi variée d'une communauté linguistique à une autre. EN *electronic mail* a une fréquence d'emploi élevée dans les revues d'informatique/Internet ou chez les jeunes et une fréquence faible dans d'autres types de revues ou chez des personnes plus âgées.

6. Conclusion

Les tests de compositionnalité ne lèvent pas toutes les incertitudes. Les unités lexicales testées dans ce chapitre révèlent l'insuffisance des critères pour déterminer la lexicalité d'un assemblage. De plus, à la suite des observations de Bauer (1998), on s'attend à ce que le statut de certaines combinaisons de lexies ne soit pas clair. Les critères considérés ne permettent pas de confirmer ou d'infirmer une division lexique-syntaxe. Une division composé-séquence libre basée sur un seul critère serait hasardeuse et non la preuve d'une frontière. Après examen des corrélations entre critères, Bauer (1998) doute de l'existence de deux constructions distinctes. Aucun exemple ne réunit toutes les particularités des composés ou des séquences libres.

Nous partageons l'avis d'Adams et de Boisson qui voient le statut de composé comme une question de degré (1973 : 58). Tournier résume bien ce qui résulte de cet assemblage de tests :

Aucun des critères proposés pour décider si une séquence est un mot composé n'est à lui seul déterminant dans tous les cas. Toutefois, l'application conjointe de ces cinq critères [accentuation, graphie, distribution, sémantique, fréquence] peut, sinon aboutir à une certitude, du moins permettre de réunir un faisceau de présomptions (1991a : 67).

Pour réaliser notre projet, nous avons besoin d'une définition opératoire pour le recensement des noms composés du domaine d'Internet. La définition utilisée sera peu restrictive⁵² puisqu'une définition étroite nuirait à la mise en place d'un système d'équivalences types d'aide à la traduction en écartant les catégories mineures. Nous pensons qu'appliquer un triple filtre⁵³ aux candidats composés dont les critères sont lâches peut réunir l'indispensable « faisceau de présomptions ».

⁵² Arnaud (2004 : 333) se demande si les auteurs qui adoptent des définitions restrictives, dont Di Sciullo et Williams 1987 pour qui le français et l'espagnol manquent de composés, ne desservent pas leur théorie.

Seront extraites du corpus et appelées *noms composés* les lexies qui, d'une part, correspondent à la définition suivante du nom composé et à la majorité des propriétés énoncées par Tournier (Part. I, Ch. 1, Sec. 1.) : un nom composé est une lexie construite désignant un concept. Les éléments qui entrent dans la formation d'un nom composé sont le plus souvent autonomes, c'est-à-dire des lexies simples, des conjonctions, des prépositions, des lexies composées, des sigles, des amalgames mais aussi des morphèmes grammaticaux. Les éléments dépourvus d'autonomie sont des troncations et des pseudo-confixes (voir Part.I, Ch.1, Sec. 2. pour les détails et les exemples). D'autre part, pour éliminer autant que possible les séquences fortuites et les collocations, les termes sélectionnés devront en outre :

- figurer dans des ouvrages de référence généraux ou spécialisés. Nous avons conscience que ces ouvrages ne se soucient pas de savoir si un assemblage est un composé ou pas, mais l'inclusion des candidats composés est un signe de lexicalité. S'ils sont répertoriés, ils ne sont normalement pas des séquences fortuites. Nous consulterons des ouvrages électroniques⁵⁴ car les mises à jour y sont plus fréquentes que celles des versions papier.
- être indexés, sous certaines conditions, par des moteurs de recherche. Google.com pour l'anglais, Google.fr pour le français et Lycos.es ou Yahoo.es puis Google.es, depuis sa création, pour l'espagnol sont utilisés comme concordanciers. Nous estimons que cinq attestations relevées dans cinq contextes Internet différents et le fait que chaque contexte dans lequel le candidat apparaît est propre au domaine d'Internet seront suffisants pour réunir un faisceau de présomptions. En effet, ces deux exigences tendent à assurer la lexicalité des candidats composés⁵⁵ et la stabilité du signe. Nous vérifierons que ces termes sont utilisés sur des sites sérieux (sites de revues Internet, portail 01Net, etc.) et dans les glossaires constitués par des instances fiables (ATI pour l'espagnol, GDT pour l'anglais et le français, sites des fabricants d'antivirus).

Si un doute subsiste sur une unité, elle sera recensée. En effet :

- l'efficacité des critères est certes controversée mais leur utilisation souligne le fait que les différences entre les composés et certaines séquences libres et collocations sont extrêmement minces. La division composé-séquence libre est très complexe car leur formation est probablement identique : un composé a été une séquence libre avant d'être lexicalisé.
- les difficultés de traduction seront les mêmes étant donné les ressemblances entre certains noms composés, séquences fortuites ou collocations. Par conséquent, éliminer un assemblage à cause de son statut incertain desservirait le traducteur.

⁵³ Une illustration du triple filtre figure en Annexe (Annexe 4).

⁵⁴ www.m-w.com , www.dictionary.cambridge.org pour l'anglais, www.granddictionnaire.com pour l'anglais et le français, GR 2001 pour le français, www.diccionario.es pour l'espagnol.

⁵⁵ Devant le manque de fiabilité des tests de lexicalité, Arnaud (2003) a retenu cette méthode pour déterminer en toute objectivité les « suites NN timbre-poste » qui sont des unités lexicales.

Chapitre 3. Quelques typologies des noms composés

À chaque auteur sa typologie. La profusion des typologies s'explique par le fait que les auteurs adaptent leur catégorisation à leur but et leur besoin. Les classements sont souvent basés sur la morphologie des composés et plus rarement sur leur sémantique. Ce troisième chapitre examinera des typologies élaborées par Tournier, Gross, Almela Perez, Depierre et Béciri. Analyser des classifications créées pour les composés de la langue générale et de certaines langues de spécialité permettra d'évaluer l'utilité des catégories préexistantes pour le fonctionnement de notre futur système d'équivalences types anglais-français et anglais-espagnol.

Nous sommes consciente que ce sont les composés anglais et leur classement qui sont déterminants dans la création d'un outil de proposition d'équivalences types. Néanmoins, étant donné les ressemblances entre les composés des trois langues étudiées,⁵⁶ des typologies réalisées pour des composés français et espagnols pourront être utiles, notamment pour déterminer le degré d'affinage des typologies et des catégories à retenir ou rejeter. Dans l'hypothèse où nous réutiliserions des classes de typologies françaises ou espagnoles, les modifications à apporter pour créer un classement anglais seront infimes (ordre de détermination). Plusieurs modèles de construction seront ainsi à notre disposition pour repérer les points indispensables à la création des typologies qui seront au cœur de notre système d'équivalences types.

1. Typologies conçues pour la langue générale

1.1 La typologie de Tournier

Dans *Introduction à la lexicogénétique de l'anglais contemporain* (1985), le but de Tournier est de décrire environ 75 000 lexies de l'anglais courant contemporain (1960-1980) pour donner matière à de futurs travaux sur la lexicogénèse de l'anglais. Le troisième chapitre, consacré à la composition, est découpé selon la nature grammaticale (nominale, adjetivale, verbale) des composés. Nous ne retiendrons ici que la typologie des noms composés, illustrée de nombreux exemples et comprenant les types attestés dans un corpus de sept dictionnaires. Cette classification est reproduite sous la forme d'un tableau :

Tableau 4 : Typologie de Tournier

⁵⁶ Le premier chapitre a montré que les définitions de Tournier créées pour les composés de la langue anglaise s'appliquent parfaitement au français et à l'espagnol.

	Types	Exemples
1	N + N	<i>honours list, bell-boy</i>
2	A + N	<i>shorthand, pink elephant</i>
3	N s + N (déterminant au génitif)	<i>statesman, fool's paradise</i>
4	N propre + s + N	<i>Parkinson's disease, Adam's apple</i>
5	N propre + N	<i>Oxford accent, Molotov cocktail</i>
6	N propre + A	<i>London special, Vandyke blue</i>
7	N + V° ⁵⁷	<i>toothpick, wind-break</i>
8	N + V° + -er	<i>story-teller, shopkeeper</i>
9	A + N + -er	<i>old-timer, freebooter</i>
10	N + déverbal en -ing	<i>data processing, wife-swapping</i>
11	Pronom + N	<i>he-goat, she-ass</i>
12	A + A	<i>mental defective, new rich</i>
13	A + V°	<i>whitewash</i>
14	Letter + N	<i>O level, X chromosome</i>
15	A /Adv+déverbal -ing	<i>well-being, ill-breeding</i>
16	Adv + V°	<i>downpour, upstart</i>
17	Adv + Vn	<i>offcut, outcast</i>
18	Vn + N	<i>fallen angel, detached house</i>
19	Ving + N	<i>developing country, revolving door</i>
20	V° + N obj	<i>pickpocket, ceasefire</i>
21	V° + N non-obj	<i>dial-tone, whetstone</i>
22	V° + -er + part adv	<i>passer-by, hanger-on</i>
23	V° + part adv	<i>stand-by, sit-in</i>
24	A /Adv + V + -er	<i>onlooker, loudspeaker</i>
25	Adv-prep + N	<i>afternoon, by-election</i>
26	Déverbal -ing + N	<i>freezing-point, drinking water</i>
27	Surcomposé à trois éléments	<i>AB + C : duty-free shop A+BC : central nervous system A+B+C : still small voice</i>
28	Surcomposé à quatre éléments	<i>sudden infant death syndrome</i>
29	Surcomposés à cinq éléments	<i>sixty-four thousand dollar question</i>
30	N + A (emprunt ou calque)	<i>princess royal, court martial</i>
31	N + A (conversion A > N)	<i>penny dreadful</i>
32	N + prep + N	<i>Secretary of State, Secretary of State for War</i>
33	N + N propre	<i>peach Melba, herb Paris</i>
34	Double déterminant	<i>Postmaster General</i>
35	VC converti en NC	<i>pull-over, make-up</i>
36	AC converti en NC	<i>first-born, evergreen</i>

⁵⁷ « V° peut parfois être interprété également comme N par conversion .» (Tournier, 1985 : 114)

37	Adv comp converti en NC	<i>hereafter, whereabouts</i>
38	2 N coordonnés	<i>cap and bells, sale or return</i>
39	2 A coordonnés	<i>young and old, mild and bitter</i>
40	2 V coordonnés	<i>hide-and-seek, tie and dye</i>
41	Prep + N	<i>at-home, plus-fours</i>
42	Adv + Vn	<i>newly-weds</i>
43	Composé à redoublement	<i>helter-skelter, hurly-burly</i>
44	N + numéral ou lettre	<i>mach one, vitamin B1</i>
45	N + adv (ou part adv)	<i>line ahead, crop-over</i>
46	V + adv	<i>know-how, die-hard</i>
47	V° + V°	<i>make-believe</i>
48	(Part) adv +(part) adv	<i>in-off</i>
49	Fragment de lexie complexe	<i>rolling stone, Kilkenny cats</i>
50	Référence implicite à une lexie complexe	<i>first string, second string</i>
51	Conversion verticale	<i>touch-me-not, beggar-my-neighbour</i>
52	NC emprunté	<i>joie de vivre, chargé d'affaires</i>
53	Divers types rares	<i>no man's land, go-getter, bride to-be, merry-go-round, first cousin once removed</i>

Cette typologie est une typologie en « râteau ». Par exemple, les composés N + N sont au même niveau que les N propre + N, les N + déverbal -ing, etc. Ainsi, V° + N obj et V° + N non-obj ne font pas partie d'une classe V° + N décomposée en V° + N obj et V° + N non-obj. Il n'y a pas d'hyperonymes, ni d'hyponymes.

Tournier prend en compte des facteurs très différents les uns des autres :

- La nature grammaticale des éléments constituants des composés : nom (N), verbe (V°), adjectif (A), adverbe (Adv), pronom (Pronom), préposition (prep), participe adverbial (part adv).
- La nature exacte de ces mêmes éléments : nom propre (n°4, 5, 6, 33), nom comprenant le suffixe -er (n°9), nom dérivé d'un verbe au moyen du suffixe -er (n°8, 22, 24,), nom formé à partir d'un verbe en -ing (n°10, 15, 26).
- L'origine des éléments : emprunt ou calque (n°30, 52).
- La catégorie première du composé : verbe, adverbe ou adjectif composé converti en nom composé (n°35, 36, 37), conversion de l'un des éléments (n°21).
- Le redoublement des éléments : n°43.
- La présence d'une lettre ou d'un chiffre : n°14, 44.
- La coordination des éléments : n°38, 39, 40.
- La troncation de lexies complexes : n°49.

- L'existence d'une référence implicite : n°50.
- Le nombre d'éléments formant le composé : n°27, 28, 29.

En conclusion, certaines des catégories de Tournier sont superflues dans le cadre du système d'équivalences (n°4 à 6, 20, 21, 38 à 40, 43, 50 à 52) car nous n'avons pas relevé, dans le lexique d'Internet, de composés contenant un nom de personne (n°4 à 6), un redoublement et une conversion verticale (n°51). Nous n'avons rencontré qu'un seul coordonné (n°38-40). Hormis le type n°52 (emprunt total) rejeté uniquement pour la typologie anglaise, tous les types mentionnés ci-dessus seront rejettés pour les trois langues étudiées. Une catégorie emprunt total sera introduite dans les classements français et espagnols.⁵⁸ Les composés dont un seul formant est emprunté seront dans la même catégorie que les composés français et espagnols de nature grammaticale identique car l'élément emprunté se comporte comme un élément endogène (FR *fenêtre pop-up* et ES *banner publicitario* seront dans les classes N + N et N + A alors que FR *bookmark* sera dans une classe 'Composés exogènes').

1.2. La typologie de G. Gross

Dans *Typologie des noms composés* (1986), l'objectif de Gross est de créer un dictionnaire électronique de noms composés français « qui prend en compte à la fois une description linguistique de la composition nominale et un traitement informatique des données permettant un accès automatique à ces données » (*ibid.* : 6). L'auteur construit une typologie fondée sur la structure des composés qui est organisée en trois parties : a) une classification première, b) un aperçu de structures plus complexes et c) une analyse rigoureuse des types N + A et N + N. Le tableau ci-dessous reprend la classification première de l'auteur (*ibid.* : 70).

Tableau 5 : Typologie de G. Gross

⁵⁸ Nous abordons ici la création des typologies françaises et espagnoles. Il nous semble pertinent de parler de toutes nos futures typologies, indépendamment de leur objectif. Rappelons, cependant, que si les classements français et espagnols ont un seul objectif – dresser le profil morphosémantique de tous les composés (Partie II, Chapitres II, III et IV) pour fournir au traducteur spécialisé dans le domaine d'Internet les informations utiles à la traduction des noms composés anglais en français et en espagnol – la classification anglaise a, elle, un double objectif – permettre l'élaboration du système d'équivalences types et améliorer la connaissance des composés.

Les noms composés anglais français et espagnols du domaine d'Internet :

Classes	N°	Types	Exemples
A	1	N N	<i>homme grenouille, chou-fleur</i>
	2	N Adj	<i>cercle vicieux, cordon bleu</i>
	3	Adj N	<i>beau-père, blanc-bec</i>
	4	N part prés	<i>poisson volant, sables mouvants</i>
B	5	N de N	<i>pomme de terre, table des matières</i>
	6	N par N	<i>preuve par neuf, moulage par pression</i>
	7	N en N	<i>arc-en-ciel, entrée en fonction</i>
	8	N à N	<i>pelle à gâteau, roue à aube</i>
	9	N prép N	<i>mise sur orbite, sculpteur sur bois</i>
C	10	V N	<i>gratte-papier, attrape-nigaud</i>
	11	V prép Inf	<i>pince-sans-rire</i>
	12	V prép N	<i>touche-à-tout, tire-au-flanc</i>
	13	V impér	<i>rendez-vous</i>
	14	V con V	<i>va-et-vient</i>
D	15	V adv	<i>frappe-devant</i> ⁵⁹
	16	à N	<i>un à-coup, un à-côté</i>
	17	contre N	<i>un contre-projet, une contre-allée</i>
	18	sur N	<i>le sur-moi, du sur-place</i>
	19	sans N	<i>un sans-culottes, un sans-abri</i>
E	20	arrière N	<i>une arrière-saison, l'arrière-train</i>
	21	Prép N	<i>la plus-value, une avant-scène</i>
	22	Prép Pron	<i>un chez-soi</i>
	23	Prép Adv	<i>un en-avant, un pardessus</i>
	24	Adv part prés	<i>un bien pensant</i>
F	25	Adj Prép N	<i>un haut-de-forme</i>
	26	Numéro N	<i>un trois-mâts, un deux-pièces</i>

Comme Tournier, l'auteur établit un classement basé sur la classe grammaticale des éléments constitutifs des composés (N A, A N, N N, V N, etc.). Ce classement appelle quelques remarques :

- Cinq grands regroupements sont opérés (classes A à E) : A) les composés formés d'un nom et un autre élément, B) les composés de deux noms reliés par une particule, C) les composés comprenant un verbe, D) les composés formés d'une préposition ou d'un adverbe et d'un nom et E) les composés autres.
- Les structures sont plus nombreuses que ne le laisse penser le tableau. Gross fait état de structures qu'il estime beaucoup plus complexes :
 - a) Les composés incluant un numéral. Cette catégorie, qui est le développement du n°26, contient sept types : card + card (*trois huit*), card + prép + card (*douze en dix*), card + N + Adj (*sept péchés capitaux*), card + N + de + N (*sept merveilles du monde*), N + card + N (*restaurant trois étoiles*), N + prép + card + N (*marchande des quatre-saisons*), N + prép +

⁵⁹ Un *frappe devant* est un gros marteau à long manche, à l'usage des forgerons.

card *preuve par neuf*). Ces lexies ne correspondent pas toutes à la définition du nom composé utilisée ici.

b) Les composés incluant un verbe. Cette catégorie est l'extension de la classe C du tableau 5. Gross dénombre treize structures supplémentaires dont : V + dét + N (*trompe-la mort*), V + prép + N (*laisser sur place*), V divers (*va-et-vient, porte-à-faux, pousse-pousse, pince-sans-rire, va-nu-pieds*).

c) Les composés comprenant des noms et des adjectifs. Cette section développe les types 2 et 3 du tableau 5. Plusieurs groupes peuvent être combinés, aussi la liste qui suit n'est-elle pas exhaustive : A A N (*haut moyen-âge*), A N A (*Grand Lac salé*), N A N (*restaurant libre service*), A et A N (*petites et moyennes entreprises*), N A et A (*propriété artistique et littéraire*), N A A (*système monétaire européen*), A N A A (*Saint Empire romain germanique*), N A A A (*pion central passé isolé*), N N A (*opération ville morte*).

d) Les composés contenant des noms, des adjectifs et des prépositions. Plusieurs types peuvent être combinés entre eux (n°2, 3, 5, etc.). Les types recensés sont : A N de N (*Haute cour de justice*), N A de N (*chapeau haut-de-forme*), N A de Dét N (*principes directeurs de la conscience*), N de A N (*loi de finance rectificative*), N à A N (*film à grand public*), N à N A (*opération à cœur ouvert*), N de Dét N A (*principe du tiers exclu*), N à Dét N A (*œil au beurre noir*), N A à Dét A N (*classes préparatoires aux grandes écoles*).

Gross précise ensuite que les langues techniques et scientifiques comportent des termes plus longs et illustre ses propos avec un terme utilisé en droit : *servitude de libre passage sur les berges des cours d'eau non navigables et non flottantes*. La plupart des catégories des paragraphes a, b, c et d sont des extensions de classes du tableau 5. Elles sont, elles aussi, basées sur la nature grammaticale des constituants. Ces catégories appellent deux remarques :

- Certaines d'entre elles se recoupent : *preuve par neuf* est inclus dans le type n°6 et dans les composés avec numéral, *pince-sans-rire* apparaît dans V divers et dans le type n°11 (V + prép + inf).
- Les types répertoriés dans le paragraphe c) sont différents de ceux des autres catégories ; ils contiennent plus de deux éléments constitutifs.

Gross analyse ensuite avec précision les composés contenant un substantif et un adjectif et ceux comprenant deux substantifs. Dans sa classification des composés N + A, Gross utilise quatre critères : noms propres/noms communs, humains/non-humains, degré de figement et avec/sans verbe support. Le degré de figement du composé est pris en compte pour les noms communs humains et non-humains. Par exemple, une *main courante* n'a aucun lien avec une *main* (figement total) alors qu'un *être humain* est un *être*. La distinction concernant le verbe support s'applique uniquement aux composés non-humains. Un verbe support peut rendre compte de la relation entre les éléments pour certains composés (*arrêté préfectoral > le préfet a pris un arrêté*) mais pas pour d'autres(*vin blanc > *ce vin est blanc* ne conserve pas le sens du composé).

N + A⁶⁰

- Noms propres

Npr Hum (Soldat Inconnu, Croix-Rouge)

Npr -Hum (Mont Blanc, Etats-Unis)

- Noms communs

- Nc Hum

NA est N (*cousin germain, détective privé*)

*NA est N (*pied noir, cordon bleu*)

- Nc- Hum

NA est N

- verbe sup (*vin blanc, accent grave*)

- + verbe sup (arrêté préfectoral, aménagement rural)

*NA est N (*main courante, cerf-volant*)

La structure de la classe N + N est différente de celle que l'on vient de voir. Elle ne tient pas compte de l'existence d'un verbe support mais donne plus de détails sur le degré de figement. L'auteur précise si le composé peut être remplacé par l'un ou l'autre de ses éléments ou par les deux : une *commode-bureau* est un *bureau* et une *commode* (N1N2 est un N1 et un N2), un *bloc-notes* est un *bloc* (N1N2 est un N1 et *N1N2 est N2), un *ciné-club* est un *club* (*N1N2 est un N1 et N1N2 est un N2), un *auto-stop* n'est pas une *auto* et n'est pas un *stop* (*N1N2 est un N1 et un N2).

- N + N

- Noms propres

Npr Hum (*Roi Soleil*)

Npr –Hum (*Tour Eiffel, Cap Horn*)

- Noms communs

- Nc – Hum

N1N2 est N1

N1N2 est N2 (*commode-bureau, bloc-répertoire*)

60

L'utilisation de puces différentes reflète les divers niveaux de la taxinomie de Gross. Hum = humain, -Hum = non humain, *= incorrect, impossible.

*N1N2 est N2 (*abri-jardin, bloc-notes*)

*N1N2 est N1

N1N2 est N2 (*auto-école, ciné-club*)

*N1N2 est N2 (*auto-stop, coin-repos*)

· Nc Hum

N1N2 est N1

N1N2 est N2 (*artiste-peintre, chauffeur-livreur*)

*N1N2 est N2 (*bébé-éprouvette, commis-voyageur*)

*N1N2 est N1

N1N2 est N2 (*ciné-reporter, coq héron*)

*N1N2 est N2 (*auto-stoppeur, apprenti-sorcier*)

Dans ces deux dernières classifications :

- Le travail de Gross présente des caractéristiques communes avec celui de Tournier. Ils sont tous deux en grande partie basés sur les catégories grammaticales des éléments des composés.
- La dernière classification de Gross, la plus précise, prend en compte la sémantique interne des composés, c'est-à-dire les relations entre les éléments (ce qui n'est pas le cas chez Tournier).

1.3. La typologie de Almela Pérez

Dans *Procedimientos de formación de palabras en español* (1999), Almela Pérez commente brièvement quelques typologies dont celles de Alvar Ezquerra (1995), Bustos Gisbert (1986), *Esbozo de una nueva gramática de la lengua española* (1973) et Darmesteter (1967) avant d'établir la sienne. Pour l'étude de la composition, l'auteur a choisi de partir d'une notion de base assez large – s'appliquant à différents groupes considérés comme des composés – et de l'affiner pour que la définition soit adaptée aux différents types de composés :

- Les composés juxtaposés
 - « Heterólogos de complementación »

V + N (*cazatalentos, lavacoches, metepatas*)

V + A (*matasanos*)

V + Adv (*mandamás*)

A + V (*altisonante*)

N + V (*terrateniente*)

Les noms composés anglais français et espagnols du domaine d'Internet :

- « Heterólogos de atribución »

N + i + A (*boquinegro, faldicorto*)

N + i + N (*palabrimujer*)

A + i + N

- « Homólogos determinativos »

N + N (*aguanieve, ajoaceite*)

N + A (*aguafuerte, nochebuena*)

A + N (*altorrelieve, malaleche*)

Numeral + N (*ciempiés*)

Numéral + A (*sietemesino*)

- « Homólogos copulativos »

A + A (*claroscuro, sacrosanto*)

A + i + A (*tragicomico, nochebuena*)

N + i + N (*coliflor, gallipavo*)

V + V (*ganapierde*)

V + i + V

- Les composés syntagmatiques

N + N (*hombre rana*)

N + A (*lengua materna, llave inglesa*)

A + N (*pájaro loco, alta mar*)

- Les composés prépositionnels

N + de + N (*conflicto de interés*)

N + a + N (*olla a presión*)

N + por + N (*televisión por cable*)

N + para + N (*chico para todo*)

N + con + N (*café con leche*)

N + en + N (*fabricación en serie*)

- Les pro-composés (*xenofobia, homeopatía, hidroterapia*)

La typologie d'Almela Pérez est, elle aussi, en partie basée sur la nature grammaticale des constituants des composés. Elle repose sur l'intégration de quatre types de composition qui ont un ou deux niveaux de composés. Le premier type possède deux

niveaux (nature du composé identique/différente de celle du déterminé et nature des éléments du composé), les deux suivants en possèdent un seul (nature des éléments) et le dernier n'a pas de sous-type.

Elle met l'accent sur les noms des différents types de formation de composés. L'auteur donne même des synonymes : les *juxtaposés* reçoivent également le nom de *composés orthographiques, propres, stricts, lexies composées* ; les *syntagmatiques* ceux de *composés libres, composés syntagmatiques binominaux* ; les *prépositionnels* ceux de *lexies complexes, synapsies, composés imparfaits, locutions dénominatives complexes* et les *pro-composés* ceux de *pseudo-affixes, formes combinées, demi-mots, composés néoclassiques, recomposés*, etc.

2. Typologies conçues pour des domaines de spécialité

2.1. La typologie de Depierre

Dans *La Composition en anglais scientifique : application au domaine de la physico-chimie* (2001), Depierre analyse les groupes nominaux anglais complexes pour faciliter la compréhension, la traduction et la production d'écrits scientifiques des étudiants et des chercheurs francophones et constituer un glossaire informatisé d'après son corpus. Depierre propose deux classifications (2001 : 107 sq.).

a) Classification morphologique

- Composés binaires sans élément verbal

A + N (*black hole, dry ice, cold fusion*)

N + N (*neon tube, electron diffraction, riprap*)

Npr + N (*Planck('s) constant, the Bohr atom*)

Lettre, sigle + N (*X-ray, UV radiation*)

- Composés binaires avec élément verbal

- avec V-er/or

N + V-er/-or (*circuit-breaker, bridge rectifier*)

Msav + V-er/-or (*microprocessor, radio receiver*)

V-er/-or + N (*starter gas, killer satellite*)

V-er/-or + V-er/-or (*emitter follower, converter reactor*)

Adj + V-er/-or (*magnetic amplifier*)

- avec V-ing

N + V-ing

- Adj + V-ing
- V-ing + Nom
- avec V-en
- V-en + N
- N + V-en
- Adj + V-en
- Adv + V-en
- Prép + V-en
- V-en + V-ing
- V-en + sigle
- Pronom + V-en
- avec V
- V + Part (*breakthrough, burn-up*)
 - Part + V
 - V + V
 - V+ V-er
 - Msav+ V
- Composés adjectivaux
 - le déterminé est un adjectif
- N + A (*salt-free, waterproof*)
 - A + A (*dark-red, white-hot*)
 - Autres (*freezing cold, all-electric*)
- le déterminé est un nom
 - N + N (*chain reaction, full-air*)
 - A + N (*high-energy, open-chain*)
 - Part/prep + N (*in phase, underweight*)
 - le déterminé est un N-ed
 - A + N-ed (*close-grained, fine-grained*)
 - Nombre + N-ed (*four-footed, five-sided*)
 - N + N-ed (*egg-shaped, lynx-eyed*)

- le déterminé est un V-ing

N + V-ing (*electron-accepting, chlorine-carrying*)

A + V-ing (*dry-cleaning*)

Adv + V-ing (*high-boiling, fast-moving*)

Pronom + V-ing (*self-catering, all-embracing*)

- le déterminé est un V-en

N + V-en (*battery-operated, air-borne*)

Adj + V-en (*long-lived, short-lived*)

Adv + V-en (*oppositely charged, randomly linked*)

Prép/part + V-en (*overheated*)

Pronom + V-en (*self-induced*)

V-en+part/prép (*grown-up, wound-up*)

b) Classification sémantique

Les noms composés anglais français et espagnols du domaine d'Internet :

	Relation syntaxique, sémantique	Types morphologiques avec déverbal	Exemples avec déverbal	Exemples sans déverbal
1	Verbe -COD	N + V-er V-er + N N + V-ing Adj + V-ing V-en + N N + V	<i>circuit-breaker</i> <i>carrier storage</i> <i>phase-mixing</i> <i>global warming</i> <i>absorbed dose</i> <i>handshake</i>	
2	Sujet -verbe	V-er + N V-ing + N V + N	<i>carrier</i> <i>concentration</i> <i>bonding energy</i> <i>surge power</i>	
3	Locative	N + V-er, V-ing + N N + V-ing Adj + V-ing	<i>fusion reactor</i> <i>melting pot</i> <i>laboratory testing</i> <i>clinical testing</i>	<i>water waves</i>
4	Locative (température)	V-ing + N	<i>melting point</i>	<i>dew point</i>
5	Moyen	V-er + N N + V-ing Adj + V-ing N + V-en	<i>computer graphics</i> <i>air-cooling</i> <i>hot-pressin</i> <i>gas-cooled</i>	<i>steam engine</i>
6	Destination, fonction	V-ing + N Adj + V-ing	<i>stirring rod</i> <i>protective clothing</i>	<i>current balance</i>
7	But	V + V	<i>test drive</i>	
8	Manière	N + V-ing Adj + V-ing V + N	<i>vacuum</i> <i>packaging</i> <i>dry</i> <i>cleaning</i> <i>strike</i> <i>force</i>	
9	Cause	N + V-ing	<i>resonance</i> <i>scattering</i>	<i>heat death</i>
10	Identification	V-er + N	<i>indicator diagram</i>	<i>daughter</i> <i>product</i> <i>particle</i> <i>physics</i>
11	Possession	V-er + N	<i>killer instinct</i>	<i>line profile</i>
12	Méronymie	N + V-er V-er + N N + V-ing	<i>well counter</i> <i>capacitor</i> <i>microphone, lung lining</i>	<i>spring balance</i> <i>dust core</i>
13	Caractéristique saillante/ recharge sémantique	V-ing + N	<i>cooling method</i>	<i>gas constant</i>

14	forme	Adj + V-ing	<i>toroidal winding</i>	<i>ring current</i>
15	extraction	V-ing + N	<i>lightning stroke</i>	<i>waveband</i>
16	résultat	V + N	<i>skim milk</i>	
17	Alternance	V + V	<i>freeze-thaw</i>	
18	Coordination			<i>volt-amperes</i>
19	Comparaison			<i>cradle guard,</i> <i>cascade shower</i>
20	Unité de mesure			<i>electronvolt</i>
21	Relations complexes	Adj + V-ing	<i>electric braking,</i> <i>spatial filtering</i>	<i>field coil, heat capacity</i>

La première classification est composée de trois types. Les deux premiers sont marqués par la présence ou l'absence d'un verbe. Les sous-types sont créés selon la nature grammaticale précise des constituants. Les composés adj ectivaux sont regroupés selon la nature grammaticale du déterminé.

La deuxième classification comprend 21 types répartis selon la relation qui unit les éléments des composés (colonne 1). Depierre énumère les types morphologiques avec déverbal (colonne 2) qui utilisent ces relations et donne des exemples de composés avec élément déverbal (colonne 3) et sans élément verbal (colonne 4). Pourquoi ne pas consacrer une colonne spéciale aux types morphologiques sans élément verbal puisque les exemples de la colonne 4 ne correspondent pas aux types morphologiques annoncés dans la colonne 2 (par exemple, relations n°4, 5, 6, 14, 15) ?

Ces typologies appellent quelques remarques :

- L'étude sémantique tient compte de la morphologie des composés mais l'étude morphologique n'inclut pas la sémantique. La typologie sémantique ne reprend pas tous les types morphologiques.
- Sont sur le même plan les composés binaires avec ou sans élément verbal qui sont autonomes dans la langue (*black hole, radio receiver*) et les composés adj ectivaux qui n'ont pas un statut indépendant (*salt-free, five-sided*) – ils sont les déterminants d'un nom. Pourquoi les inclure ici alors que les surcomposés et les composés néo-classiques sont étudiés séparément ?
- Pourquoi créer une catégorie pour les noms ayant les suffixes *-er* et *-ing* (V-*er*, V-*ing*) et se désintéresser des noms comportant d'autres suffixes (*-ion, -ement, -age*) comme dans *carrier concentration* (type n°2), *carrier storage* (type n°1) ?

2.2. La typologie de Béciri

La typologie de Béciri concerne un domaine connexe à Internet, l'informatique. Dans sa thèse, *La Néologie dans le domaine microinformatique professionnel en français et en anglais : création lexicale et explication en contexte* (1999), Béciri analyse les particularités de la création lexicale et les procédés de transmission du sens dans les discours techniques intermédiaires de la communication spécialistes/utilisateurs en

français et en anglais. L'auteur consacre quelques pages à la structure morphosémantique des noms composés dans la partie consacrée aux procédés de création lexicale. Elle dégage les types suivants :

a) Typologie des types formels

- Composition syntagmatique

N + prép + N (*table d'allocation de recherche*)

- Composition par juxtaposition de deux éléments

N + N

N + A

N + participe adj

b) Typologie des relations syntaxico-sémantiques

- Composés situatifs

- Description

N + A (*coupure automatique, menu déroulant*)

N + N (*document maître, logiciel compagnon*)

Certains N + Adv (*dictionnaire en ligne*)

Identité (X-Y = un X qui est aussi un Y)

N + N (*ligne-repère, mot-bouton*)

- Qualification

N + A (*document composite, version intégrale*)

N + Vpp (*aide intégrée, fenêtre indentée*)

N + prép + N (*aide en ligne, menus en cascade*)

N + N (*composition type, page maîtresse*)

N + Vppr (*fenêtre recouvrante, menu déroulant*)

NC + A (*interface utilisateur*)

- Relations complexes

N + prép + N (*champ de texte, attribut de caractères*)

N + N (*interface utilisateur, cadre-texte*)

NC + prép + N (*entrée de texte au kilomètre*)

N + A (*mode graphique*)

N + NC (*mode noir au blanc, vue pleine page*)

N + p + N + p + N (*table d'allocation de fichier*)

N + p + NC (*virus du secteur de démarrage*)

- Composés actanciels (relation prédicative)

- Action

N + p + N (*capture d'écran, gestion de plan*)

N + N (*compression-décompression, mode apprentissage*)

Vcomp + A (*couper/coller glissant*)

N + p + NA (*échange de données dynamique*)

N + A (*correction orthographique, enregistrement automatique*)

N + et + N + p + N (*lien et encapsulation d'objets*)

NC +A (*mise en forme directe*)

N + N + N (*mode Affichage-Révision*)

N + NC (*mode d'entrée de texte, mode mise en page*)

- Outils

N + p + N (*accélérateur de clavier, case de pointage*)

N + p + N + p + N (*barre d'aire à la fusion, marque de fin de paragraphe*)

N + N (*case Icône, icône texte*)

N + NC (*case plein écran*)

N + A (*correcteur orthographique, lien dynamique*)

N + p + NC (*éditeur de fonctions scientifiques, éditeur de formules mathématiques*)

NC + A (*environnement d'exploitation graphique*)

N + p + N + N (*fichier d'information programme*)

N + A + p + N (*interface graphique d'exploitation*)

Dans sa classification des types formels, l'auteur commente uniquement les deux grands types et quelques sous-types fréquents. À l'instar d'Almela Perez (1999), Béciri distingue la composition syntagmatique de la composition par juxtaposition.

L'étude syntaxico-sémantique de Béciri est différente de celles menées habituellement. La plupart des typologies sémantiques retracent la diversité des relations sémantiques simples (moyen, but, manière, possession, etc. ; voir Depierre, Part. I, Ch. 1, Sec 2.1.). Béciri (1999 : 141) précise que les composés situatifs (*coupure automatique, logiciel compagnon*) se caractérisent par « un rapport de qualification de nature statique : il s'agit d'une structure coordonnante qui va servir par exemple à décrire un objet, un produit, une fonction ». Les composés actanciels (*gestionnaire de X, capture de X*) se caractérisent par « la présence entre les éléments d'une relation prédicative de type

actanciel : le schéma sous-jacent réfère dans ce cas à une opération. Les éléments [...] sont le plus souvent des noms d'agent ou d'action, ou des dérivés verbaux » (1999 : 141).

Deux remarques s'imposent :

- La première catégorie du classement des types formels – N + prép + N – englobe les composés binaires et les surcomposés. La même remarque est valable pour les juxtaposés.
- Tous les types formels qui figurent dans la typologie syntactico-sémantique ne figurent pas dans la première qui est morphologique.

L'analyse de cinq typologies démontre que si les auteurs parviennent à s'entendre sur les matrices lexicogéniques, ils sont en désaccord en ce qui concerne l'articulation et le contenu des divers types qu'elles renferment. Des auteurs ayant des buts semblables – la description pour Tournier et Almela Perez, par exemple – construisent parfois des typologies totalement différentes.

En conclusion, grâce au travail effectué dans ce chapitre, lors de la création de nos typologies (Part. II, Ch. 1, Sec. 3) nous disposerons de plusieurs critères utilisables conjointement : le nombre d'éléments, les relations syntaxiques et sémantiques entre les éléments, les catégories grammaticales des formants, la catégorie grammaticale du composé, l'ordre de détermination, la productivité des types (Almela Perez, 1999 : 25). Nous pourrons opérer une pré-sélection parmi ces critères. Le choix des catégories des classifications et la définition du degré de précision à adopter sera fonction de nos deux objectifs : la création d'un système d'équivalences types et la description des composés du lexique d'Internet.

Deuxième partie. Les noms composés anglais, français et espagnols du lexique d'Internet

Chapitre 1. Le corpus : constitution et contenu

1. Travail terminologique et corpus

The decisions that are taken about what is to be in the corpus, and how the selection is to be organized, control almost everything that happens subsequently. The results are only as good as the corpus (Sinclair, 1991 : 13).

1.1. La nécessité d'un corpus

La notion de corpus, bien qu'utilisée depuis quelques années, est loin de faire l'unanimité. Dans *Corpus*, *Concordance*, *Collocation*, Sinclair définit le corpus comme « a collection of naturally-occurring language text, chosen to characterize a state or variety of a language » (1991 : 171-172). Bowker et Pearson insistent sur quatre caractéristiques : un corpus est

un assemblage **important** de textes **authentiques** compilés sous format **électronique** selon des critères **précis** (2002 : 9).

Étudier les noms composés de la terminologie d'Internet sans un corpus de textes serait insensé pour deux raisons. Premièrement, en collectant des composés au hasard des rencontres, des termes dénommant des concepts fondamentaux pourraient être omis. Un corpus bien construit est un fil conducteur permettant de voyager à travers les différents sous-domaines. Pour analyser le vocabulaire d'un domaine il faut une source dans laquelle puiser le lexique. Les phases de repérage et de description des éléments recherchés passent généralement par le dépouillement systématique de revues, de dictionnaires ou de livres dédiés au sujet car les termes y sont repérables et le contexte éclaire leur sens. C'est ainsi qu'ont procédé Béciri, Depierre, Gross, Tournier. Leurs travaux sont basés sur des corpus. Deuxièmement, un corpus est une source d'information contenant des extraits de textes, de discours qui reflètent ainsi l'usage. Représentant le discours de plusieurs personnes, un tel assemblage de données authentiques est une excellente base pour procéder à des analyses, fonder une méthode et des résultats.

Les corpus de textes spécialisés présentent, en outre, l'avantage de fournir des explications sur les termes (Kocourek, 1991 : 23). On ne peut pas étudier le vocabulaire d'un domaine sans en avoir une connaissance approfondie. Cela est d'autant plus important pour les disciplines techniques dans lesquelles les termes désignent des concepts précis qu'il faut identifier pour pouvoir reconnaître les termes qui sont des synonymes, des variantes, des raccourcissements. Les auteurs de termes techniques, soucieux de leur précision et de la compréhension des lecteurs, les définissent, utilisent des introducteurs, recourent aux guillemets, à l'italique.

Lors de la constitution d'un corpus, la sélection de textes doit tenir compte de l'objectif poursuivi. La collecte peut être nécessitée par des buts très variés comme des « divulgations, dans les revues qui informent systématiquement les spécialistes [...] », la « confection de dictionnaires » (Sablayrolles, 2000 : 250) ou la description du lexique d'un domaine et la création d'un outil d'aide à la traduction dans notre cas.

1.2. Quel type de corpus constituer ?

1.2.1. Les différents types

Bowker et Pearson (2002 : 11-13) distinguent plusieurs genres de corpus :

- Corpus de référence (« general reference corpus ») et corpus spécialisé (« special purpose corpus »). Un corpus de référence reflète une langue et permet de faire des observations d'ordre général. Ce type de corpus contient des données orales et écrites, c'est un mélange de plusieurs textes de différentes natures (journaux, fiction, reportages, débats, etc.). Un corpus spécialisé est axé sur l'aspect particulier du vocabulaire d'un domaine, sur un certain type de textes, sur le langage des membres d'un groupe social (les jeunes, la communauté hispanophone de la Californie, etc.). Utilisés conjointement, un corpus de référence et un corpus spécialisé peuvent

permettre d'identifier les différences entre un langage spécialisé et la langue générale.

- Corpus écrit et corpus oral. Un corpus oral est constitué de transcriptions de discours oraux (conversations, émissions, cours, etc.) et un corpus écrit contient des textes destinés à être lus. Certains corpus, comme le British National Corpus, sont un mélange des deux.
- Corpus monolingue et corpus multilingue. Contrairement à un corpus monolingue, un corpus multilingue rassemble des textes dans au moins deux langues. Les corpus multilingues peuvent être divisés en corpus parallèles et en corpus comparables. Un corpus parallèle est constitué de textes en langue source accompagnés de leurs traductions en langue cible. Un corpus comparable ne contient pas des traductions mais des textes écrits en langue source qui ont tous le même sujet, la même fonction de communication, qui sont de même nature (manuel d'instruction, rapport technique, etc.) et ont des dates de publication proches.
- Corpus synchronique et corpus diachronique. Un corpus synchronique est une « photo » de l'usage d'une langue pendant un temps limité. Un corpus diachronique sert à mesurer l'évolution d'une langue sur une longue période.
- Corpus ouvert et corpus clos. Un corpus ouvert (« monitor corpus ») est constamment étendu. C'est le type généralement utilisé en lexicographie. Un corpus clos reste tel quel. Une fois qu'il est compilé aucun texte n'est ajouté.
- Corpus « apprenants » (« learner corpus »). Ce type de corpus contient des textes écrits par les apprenants d'une langue étrangère. Il est intéressant pour effectuer des comparaisons avec des textes écrits par des natifs. Il fait ressortir les erreurs types des apprenants.

Les termes, qui seront au cœur de la présente étude, seront extraits d'un corpus :

- trilingue. Le corpus est composé de textes anglais, français et espagnols,
- comparable. Les textes anglais, français et espagnols abordent les mêmes thèmes,
- synchronique. Les revues datent de 2001 et 2002,
- spécialisé. Les spécialistes sont les internautes,
- écrit. Les textes ont été produits dans le but d'être lus,
- clos. Aucun ajout de nouveaux textes n'aura lieu avant la fin de cette recherche. Les articles constituant notre corpus ont été choisis car, au travers des thèmes abordés – navigation, moteur de recherche, sécurité, piratage, e-commerce, réseau, messagerie instantanée, courrier électronique, chat, etc. –, ils touchent un très vaste public.

1.2.2. La nature des textes

Le domaine considéré ici se prête mal à un corpus oral tant sur le plan de la faisabilité que de la pertinence. Les internautes communiquent par le biais de leur ordinateur. Leur communication orale est réduite, contrairement à celle de biologistes travaillant en

laboratoire, par exemple. Enregistrer des cours d'informatique prendrait plusieurs mois pour obtenir un nombre suffisant de composés du domaine d'Internet. Visiter des forums de discussion ou des messageries instantanées et noter les termes employés aurait pu être une solution. L'obstacle majeur de cette méthode est que le registre de discours de ces sites est assez familier⁶¹ et les discussions, elles aussi décontractées, regorgent d'allusions, de « private jokes », de troncations et d'abréviations qui ne sont pas toujours connues ou comprises par la majorité des internautes (Brolles 2001 : 5).

La nature des textes à sélectionner est liée aux objectifs à atteindre. Pour que notre système d'équivalences types soit pertinent, il doit contenir les termes du lexique réel, ceux utilisés par les internautes. Les glossaires et les dictionnaires seront donc exclus de nos sources. Ils peuvent avoir des ambitions normatives, prescriptives et faire figurer des termes non employés par les internautes.⁶² De plus, les dictionnaires spécialisés ne saisissent pas tout le lexique spécialisé. Dans un dictionnaire, général ou spécialisé, ne figurent pas tous les termes utilisés dans une revue technique ou scientifique (Kocourek, 1991). Compiler un dictionnaire prend des années et les domaines technico-scientifiques évoluent avec une telle rapidité que lorsque le dictionnaire est publié des termes nouveaux sont absents et des termes inusités sont répertoriés. Peu de dictionnaires sont entièrement consacrés à Internet. En règle générale, les termes d'Internet sont inclus dans les dictionnaires informatiques.⁶³ La macrostructure du *Dictionnaire du Multimédia* (Colombain, 1998) comporte 469 entrées et toutes ne concernent pas Internet (et moins de deux cents entrées sont des composés). La taille du dictionnaire est une autre contrainte des listes de termes basées sur dictionnaires (Bowker et Pearson, 2002 : 15). Un dictionnaire qui couvre la totalité d'un domaine peut être constitué de plusieurs volumes. Son prix et sa taille sont des obstacles à son achat et à son utilisation.

Une autre solution serait de se tourner vers les livres, les manuels, les notices d'utilisation et les cours plutôt que vers les dictionnaires. Collecter les composés dans des manuels d'apprentissage ou des cours ne nous conviendrait pas, même si leur vocabulaire est assez abondant, car ils utilisent une terminologie de base. Ce sont des ouvrages de vulgarisation qui excluent les notions complexes. Quant aux livres, ils ne seront pas pris en compte pour limiter les risques d'influence individuelle (par exemple, pour désigner un concept un auteur français peut préférer utiliser un terme endogène alors que le terme le plus fréquemment employé dans le domaine est un emprunt). Pour

⁶¹ Joos (1967) décrit cinq styles de discours basés sur le degré de formalité (ordre croissant) : « intimate » (intime ; style employé entre époux, enfants, amis de longue date), « casual » (familier, décontracté ; emplois : entre amis, collègues, connaissances, lors d'activités de loisirs), « consultative » (consultatif ; emplois : entre inconnus, lors de réunions ou de transactions), « formal » (soutenu ; emplois : présentations, exposés) et « frozen » (oratoire, « glacé » ; emplois : publications, discours publics, œuvres littéraires). Le style familier se caractérise par l'utilisation de mots d'argot et d'ellipses.

⁶² Par exemple, les glossaires de l'OLF et la DGLF évitent les anglicismes.

⁶³ En mai 2004, une recherche avec le mot clé « dictionnaire Internet » sur le SCD de la bibliothèque de l'université Lyon 2 a comme résultat : *Dictionnaire bilingue de la micro-informatique et de l'Internet* (Marabout, 1997), *Dictionnaire de l'informatique et de l'Internet* (Micro Application, 1999), *Dictionnaire français/anglais, anglais/français : cinéma, audiovisuel, son, multimédia, réseaux* (Dirit, 2001), *Grand dictionnaire Marabout de la micro-informatique et de l'Internet* (Marabout, 1996).

être sûr que les termes sélectionnés sont reconnus et employés par les experts du domaine, il faut consulter plusieurs livres. Or, les particularités linguistiques peuvent concerner plusieurs documents (Bowker et Pearson, *ibid.* : 17). Sélectionner de courts extraits est une solution mais les livres traduits étant nombreux, elle n'est pas envisageable. De plus, dans les livres comme dans les dictionnaires, certains termes peuvent avoir disparu avant la publication du livre (qui a lieu plusieurs mois au moins après l'écriture). En ce qui concerne les notices d'utilisation, les usagers les lisent rarement ou les parcoururent seulement. Le discours étant très soutenu, les termes rencontrés ne sont pas forcément ceux employés par la majorité des internautes.

Parmi les différentes sources mentionnées, seules restent les revues spécialisées. Les composés étudiés seront recueillis dans des revues écrites par des experts du domaine : les internautes. Les thèmes variant en fonction des rubriques, des mois et des revues, les articles permettront de recueillir l'essentiel des termes du domaine.

2. La constitution du corpus

2.1. Qui sont les experts d'Internet ?

Contrairement à des domaines comme la radiologie, la biologie ou la météorologie, il est délicat d'identifier les spécialistes du domaine d'Internet. Les acteurs du Réseau ont des buts et des préoccupations très hétérogènes – création de protocoles, de logiciels de navigation, de sites Web, etc. Brolles (2001 : 3-4) distingue deux types de candidats susceptibles d'endosser le rôle d'experts : les concepteurs de logiciels Internet et les pirates informatiques.

Les concepteurs de logiciels Internet, comme Napster, Netscape ou Eudora Email, peuvent être assimilés aux concepteurs de logiciels informatiques. Tous deux se caractérisent par leur qualité de programmateur. Il n'y a donc aucune garantie qu'ils soient des experts d'Internet et les sociétés informatiques ne font pas autorité sur le Réseau. Quant aux hackers, ils créent de nombreux néologismes, principalement pour nommer les acteurs du Net (EN *cracker*, FR *développeur Web*, etc.), les virus (Adshow, african 109, loveletter, Christmas Tree, etc.) ou les infractions commises (FR *attaque*, *intrusion*, EN/FR *spamming*, etc.) mais leur communication passe beaucoup par un langage crypté qui semble réservé à une minorité d'utilisateurs. Devant les difficultés à nommer un groupe d'acteurs 'experts d'Internet', la solution est de considérer les internautes – les véritables utilisateurs – comme des spécialistes. En arpantant la Toile, ce sont eux qui font le succès d'un site, d'un logiciel et diffusent le vocabulaire.

2.2. Le degré de spécialisation des textes

Lorsqu'ils communiquent entre eux, les spécialistes utilisent un vocabulaire spécifique. Ils comprennent le sens précis des termes du domaine et ne donnent pas d'explications aux lecteurs. Les spécialistes peuvent aussi s'adresser à des semi-spécialistes (étudiants, spécialistes de domaines connexes). Ils utilisent les mêmes termes que précédemment mais ils les expliquent pour permettre aux semi-spécialistes de comprendre et d'assimiler

les concepts. Enfin, la communication peut avoir lieu entre des spécialistes et des non-spécialistes. Les experts utilisent moins de termes ou emploient des mots du langage général pour décrire les concepts.

Les internautes n'ont pas tous le même niveau de compétence et les magazines d'Internet n'ont pas tous le même objectif et ne s'adressent pas au même lectorat. Voici une brève présentation des types rencontrés :

- Type n°1 : Magazines adressés à un public de débutants.

Ce sont principalement des aides à la navigation. Leurs thèmes de prédilection sont la présentation de sites Web, les astuces pour maîtriser les logiciels de courrier électronique, les moteurs de recherche, etc. Exemples de magazines français : *Web magazine*, *Internet Today*, *Netscope*.

- Type n°2 : Magazines destinés à un public expérimenté.

Leur lecture nécessite des connaissances en informatique. Ces revues sont construites sur le modèle des revues informatiques et sont axées sur la pratique. Les thèmes qu'elles affectionnent particulièrement sont la création de sites web, la maîtrise des codes (HTML, Java), la sécurité, etc. Exemples de magazines français : *Netsurf*, *.Net*, *Officiel du Net*.

- Type n°3 : Revues adressées à un public adepte de l'« underground ». ⁶⁴

Il est conseillé d'avoir des notions en informatique pour lire ces revues qui traitent du piratage informatique. Elles relatent les « exploits » et les attaques exécutées par des hackers, parlent des sites qui présentent des failles de sécurité, des logiciels indispensables pour se protéger contre les virus, les attaques et les intrusions. Exemples de magazines français : *Pirates mag*, *Zataz*, *Hackermag*, *NetBug*, *Piratez*.

D'après Brolles (2001), les deux premiers types de revues présentent, en règle générale, les mêmes termes. Tous deux seront donc retenus dans cette étude. Les revues destinées aux néophytes contiennent moins de termes que les autres mais cela ne constitue pas un inconvénient. Le milieu « underground » touchant moins d'internautes que les autres revues, le type n°3 n'entrera pas dans la constitution de notre corpus trilingue comparable.

2.3. La dimension du corpus

Pour la plupart des chercheurs, la taille du corpus est un des premiers facteurs à considérer. Sinclair préconise, pour des recherches sur le lexique, de composer un corpus ouvert qui contient des millions de mots parce que la moitié du vocabulaire d'un texte, même très long, peut être constituée de lexies qui n'apparaissent qu'une seule fois (1991 : 18). Pour étudier leur comportement en contexte, il faut disposer d'un grand nombre d'occurrences. De plus, dans un texte, certains des mots qui reviennent le plus

⁶⁴ Plusieurs types de personnages constituent le milieu underground : les hackers, les script kiddies, les phreakers, les crackers, les carders, les lamers et les crashers. Toutes ces catégories se divisent en trois niveaux : les newbies (débutants), les wild (cloudy (36), partly (29), low (25), and (22), highs, lows (21), to (19), mph, upper, winds (expérimentés), les élites. Source : <http://www.infoshackers.com/> le 03/07/2003.

(12), *near* (11), *with* (10), *of* (9) (Bowker et Pearson, 2002 : 31-32).

Bowker et Pearson nuancent le critère de la taille du corpus : « It is very important not to assume that bigger is always better » (2002 : 45). Un petit corpus bien construit peut contenir plus d'informations qu'un vaste corpus ne répondant pas aux besoins du projet. Étudier le vocabulaire de l'industrie à l'aide du British National Corpus (BNC) n'est pas très pertinent : les occurrences de EN *nuts* sont douze fois plus nombreuses dans le BNC que dans un corpus spécialisé mais le BNC ne donne pas d'indications sur les types de *nuts* (*collar nut*, *compression nut*, *flare nut*, *wing nut*) (Bowker et Pearson, 2002 : 46-47). Il n'existe pas de règles pour déterminer la taille idéale d'un corpus. Il faut prendre en considération les besoins du projet, la disponibilité des sources, le temps et les moyens dont on dispose. Cependant, il est reconnu que les corpus utilisés pour étudier les langues de spécialité peuvent être de taille plus modeste que ceux utilisés pour analyser la langue générale.

Dans le présent travail, le but est différent de celui exposé par Sinclair. Il n'est pas question de calculer la fréquence des termes d'Internet, ni d'étudier leur comportement en contexte. Il est inutile de rassembler des milliers de textes extraits de centaines de magazines. Quelques revues seront suffisantes. Ensuite, le repérage et l'extraction des termes seront effectués de manière empirique⁶⁵ – comme les saisies dans les différentes typologies et la base de données du système d'équivalences – la taille du corpus ne pourra être que modeste.

Les textes seront extraits de 22 revues anglaises, 12 revues françaises et 13 revues espagnoles (voir Annexe 5 et Annexe 6 pour le détail des numéros et des rubriques). Le gain de temps obtenu par la non-informatisation de nos corpus, c'est-à-dire par la suppression des phases de numérisation,⁶⁶ de nettoyage, de segmentation des unités sera mis à profit pour la construction du système d'équivalences types et son informatisation.

2.4. La période

Il a semblé judicieux de constituer un corpus très récent dans la mesure où le réseau des réseaux évolue constamment. Un flot incessant de néologismes apparaît pour nommer les innovations et les techniques. La phase de sélection des textes a débuté en 2001, année marquant le début de cette étude. Pour des raisons pratiques, les numéros de 2001 de certains magazines espagnols n'ont pu être obtenus.⁶⁷ Quelques parutions sont donc plus récentes et datent du début de 2002.

⁶⁵ Avoir un corpus informatisé n'est pas essentiel pour nous, surtout si l'on considère que l'utilisation d'un concordancier ne permettrait pas d'améliorer le repérage des noms composés.

⁶⁶ Nous avons été confrontée à un problème d'accès aux sources pour la constitution du corpus espagnol. Il est presque exclusivement composé de photocopies de revues commandées à la BNE. Il serait chronophage de numériser chaque article et de procéder aux corrections puisque nous ne travaillons pas sur la fréquence.

⁶⁷ La BNE ne dispose pas de tous les numéros de l'année 2001 pour les revues qui nous intéressaient.

2.5. Les revues

Comme dans Ahronian (2001), pour chaque langue, les recherches reposent sur plusieurs magazines pour éviter les politiques puristes ou laxistes en matière d'emprunts. Les articles sont rédigés par des spécialistes/internautes travaillant pour des revues différentes et publiées par des maisons d'édition distinctes. Des fiches présentant les revues anglaises, françaises et espagnoles sélectionnées se trouvent en Annexe. Elles contiennent cinq rubriques : numéros retenus, type de la revue, lectorat ciblé, raisons du choix de la revue et nom de la maison d'édition (Annexe 5).

2.6. Les rubriques

Les textes dont seront tirés les composés seront sélectionnés d'après deux critères :

- Leur thème : la navigation, les moteurs de recherches, le téléchargement, les modes de connexion, la création de sites, le piratage, la sécurité, les virus, l'e-mail, les forums, le chat, le peer-to-peer, les startups, l'e-business, les articles « trucs et astuces ». Les thèmes retenus sont les thèmes essentiels du domaine et ceux qui intéressent le plus les internautes.
- Les publicités et les offres commerciales ne seront pas analysées en raison des jeux de mots ou des formules journalistiques utilisées pour attirer l'attention des acheteurs.

En ce qui concerne la structure des articles, ils peuvent être courts et très riches en termes (trucs et astuces) ou être plus longs, entre deux et quatre pages (dossiers relatifs au thème d'un numéro, interviews d'acteurs du Web). Quant à leur visée, les articles choisis ont pour but d'informer le lecteur (nouveaux sites, nouveaux modes de piratage, nouveaux logiciels, effets d'un virus), de le former (comment créer un site, comment utiliser un langage), d'expliquer (méthode utilisée par des pirates) ou de comparer (des logiciels, des sites).

Les rubriques retenues et rejetées pour chaque revue sont listées en Annexe (Annexe 6).

2.7. Le contenu du corpus

Notre corpus nous a permis de constituer une liste de 553 noms composés anglais, 649 composés français et 576 composés espagnols.⁶⁸

Les données descriptives et statistiques sur les noms composés (Part. II, Ch. 2, 3 et 4) et le système d'équivalences types (Part. III) seront pertinents pour leurs utilisateurs qui, rappelons-le, sont toutes les personnes qui ont besoin de produire – et parfois de comprendre – des textes en rapport avec Internet.

⁶⁸ Le nombre plus important de composés français et espagnols par rapport au nombre de composés anglais s'explique notamment par la présence de dénominations synonymiques dans les deux langues cibles. Par exemple, le terme anglais *e-mail* a deux équivalents français : *e-mail*, *e-courrier*. Une liste trilingue des noms composés figure en Annexe (Annexe 7).

3. Typologie morphosyntaxique des noms composés du corpus

À partir de l'étude définitionnelle et de l'analyse de typologies (Part. I), nous allons créer une typologie morphologique trilingue qui sera la pierre angulaire de notre étude. Elle sera commentée et complétée par un triple classement sémantique dans les trois chapitres suivants (Ch. 2, 3 et 4) et la section anglaise de cette typologie morphosyntaxique sera au cœur du système d'équivalences types présenté dans la troisième partie.

3.1. Les difficultés de classement

Quiconque souhaite organiser les noms composés d'une langue ou d'une terminologie se heurte à différents obstacles. Certains sont intrinsèques à la langue, d'autres sont liés aux buts que la typologie doit permettre d'atteindre. La principale difficulté de classement des constituants provient d'une ressource qu'ont certaines langues : la conversion. Par exemple, pour le terme EN *startup*, l'étiquette appropriée à l'élément *start* est-elle V ou N ? Nous ne faisons pas l'impasse sur ce problème pour établir notre typologie mais comme il n'est pas spécifique d'Internet, nous ne donnerons pas de détails.

Parmi les critères dont nous disposons (syntaxique, grammatical, endocentrique/exocentrique, humain/non humain, le nombre d'éléments, la relation entre les éléments, l'ordre de détermination, la fréquence des patrons syntaxiques), quels sont ceux qui seront utiles pour l'élaboration du système d'équivalences types ?

3.1.1. Critères grammaticaux

Ces critères sont au nombre de deux :

- La nature grammaticale du composé. La seule catégorie étudiée ici est la catégorie nominale. Elle sera ensuite découpée en sous-catégories.
- La nature grammaticale des éléments formant le composé. Le point de départ de la création des équivalents français et espagnols proposés par le système d'équivalences est la nature des formants des composés anglais. Ainsi, notre typologie sera-t-elle basée sur la nature grammaticale du déterminé et du déterminant des composés : N + N, N + A, N + prép. + N, par exemple.

3.1.2. Critère de l'ordre de détermination

Peu de classements tiennent compte de l'ordre de détermination (Part. I, Ch. 3, Sec. 1.2.). Pour la majorité des noms composés, l'ordre est déterminant-déterminé (N2 + N1) en anglais et déterminé-déterminant (N1 + N2) en français et en espagnol. Inclure cet aspect dans notre typologie n'est pas indispensable pour deux raisons. Premièrement, dans la plupart des cas l'ordre des éléments des équivalents cibles sera l'inverse de celui des composés sources si l'équivalent est un composé. Dans le cas où le composé anglais est de type N2 + N1, l'équivalent français ou espagnol sera N1 + N2 (EN *electronic mail* = FR *courrier électronique* et ES *correo electrónico*). Deuxièmement, nous envisageons

l'informatisation du système d'équivalences types et il semble impossible de programmer une application de façon à ce qu'elle élabore des équivalents cibles en fonction de la position tantôt modifieur, tantôt modifiant qu'une même lexie peut avoir à l'intérieur d'un composé source (dans EN *book mark*, l'équivalent du déterminant EN *book* est FR *page* alors que dans EN *guest book*, l'équivalent du déterminé EN *book* est FR *livre*).

3.1.3. Critères de centricité

L'endocentrisme/exocentrisme des composés est un facteur important dans le cadre de la création d'un système trilingue d'équivalences types. Notre système reposera sur une typologie des patrons syntaxiques des noms composés. Du fait de leur particularité sémantique, les exocentriques, comme EN *bandwidth*, *bottleneck*, *thumbnail*, *firewall*, ne seront pas traités de la même manière que les endocentriques par le système d'équivalences types, qui proposerait des termes souvent inutilisables. Les éléments des exocentriques ne seront pas traduits séparément comme il est prévu de le faire avec les endocentriques car la présence de métaphores fait qu'on aboutirait à des termes opaques et difficilement implantables comme FR**largeur de bande*, **goulot de bouteille*, **ongle de pouce* et **mur de feu*. Nous anticipons les problèmes en envisageant de créer une catégorie « Exceptions » contenant les exocentriques. Pour repérer ces derniers, une triple typologie sémantique sera créée. Elle sera analysée dans les parties descriptives et servira tout particulièrement pour le travail préliminaire du système d'équivalences mais elle ne sera pas incluse dans l'outil d'aide à la traduction pour ne pas en alourdir le fonctionnement.

3.1.4. Critères syntaxiques

Aucune distinction ne sera opérée entre les différentes sortes de composition : syntagmatique, juxtaposition, néo-classique, amalgame, etc. Cela est sans importance pour la création d'équivalents telle que nous l'entendons. Seul compte le patron syntaxique des composés. Les relations syntaxico-sémantiques – composés actanciels/situatifs (Part. I, Ch. 3, Sec. 2.2.) – ne seront pas non plus introduites dans nos typologies.

3.1.5. Critère humain/non-humain

Ce critère, très bien exploité par Gross (Part. I, Ch. 3, Sec. 1.2.), ne sera pas utilisé ici. La grande majorité des termes de nos corpus désignent des non-humains.

3.1.6. Critère de fréquence

Une description des noms composés du domaine d'Internet en anglais, français et espagnol (Part. III) serait incomplète sans une analyse de la fréquence des types compositionnels⁶⁹ et, pour être efficace, le système d'équivalences types doit proposer des termes prenant en compte l'ordre de fréquence de types français et espagnols utilisés pour rendre tel ou tel patron syntaxique anglais. Deux types de fréquence seront donc

⁶⁹ Dans notre étude les termes *patron syntaxique*, *type compositionnel*, *structure compositionnelle* sont synonymes.

calculés :

- Dans l'analyse descriptive, nous calculerons la fréquence « type » des patrons syntaxiques pour chaque langue. Par exemple, les composés anglais N + N ont une fréquence x et les surcomposés A + N + N présentent une fréquence y.
- Pour le système d'équivalences types, nous établirons la fréquence des structures françaises et espagnoles utilisées pour rendre un patron syntaxique anglais donné. Par exemple, pour rendre les composés anglais N + N, x est la fréquence du type compositionnel français N + de + N, y est la fréquence de la structure N + A et z celle du patron N + N.

3.2. La constitution de notre typologie

3.2.1. Typologie morphosyntaxique : la base du système d'équivalences types

Notre typologie devra être opérante pour l'analyse descriptive et surtout pour l'élaboration du système d'équivalences types. Elle en constituera la toile de fond puisque les équivalents français et espagnols des noms composés anglais seront créés d'après le patron syntaxique de ces derniers : si NC⁷⁰ est de structure S_a alors la structure de E_f sera S_f et celle de E_e sera S_e. Le travail sera effectué à partir d'une typologie des noms composés anglais recueillis dans le corpus. Nous étudierons les équivalents français et espagnols correspondant aux noms composés anglais recueillis pour dégager les tendances de traduction en français et en espagnol. Ces tendances seront ensuite utilisées pour créer des équivalents aux néologismes composés anglais qui n'ont pas d'équivalents en français et en espagnol. Nous pouvons d'ores et déjà avancer, en ce qui concerne le vocabulaire existant, que les composés anglais sont le plus souvent rendus en français et en espagnol au moyen du calque.⁷¹ Si le patron de NC_a est N + N alors E_f et E_e seront créés d'après le patron N + de + N. Le tableau 6 résume la démarche envisagée.

Tableau 6 : Utilisation des patrons syntaxiques des noms composés anglais dans le système d'équivalences types

⁷⁰ NC_a = composé anglais, E_{f/e} = équivalent français/espagnol, S_a = structure syntaxique du composé anglais, S_{f/e} = structure syntaxique de l'équivalent français/espagnol.

⁷¹ Le calque est une variété d'emprunt sémantique où le signifiant du mot étranger n'est pas emprunté tel quel, mais où le mot étranger est traduit littéralement. [...] Les calques sont très souvent des mots construits, surtout des composés [...]. (Tournier, 1991b : 29). Un test du système d'équivalences sur les N + N anglais révèle que les équivalents français attestés sont presque exclusivement des calques (Part. III, Ch. 2, Sec. 2.1.).

	Composé anglais (NC_a)	Patron syntaxique anglais (S_a)	Structure française (S_f)⁷²	Equivalent français (E_f)	Remarques
1	<i>e-mail</i>	PS ⁷³ + N	PS + N	<i>e-courrier</i>	Le composé anglais présente l'ordre de détermination des composés français donc le patron syntaxique français peut être identique.
2	<i>address book</i>	N + N	N + de + N	<i>carnet d'adresses</i>	Le patron syntaxique N + de + N est très fréquemment utilisé en français pour rendre les N + N anglais.
3	<i>e-mail address</i>	PS + N + N	N + de + PS + N	<i>adresse de e-courrier</i>	S _{a3} est une combinaison de S _{a1} et S _{a2} . S _{f3} est calqué sur S _{a3} + modification de l'ordre de détermination.

Étant donné la méthode envisagée, il semble primordial de considérer tous les formants des composés, sans tenir compte de leur utilisation avant Internet ou pas, ni du fait que des composés sont réutilisés dans la formation de surcomposés. Par exemple, si EN *e-mail address* était de type N + N plutôt que PS + N + N parce que la combinaison PS + N ou N + N existait avant son inclusion dans le surcomposé alors le terme français serait de structure N + de + N, comme *adresse de courrier* et non N + de + PS + N, comme *adresse de e-courrier*. Le signifiant du terme français serait alors incomplet et désignerait un concept différent de celui auquel le terme anglais renvoie. Lorsque nous avons extrait les composés de notre corpus, nous avons inclus tous les composés rencontrés sans considérations diachroniques. Écarter les composés dont le signifié existait avant l'utilisation du signe pour Internet nous aurait conduit à éliminer des concepts importants du domaine désignés par des termes empruntés à d'autres domaines (métaphores) comme EN *firewall*, *mailbox*, *bandwidth*, *incoming mail*, *address book*, etc.

3.2.2. La structure principale

3.2.2.1. Les lexies incluses

⁷² Les structures y_f figurant dans cette colonne ne sont pas issues de statistiques. Elles ont uniquement une valeur d'exemple.

⁷³ PS = pseudo-confixe (Part. I, Ch. 1, Sec. 1.1.)

Rappelons que, conformément à notre définition du nom composé, seront inclus dans nos typologies :

- Les composés contenant au moins deux lexies, c'est-à-dire des lexies simples (primaires ou affixées) et des lexies composées ainsi que des :
- sigles ou acronymes (EN *cookie ID*, FR *FTP anonyme*, ES *módem ADSL*),
- amalgames (EN *Internet provider*, FR *connexion Internet*, ES *módem satelite*),
- troncations (EN *techno -nerd*, FR *site perso*, ES *punto com*),
- Les composés pseudo-confixés, c'est-à-dire les composés qui contiennent une (des) lexie(s) et un (des) pseudo-confixe(s) (EN *e-mail*, EN/FR *cybercafé*, FR *e-tourisme*, ES *hipervínculo*) ou les composés formés de deux pseudo-confixes (EN *cybernaut*, *webzine*).⁷⁴

3.2.2.2. Les lexies exclues

Toujours selon notre définition du composé, seront écartés de nos catégories :

- Les composés néoclassiques (FR *cardiologue*). Ce type de composition n'est pas fréquent dans le domaine d'Internet. Les éléments *cyber-* et *hyper-* sont, dans la terminologie d'Internet, des pseudo-confixes : des lexies tronquées entrant dans la formation de composés homographes de confixes (Part. I, Ch. 1, Sec. 1.1.). Par exemple, dans ES *jubilnauta*, *-nauta* est la troncation de *internauta* et non le confixe *-nauta*.
- Les dérivés. Contrairement à certains auteurs (Alvar Ezquerra, Catach), nous avons estimé que la dérivation n'est pas un type de composition. En revanche, un terme formé par dérivation peut faire partie d'un composé (EN *archive moderator*, FR *gestionnaire de courrier*, ES *proveedor de servicios*).

3.2.2.3. Échantillon de la structure principale

La structure principale sera construite sur le principe suivant (exemple des composés binaires et surcomposés à trois éléments) :

Tableau 7 : Structure principale de la typologie des patrons compositionnels anglais

⁷⁴ Pour l'étude des pseudo-confixes se reporter aux Ch. 2, 3 et 4 (Sec. 1.2.).

Les noms composés anglais français et espagnols du domaine d'Internet :

Catégorie	Patrons	Composés
1- Composés anglais binaires		
1	N + N	<i>domain name, DNS server</i>
2	A + N	<i>temporary cookie, analog modem</i>
3	PS + N	<i>cyberattack, e-tourism</i>
4	A + PS	<i>freeware, openware</i>
5	in + N	<i>inbox</i>
6	up+ N	<i>uplink</i>
7	N+to+N	<i>business to business</i>
8	N+and+N	<i>click and mortar</i>
...
2- Surcomposés anglais à trois éléments		
1	N + N + N	<i>Application Service Provider</i>
2	A + N+ N	<i>blind carbon copy</i>
...
3- Surcomposés anglais à x éléments		
...

Tableau 8 : Structure principale de la typologie des patrons compositionnels français

⁷⁵ Les composés soulignés sont hybrides.

Composés français binaires		
1- Composés endogènes purs et <u>hybrides</u> ⁷⁵		
Catégorie	Patrons	Composés
1	N + de + N	<i>hébergement de site, fournisseur d'accès</i>
2	N + N	<i>site marchand, adresse e-mail</i>
3	A + N	<i>jeune pousse, haut débit</i>
4	PS + N	<u><i>cyberacheteur, e-métier, Net économie</i></u>
5	PS + A	<u><i>Web mobile, Web invisible</i></u>
6	N + de + PS	<u><i>citoyen du Net, communauté du Web</i></u>
7	N + à + N	<i>accès à distance, Foire aux Questions</i>
9	Flud	<i>e-picerie, site hacké, start-upeur</i>
...
2- Composés exogènes		
8	Emp2	<i>adware, backdoor, web designer</i>
Surcomposés à trois éléments		
1-Surcomposés endogènes purs et hybrides		
1	N + de + N + N	<i>outil de création de site, détecteur de chevaux de Troie</i>
2	N + N + A	<i>copie carbone invisible, adresse IP partagée</i>
3	N + sans + N	<i>connexion sans fil</i>
2-Surcomposés exogènes		
4	Emp3	<i>World Wide Web, Frequently Asked Questions</i>
...

Tableau 9 : Structure principale de la typologie des types compositionnels espagnols

⁷⁵ Les composés soulignés sont hybrides.

Les noms composés anglais français et espagnols du domaine d'Internet :

Composés espagnols binaires		
1-Composés endogènes purs et hybrides		
Catégorie	Patrons	Composés
1	N + de + N	<u>alerta de correo</u> , <u>búzon de mail</u> , <u>proveedor de Internet</u> , <u>software de conexión</u>
2	A + N	<u>recién conectado</u>
3	PS + N	<u>ciberterrorismo</u> , <u>e-turismo</u> , <u>netiqueta</u>
4	PS + A	<u>web musical</u> , <u>web personal</u>
5	V + N	<u>cortafuegos</u>
6	N + a + N	<u>acceso a Internet</u> , <u>negocio a negocio</u>
8	Flud	<u>chequeador de mail</u> , <u>página hackeada</u>
...
2-Composés exogènes		
7	Emp2	<i>bookmark, chatbox, message board</i>
Surcomposés espagnols à trois éléments		
1-Surcomposés endogènes purs et hybrides		
1	N + de + N + de + N	<i>sistema de nombre de dominio</i> , <i>programa de compartido de ficheros</i>
2	N + N + A	<u>conexión ADSL multipuesto</u> , <u>telaraña electrónica</u>
3	N + sobre + N	<u>voz sobre IP</u>
...
2-Surcomposés exogènes		
4	Emp3	<i>Domain Name System, junk e-mail, shadow password</i>

Les tableaux 7, 8 et 9 appellent quelques commentaires :

- Les prépositions, conjonctions et adverbes n'ayant pas les mêmes fonctions, ces éléments seront détaillés et non regroupés dans une catégorie 'prép', 'conj' ou 'adv'. De plus, il peut y avoir ambiguïté. EN *in*, par exemple, est une préposition et un adverbe.
- Les composés contenant un sigle seront inclus dans des catégories de composés avec N car, dans les termes relevés, les sigles se comportent comme des noms. Par exemple, EN *ADSL modem*, FR *modem ADSL*, ES *módem ADSL* sont des composés N + N.
- Les pseudo-confixes (EN *cyberspace*, FR *e-courrier*, ES *metaportal*) seront inclus dans des catégories spécifiques (PS + N pour les exemples ci-dessus). Ils sont des troncations et ne se comportent pas comme des noms⁷⁶ (contrairement aux sigles).
- Les formations ludiques seront regroupées dans une catégorie spécifique (Flud) : EN *Napsterized file*, FR *e-picerie*, *netmarketeur*, *sité hacké*, ES *chequeador de e-mail*. Ces termes ne peuvent pas être supprimés, même s'ils ne sont pas répertoriés dans les ouvrages de référence, puisqu'ils répondent aux critères des composés.

⁷⁶ Seul cyber a acquis un emploi autonome. On peut dire FR *e-cyber* mais pas *le e-cyber*.

- Dans le tableau 7, le double traitement de *modem* et *Internet* peut surprendre. Les formants de *Internet* et *modem* seront distingués uniquement lorsque ces deux lexies sont employées seules (catégories 1 et 2). Ce sont des amalgames formés par troncation de deux noms (*modem*) et par troncation d'un adjectif et d'un nom (*Internet*). Ces unités lexicales ont très vite été assimilées et ne sont plus senties comme des composés, d'où leur non découpage dans les surcomposés. Par exemple, *Internet content provider* est de type N + N + N (et non A + N + N + N) et *satellite modem* est un N + N (et non un N + N + N). Ce choix ne sera pas un obstacle pour le système d'équivalences puisque EN *Internet* et *modem* possèdent déjà des équivalents français et espagnol implantés (qui sont des emprunts).
- Les tableaux 8 et 9 montrent que les composés français et espagnols seront classés selon leur patron syntaxique, tout comme les composés anglais, mais ils seront en plus répartis dans deux catégories : composés endogènes (purs et hybrides) et composés exogènes. La première catégorie sera constituée des composés dont tous les formants sont d'origine française (endogènes purs) ou dont un formant est d'origine anglaise (endogènes hybrides). La deuxième sera consacrée aux composés exogènes, c'est-à-dire aux emprunts à l'anglais. Cette distinction nous permettra, dans la partie analytique, d'étudier l'ensemble des composés des lexiques français et espagnol d'Internet et d'isoler les composés endogènes lorsque cela sera nécessaire pour ne pas brouiller les résultats avec des termes du xénolexique (Ch. 3 et 4, Sec. 1.1.).
- Les emprunts intégraux seront regroupés dans la catégorie 'Composés exogènes' ou 'Surcomposés exogènes'. De cette façon, leur patron ne sera pas détaillé. En effet, les internautes empruntent le signe parce qu'ils ont besoin du signifié. Ils connaissent la nature grammaticale de l'emprunt mais pas celles des formants. Les internautes français savent que FR *adware* est un nom, mais ils ne se soucient pas de savoir que *ad* est une troncation et *ware* un pseudo-confixe (*software*).
- Une catégorie 'Composés endogènes' distinguant les formations hybrides permet d'inclure les composés contenant un emprunt anglais dans les catégories correspondant à la nature grammaticale de l'emprunt. Ils ne seront pas traités différemment des unités lexicales françaises et espagnoles endogènes pures. Par exemple, FR *gestion de cookies* est de type N + de + N, tout comme *page d'accueil*, et ES *banner publicitario* est N + A comme *mensaje cifrado*. Pour inclure un emprunt dans une formation endogène et respecter les règles de formation des composés de la langue, le locuteur doit connaître la nature grammaticale de l'emprunt. Par exemple, dans FR *bloqueur de popup* et *e-mail chiffré* le déterminé précède le déterminant. L'internaute a conscience que *popup* et *e-mail* sont des noms : *popup* est précédé de la préposition *de* et *e-mail* est suivi d'un adjectif. N + de + N et N + A sont parmi les patrons les plus fréquents en français.
- Dans les surcomposés hybrides, les composés anglais empruntés et réutilisés seront considérés comme des lexies simples. Par exemple, le patron syntaxique de FR *bloqueur de pop-up*, *chiffrement d'email*, ES *dirección de email*, *sitio de e-commerce* est N + de + N. Ce sont les signes (signifiant + signifié) *email*, *popup* et *e-commerce*

qui sont empruntés et réutilisés dans la formation de termes hybrides. Tous les formants ne sont pas énumérés. En effet, les internautes empruntent un tout et non plusieurs signifiés indépendants.

3.2.3. Une structure secondaire est-elle profitable ?

Après avoir décidé de la structure principale, nous avons introduit une sous-structure dans nos typologies pour en tester l'utilité. La structure secondaire est découpée selon la nature grammaticale précise des constituants (EN *measurement* = V-*ment*, *definition* = V-*ion*, etc.). Le tableau 10 en est une illustration.

Tableau 10 : Structure secondaire de la typologie des patrons syntaxiques

Cat.	Patrons	Sous-cat	Sous-patrons	Termes
1	N + N	0	N + N	<i>address book, chatroom</i>
		1	N + V-ment	<i>audience measurement</i>
		2	N + V-ion	<i>virus definition, spam protection</i>
		3	V-ion + N	<i>discussion board, navigation bar</i>
		4	N + V-er/or	<i>bug hunter, mail checker</i>
		5	V-age + N	<i>storage site</i>
		6	V-isom + N	<i>comparison site</i>
		7	Am + N	<i>Internet access, Internet portal</i>
		8	Am + V-ion	<i>Internet connection</i>
		9	Am + emprunt	<i>Internet cache</i>
		10	Am + V-er	<i>Internet server</i>
		11	N + Am	<i>Internet modem</i>
		12	N + S	<i>cookie ID, voice IP</i>
		13	N + Nt	<i>dot com, banner ad</i>
		14	S + N	<i>FTP access, POP account</i>
		15	Nt + N	<i>techno-nerd</i>
		16	Nt + V-er	<i>ad-blocker</i>
		17	Nt + Nt	<i>emoticon</i>
		18	V-ort + V-ort	<i>modem</i>
		19	N + V-ing	<i>banner advertizing</i>
		20	V-ing + N	<i>browsing habit</i>
		21	V-ing + V-ion	<i>blocking access</i>
		22	V-ing + S	<i>hosting URL</i>
		23	V-ingt + Nt	<i>shopbot</i>
		24	V-er + N	<i>server attack</i>
		25	V-er + V-er	<i>browser sniffer</i>
		26	V-er + V-ion	<i>server connection</i>
		27	Npr + Am	<i>Java applet</i>
		28	Npr + N	<i>Usenet group</i>
		29	Npr + V-ing	<i>Usenet posting</i>
		30	N-less + V-ion	<i>Wireless connection</i>
		31	S + V-er	<i>ADSL provider</i>
		32	S + Am	<i>ADSL modem</i>
		33	V-er + emp	<i>browser cache</i>
		34	V-ity + N	<i>security hole</i>
2	A + N	0	A + N	<i>blacklist, viral attack</i>
		1	V-able + N	<i>executable attack</i>
		2	A + V-er/or	<i>central server</i>
		3	A + V-ion	<i>viral infection</i>
		4	Apr + N	<i>Trojan horse</i>
		5	A + S	<i>anonymous FTP</i>
		6	A + Am	<i>analog modem</i>

	7	V-n + N	<i>encrypted virus</i>
	8	V-n + V-er	<i>dedicated server</i>
	9	A + V-ing	<i>virtual hosting</i>
	10	V-nt + Nt	<i>Internet</i>
	11	Apr + V-er	<i>Boolean operator</i>
	12	A + V-ure	<i>digital signature</i>
	13	V-n + V-ing	<i>shared hosting</i>
	14	S + V-ing	<i>WAP locking</i>

Ce tableau représente un extrait du travail accompli pour la langue source.⁷⁷ Dans cette structure secondaire :

- les éléments déverbaux (V-ion, V-ment, V-er/or, V-isom, V-ure, V-age) tronqués (*t) ou non, les sigles (S), les amalgames (Am), les emprunts (Emp), les dérivés (N-less) et les noms propres (Npr) sont des sous-types de substantifs (N) puisqu'ils se comportent comme des substantifs.
- les adjectifs ayant la forme d'un participe passé (V-n) tronqué (*t) ou non, ceux correspondant à un nom propre (Apr) et ceux comportant un suffixe (V-able) sont des sous-catégories d'adjectifs (A).

Un test révèle que les catégories générales sont plus que suffisantes. Les sous-catégories sont néfastes pour le fonctionnement du système d'équivalences types car elles entraînent :

- un problème de calcul de la fréquence des types mineurs. Diviser les catégories en sous-catégories augmente considérablement le nombre de patrons compositionnels. Le patron N + N anglais, par exemple, est divisé en 35 sous-types dans le tableau 10. Si le nombre de composés par sous-types est faible – par rapport au nombre de composés par types – alors les chiffres sont peu représentatifs. Cela est d'autant plus vrai dans le cas des surcomposés.
- une trop grande restriction dans le choix des constituants pour construire les équivalents en langue cible. Par exemple, le sous-type du composé anglais *encryption key* est V-ion + N. Le sous-type français correspondant le plus fréquent est N + V-ion. Or, il est impossible de créer un équivalent français calqué qui soit N + V-ion puisque le nom désignant l'action du verbe *chiffrer* porte le suffixe -ment. De plus, cette équivalence est irrégulière puisqu'elle fonctionne pour EN *navigation bar* (FR *barre de navigation*), par exemple. Le sous-type de EN *archive manager* est N + V-er. En français, le sous-type le plus fréquent pour rendre les termes anglais N + V-er est V-eur + de + N. Le terme français correspondant à EN *manager* est *gestionnaire* qui est V-aire. Au sous-type anglais V-ing + N (*indexing engine*), par exemple, correspondent six sous-types français. Le plus fréquent est N + de + V-ion (4 occurrences). Or, on ne peut pas créer un équivalent français N + de + V-ion pour

⁷⁷ Les autres catégories – PS + N, N + PS, A + PS, P + N, N + P, V + N, N + Adv, Adv + V et N + L + N – ont été découpées de la même manière.

plusieurs composés anglais (*filtering rule, hosting company*) et il n'est parfois pas non plus possible de recourir à N + de + V-age, la deuxième équivalence la plus fréquente. En somme, plus on détaille la morphologie des composés anglais, plus on est confronté à l'impossibilité de trouver les éléments morphologiques adéquats en langue cible. Si l'on considère uniquement le type de *encryption key* et de *archive manager* (N + N), on peut créer un équivalent français puisque le patron le plus fréquent pour rendre cette structure anglaise est N + de + N et qu'aucun suffixe n'est imposé : *clé de chiffrement, gestionnaire d'archive*.⁷⁸

Nous excluons l'insertion d'une structure secondaire dans nos typologies car les pourcentages trop faibles des sous-types équivalents auraient peu de valeur, ce qui nuirait à la fiabilité de la base du système d'équivalences. D'autre part, le nombre d'exceptions serait très élevé. Or, le but est d'élaborer un outil utile et utilisable sans trop de difficultés. Le travail décrit dans cette troisième section est à l'origine des typologies anglaises, françaises et espagnoles situées en Annexe (Annexe 8).

Chapitre 2. Les noms composés anglais

La suite de cette partie est consacrée à l'analyse morphosémantique des 1 778 noms composés d'Internet (553 composés anglais, 649 composés français et 576 composés espagnols). Dans la mesure où la terminologie d'Internet est créée en anglais, procéder en premier lieu à l'analyse de la langue source permettra d'évaluer son éventuelle influence sur les deux langues cibles. Cette organisation permet également de consulter les données d'une seule terminologie.

1. Le profil morphosyntaxique des noms composés collectés

⁷⁸ Les suffixes de la langue cible n'étant pas modifiables (l'action du verbe FR *chiffer* est dénommée par *chiffrement* (V-ment) et non **chiffra*tion (V-ion)), une alternative consistait à insérer 'action du verbe x' dans l'équivalence en langue cible au lieu de détailler le suffixe lorsque le patron syntactique anglais contient un déverbal. Par exemple, le patron de EN *hosting company* est V-ing + N. Le patron français serait N + de + 'action de x' N + de + 'action d'héberger' (et plus N + de + V-ion). On pourrait alors créer l'équivalent FR entrep⁷⁹e d'hébergement qui ne serait plus contraire aux équivalences proposées puisque le suffixe n'est plus pris en compte. Une fois le système d'équivalences types informatisé, on pourrait même tenter de le relier à un dictionnaire de la langue cible, par un moyen informatique indéterminé, pour obtenir les équivalents. Reprenons l'exemple N + de + 'action d'héberger'. Le système pourrait rechercher 'action d'héberger' dans un dictionnaire électronique et ainsi fournir l'équivalent. Le problème est que les définitions ne sont pas toujours construites sur le modèle 'action de x' comme c'est le cas pour

1.1. Les structures compositionnelles et leur fréquence « type »
comparaison, chiffrement, navigation et connexion dans le PR 1996 (respectivement) : « le fait d'envisager ensemble [...] », « opération par laquelle on compare [...] ; le fait de naviguer [...] » et « liaison entre deux points [...] ». Cette solution est occultée dans *Morphological Productivity* (2001). Bauer met l'accent sur l'ambiguité de certains termes utilisés pour parler de fréquence et de productivité. Nous nous sommes penchée sur les commentaires de l'auteur sur ces deux notions ayant de procéder à l'analyse des fréquent pour rendre les noms français. Des équivalences créées à partir de action de x ne seraient pas plus détaillées.

⁷⁹ Nous serons vigilante quant à la définition du nom composé adoptée par les auteurs (inclusion/exclusion des affixés, des composés néoclassiques, des amalgames, etc.) et au degré d'affinage de leurs typologies.

typologies.

Pour certains auteurs, *fréquence* et *productivité* sont synonymes. Or, un procédé morphologique peut a) être productif mais ne créer qu'une faible quantité de termes nouveaux ou b) avoir créé de nombreuses lexies mais ne plus être productif. Bauer (ibid. : 48) illustre ses propos avec les exemples du préfixe EN *a-* (*ablaze*, *aflutter*) et du suffixe EN *-ment* (*entwinement*, *underlayment*). Marchand (1969) enregistre 90 mots en *a-* et Barnhart (1990) en relève 94. Cet affixe est donc productif mais pas fréquent puisque peu de mots contiennent le préfixe *a-*. Marchand (1969) tente de trouver des mots en *-ment*, Lenhert (1971) en relève plus de 700 et Barnhart (1973) en note 701. Ce procédé morphologique a une « type frequency » élevée mais n'est pas productif. La fréquence est le résultat d'une productivité passée plutôt qu'une indication sur la productivité actuelle.

La productivité est relative à la quantité de néologismes qui peuvent être créés en utilisant un procédé morphologique précis. Elle oscille entre la disponibilité (« availability ») et la rentabilité (« profitability ») (Bauer, 2001 : 211). La disponibilité d'un procédé morphologique est sa capacité à générer des créations morphologiques. Elle est déterminée par le système linguistique. La rentabilité d'un procédé morphologique reflète l'exploitation de sa disponibilité dans la langue.

Ce qui nous intéresse ici, c'est de connaître le nombre de composés créés par les patrons syntaxiques réalisés et répertoriés dans notre corpus. Nous calculerons la fréquence « type »⁸⁰ des patrons syntaxiques. En effet, nous ne chercherons pas à connaître le nombre de composés qui peuvent être créés par un patron (en fonction de leur disponibilité et rentabilité) mais le nombre de composés effectivement créés par celui-ci, c'est-à-dire sa productivité passée. Nous recenserons le nombre de composés formés par chaque structure compositionnelle. Par exemple, dans le corpus, la fréquence « type » du patron anglais N + N est 201 puisque 201 composés anglais sont ainsi formés. D'autre part, la fréquence « type » est à distinguer de la fréquence « token »⁸¹ qui est le nombre d'occurrences des composés formés par une structure. Reprenons l'exemple du patron N + N, sa fréquence « token » peut atteindre 1 500 ou 2 000 si on prend en compte chaque occurrence des composés N + N dans un ensemble de textes. Dans le présent travail, lorsque les termes *fréquence* et *fréquent* seront employés sans autre précision ce sera toujours pour renvoyer à la fréquence « type ».

1.1.1. Les composés binaires⁸²

⁸⁰ “Type frequency [...] is concerned with the number of items in the language that contain the item or process under consideration” (Bauer, ibid.: 47).

⁸¹ “Token frequency [...] is concerned with the number of times a particular item occurs in a given text [...] that is to say, repetitions of the same word count as separate items [...]” (Bauer, ibid.: 47).

⁸² Abréviations utilisées pour l'analyse des patrons compositionnels : A = adjectif, Adv = adverbe, Emp2 = emprunt de deux éléments, Emp3 = emprunt formé de trois éléments, Emp4 = emprunt formé de quatre éléments, Fdiv = formations diverses, Flud = formations ludiques, N = nom, PS = pseudo-confixe, S = sigle, V = verbe.

Tableau 11 : Structures compositionnelles des composés anglais

	Structures	Fréquence « type »	%
1	N + N	201	50,50%
2	PS + N	96	24,12%
3	A + N	72	18,09%
4	13 structures mineures ⁸³	29	7,29%
	Total	398	100,00%

Globalement, les composés anglais se découpent en seize catégories. Trois d'entre elles représentent 92,71% de l'ensemble. Il s'agit des composés formés de (a) deux substantifs, (b) d'un pseudo-confixe et d'un substantif et (c) d'un adjectif et un substantif. Les treize autres structures ne représentent donc que 7,29% du total.

Ce n'est pas une spécificité du lexique d'Internet que le patron syntaxique le plus fréquent soit N + N, avec 50,50% (type *snail mail*, *virus definition*). En effet, Tournier (1991b : 127) et Jouneau (1997 : 9) ont observé la même chose dans la langue générale. Les N + N représentent 58,94% du corpus de Jouneau et 59,80% de celui de Tournier. Les N + N sont aussi majoritaires dans d'autres études terminologiques : 17 ou 20% de la physico-chimie (variation selon le corpus)⁸⁴ (Depierre, 2001 : 106-108), 70,95% de l'informatique (Béciri, 1999 : 138), 50,3% de la science fiction (Gindre, 1998 : 37). 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 (type *cyberspace*, *Web access*) est non seulement élevée (24,12%) mais elle est également inhabituelle. Au total, les patrons compositionnels avec pseudo-confixes (PS + N, N + PS et A + PS) représentent 26,14 % des composés. La création de termes à partir de pseudo-confixes semble être caractéristique du domaine d'Internet.⁸⁵ Nous n'avons pas relevé de travaux sur les lexies de la langue générale ou sur les vocabulaires spécialisés faisant état de pseudo-confixes. Dans sa récente étude du vocabulaire informatique, un domaine connexe, Béciri (1999) ne mentionne aucune création de ce genre.

Les composés A + N (type *intelligent agent*, *secure server*) – au troisième rang de

⁸³ 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%).

⁸⁴ Dans son corpus dictionnaire (*Penguin Dictionary of Physics*, 1990), Depierre recense 17% de composés N + N au total et 26% de N+N chez les composés binaires uniquement. Le chiffre pour les binaires atteint 55% lorsque l'auteur regroupe les pourcentages des catégories suivantes : N + N (26%), N propre + N (15%), morphème savant + N (8%), morphème savant + morphème savant (4%), lettre + N (2%), sigle + N (- de 1%). Dans son corpus de catalogues publicitaires, Depierre relève 15 % de composés N + N. On obtient 22,70% de noms composés chez les binaires si on additionne les catégories suivantes : N + N, V-ing + N (4,5%), N + V-er (3,20%). Dans son corpus bilingue (*Do you speak Chemistry ?*, 1991), Depierre relève 20% de N + N.

⁸⁵ Les composés pseudo-confixés sont étudiés en détail dans la section 1.2.

notre 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 sont majoritaires avec 18% du total des composés (27% des composés binaires)⁸⁶ (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 plus élevée sans la présence des pseudo-confixes. Par exemple, la forme e- qui produit de nombreux termes (*email*, *ebook*, *e-card*, etc.) est la troncation postérieure de l'adjectif *electronic*. Le nombre de composés A + N se voit ainsi réduit.

Les noms composés comprenant un verbe sont peu fréquents dans notre corpus. Ils font partie des quatorze types mineurs. Et les deux types rencontrés *down + V* et *up + V* (*download*, *upload*, *uploading*) ont produit seulement 0,75% des termes du corpus. La présence des verbes est plus marquée dans la langue générale. Tournier (1985 : 114-115) a recensé les types suivants : N + V (*toothpick*), A + V (*whitewash*), Adv + V (*upstart*), V + N (*pickpocket*), V + part. adv. (*stand-by*), 2 V coordonnés (*hide-and-seek*), V + V (*make-believe*). 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, 10% des surcomposés et 14% des composés binaires (Depierre, 2001 : 103). Pour la langue générale, Brocard (1998 : 21) et Paillard (2000 : 51) mentionnent l'existence des V + N mais ne donnent aucun chiffre. 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. La faible fréquence des composés comprenant un verbe dans notre corpus s'explique, en partie, par la différence de structure des typologies. Contrairement à Jouneau, nous excluons les éléments déverbaux (V-er, V-or, V-ion, V-ing, V-n) qui sont pour nous des sous-types de N ou de A. Ils sont donc comptabilisés dans les types comprenant un N ou un A comme N + N, A + N, PS + N, etc.

D'autre part, nous constatons la très faible représentation des structures N + to + N (1,01%), N + and + N (0,25%), N + of + N (0,76%) (type *business to business*, *Denial of Service*, *click and mortar*, *peer-to-peer*, *business-to-consumer*) par rapport à la langue générale. Tournier (1985 : 118) relève 1 145 lexies de ce type (dont 84,03% de N + of + N) dans la sixième édition de *The Concise Oxford Dictionary of Current English* (1976). Selon l'auteur, cette structure est généralement utilisée dans les registres très soutenus, les langues spécialisées comme les terminologies officielles, les vocabulaires juridique, religieux ou littéraire. Par conséquent, ce serait la nature même du lexique d'Internet, bien différente de celle des domaines cités par Tournier, qui explique la faible fréquence de ce type de composition.

1.1.2. Les surcomposés à trois éléments

⁸⁶ Ces chiffres correspondent au corpus dictionnaire de l'auteur. Si on regroupe les catégories contentant 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%).

Tableau 12 : Structures compositionnelles des surcomposés anglais à trois éléments

	Structures	Fréquence « type »	%
1	N + N + N	34	26,98%
2	A + N + N	26	20,63%
3	PS + N + N	25	19,84%
4	17 structures mineures ⁸⁷	41	32,54%
	Total	126	100,00%

Les cent vingt-six surcomposés sont créés par vingt structures différentes. Avec 67,45% du total, trois d'entre elles sont plus fréquentes que les autres : les surcomposés de trois substantifs N + N + N (type *Domain Name System, ad-blocking program*), d'un pseudo-confixe et de deux substantifs PS + N + N (type *Net connection speed, email virus*) et d'un adjetif et deux substantifs A+ N + N (type *electronic mail account, shared IP address*). Les dix-sept autres patrons se partagent les 32,55% restants.

Les structures arrivant en tête ressemblent à celles de la typologie des composés binaires. Cependant, leurs fréquences diffèrent de celles des binaires. En effet, la fréquence du patron N + N est beaucoup plus élevée (50,50 %) que celle du patron N + N + N (26,98 %). La différence est plus faible pour les PS + N (24,12 %) et les PS + N + N (19,84 %). La tendance s'inverse dans le cas des composés et surcomposés construits avec un adjetif : les A + N + N (20,16 %) sont plus fréquents que les A + N (17,88 %). Que peut-on tirer de ces faits ? Les A + N + N sont, en principe, des N + N auxquels on a ajouté une restriction sémantique : *newsletter* et [electronic [newsletter]], *home page* et [personal [home page]], *IP address* et [shared [IP address]], *email* et [encrypted [email]], etc. Pour dénommer un nouveau concept d'après un concept existant, il est possible d'ajouter une restriction sémantique au terme désignant le concept de départ. Lorsqu'un composé dénomme le concept de départ, pour la compréhension du nouveau concept, peut-être est-il préférable de créer un surcomposé avec un adjetif plutôt qu'avec un troisième substantif (le découpage d'un surcomposé formé de plusieurs noms est complexe).

La prédilection des internautes pour les pseudo-confixés se confirme. Une part importante des surcomposés à trois éléments contient un pseudo-confixe. PS + A + N (type *cybersupermarket*), PS + N + N (type *e-commerce site, email provider*), A + PS + N (type *static Web page, encrypted email*) et N + PS + N (type *voice email*) représentent plus d'un quart du total (26,98%) et deviennent ainsi les surcomposés les plus nombreux.

Les surcomposés comprenant un élément de liaison, N + to + N + N, N + of + N + N et N + and + N + N (type *Denial of Service attack, click-and-mortar business*) sont peu présents (4,75 %) tout comme les N + to + N, N + of + N et N + and + N (2,01 %).

⁸⁷ Il s'agit de : A+A+N (0,78%), on+N+N (3,10%), off+N+N (1,55%), under+N+N (0,78%), N+A+N (1,55%), N+up+N (4,66%), N+down+N (0,78%), N+under+N (0,78%), PS+A+N (1,55%), A+PS+N (3,88%), Adv+V+N (3,10%), Adv+A+N (0,78%), N+PS+N (0,78%), Adv+on+N (1,55%), N+of+N+N (0,78%), N+to+N+N (3,10%), N+and+N+N (0,78%).

Les données concernant d'autres vocabulaires sont peu abondantes. D'après Béciri (1997 : 138), les surcomposés représentent 17% du total des composés en informatique et 11,50% de la langue générale. Elle ne procède pas à une analyse détaillée. En physico-chimie, un tiers des composés comportent plus de deux éléments (Depierre, 2001 : 171). Selon Brocard (1998 : 22), les surcomposés à trois éléments représentent une part assez grande de la terminologie du cinéma. Seulement 0,02% des listes de Jouneau (1997 : 10) est constitué de surcomposés. Tournier (1991b : 171-172) décrit l'ordre des éléments mais ne donne pas de chiffres. Gindre (1998) n'évoque pas le sujet.

1.1.3. Les surcomposés à quatre, cinq et six éléments

Tableau 13 : Structures compositionnelles des surcomposés anglais à quatre éléments

	Structures	Fréquence « type »	%
1	A + N + N + N	4	19,05%
2	N + N + A + N	2	9,52%
	A + A + N + N	2	9,52%
	PS + N + N + N	2	9,52%
3	N + over + N + N	1	4,76%
	N + N + N + up	1	4,76%
	on + N + N + N	1	4,76%
	in + N + N + N	1	4,76%
	out + N + N + N	1	4,76%
	Adv + on + A + N	1	4,76%
	N + A + A + N	1	4,76%
	A + N + PS + N	1	4,76%
	N + PS + N + N	1	4,76%
	PS + N + A + N	1	4,76%
	A + N + up + N	1	4,76%
	Total	21	100,00%

Tableau 14 : Structures compositionnelles des surcomposés anglais à cinq et six éléments

	Structures	Fréquence « type »	%
Surcomposés à cinq éléments			
1	N + N + N + A + N	2	33,33%
2	Adv + on + A + N + N	1	16,67%
	N + N + N + N + N	1	16,67%
	PS + N + N + up + N	1	16,67%
	N + N + V + N + N	1	16,67%
	Total	6	100,00%
Surcomposés à six éléments			
	A+ PS + N + N + up + N	1	100,00%
	Total	1	100,00%

La première observation est le faible nombre de surcomposés créés à partir des structures des tableaux 13 et 14. Ce nombre peu élevé est-il une particularité de la terminologie d'Internet ? On constate ensuite que sur les quinze procédés du tableau 6 aucun ne se détache. Cela est-il lié au manque d'occurrences ?

Les pourcentages de ces tableaux n'ont pas de signification. Il est impossible de tirer des conclusions acceptables même en se basant sur Tournier qui précise que les termes longs – quatre éléments ou plus – sont rares en langue générale (1991b : 171). Ce fait de la langue générale est peut-être partagé par d'autres domaines. Cependant, il n'est pas étonnant que les termes très longs soient peu représentés dans notre corpus. Les termes les plus longs sont techniques, or le côté technique d'Internet n'est pas le thème majeur des revues (*l'actualité*). De plus, ils sont généralement rapidement raccourcis (réductions, sigles) pour faciliter la communication.

En ce qui concerne les surcomposés à cinq et six éléments, on dénombre six patrons seulement. Comme pour les surcomposés précédents, les patrons sont très peu fréquents (un terme par type). Un corpus plus étendu ne fournirait sans doute pas beaucoup plus de termes. En revanche, le nombre de structures pourrait augmenter puisqu'il est fonction de la longueur des surcomposés. Plus un surcomposé est long, plus il a de chances de contenir des constituants de nature grammaticale différente.

Plus en détail, les quinze patrons du tableau 13 donnent vingt et un termes surcomposés. Contrairement aux composés et surcomposés à trois éléments, aucune structure n'est plus fréquente que les autres. Dans notre corpus, aucune n'est très fréquente. Les études qui nous ont permis d'effectuer les comparaisons précédentes sont inutiles ici (seule Depierre calcule qu'environ 1% des composés de la physico-chimie ont quatre éléments ou plus (2001 : 175)). Les autres études n'abordent pas les surcomposés de plus de trois éléments et certaines ne mentionnent même pas leur existence (Gindre, Brocard, Béciri, Paillard, Jouneau). Cela tend à confirmer que la faible productivité de ces procédés n'est pas une particularité d'Internet.

Cinq surcomposés à quatre éléments contiennent un pseudo-confixe (23,08 % du total). Le surcomposé à six éléments ainsi qu'un surcomposé à cinq éléments sont pseudo-confixés. La « pseudo-confixation » est un type de composition fréquent. Cela renforce la tendance déjà dégagée – mais comme dit précédemment la prudence

Les noms composés anglais français et espagnols du domaine d'Internet :

s'impose.

1.1.4. Bilan

Tableau 15 : Structures compositionnelles des composés et surcomposés anglais

	Structures	Fréquence « type »	%
	6 Structures majeures :	454	82,10%
1	N + N	201	36,35%
2	PS + N	96	17,36%
3	A + N	72	13,02%
4	N + N + N	34	6,15%
5	A + N + N	26	4,70%
6	PS + N + N	25	4,52%
	52 structures mineures :	99	17,82%
7	N + up + N	5	0,90%
	A + PS + N	5	0,90%
8	N + PS	4	0,72%
	A + PS	4	0,72%
	N + to + N	4	0,72%
	Fdiv2	4	0,72%
	on + N + N	4	0,72%
	Adv + V + N	4	0,72%
	N + to + N + N	4	0,72%
	A + N + N + N	4	0,72%
9	N + of + N	3	0,54%
	N + A + N	3	0,54%
	PS + A + N	3	
10	N + up	2	0,36%
	down + V	2	
	off + N + N	2	
	Adv + on + N	2	
	N + N + A + N	2	
	A + A + N + N	2	
	PS + N + N + N	2	
	N + N + N + A + N	2	
11	in + N	1	0,18%
	up + N	1	
	N+ to	1	
	N + over	1	
	up + V	1	
	N + and + N	1	
	A + A + N	1	
	under + N + N	1	
	N + down + N	1	
	N + under + N	1	
	N + up + N	1	
	Adv + A + N	1	
	N+ PS + N	1	
	N + of + N + N	1	

N + and + N + N	1	
N + over + N + N	1	
N + N + N + up	1	
on + N + N + N	1	
in + N + N + N	1	
out + N + N + N	1	
Adv + on + A + N	1	
N + A + A + N	1	
A + N + PS + N	1	
N + PS + N + N	1	
PS + N + A + N	1	
A + N + up + N	1	
Adv + on + A + N + N	1	
N + N + N + N + N	1	
PS + N + N + up + N	1	
N + N + V + N + N	1	
A+ PS + N + N + up + N	1	
Total	553	100,00%

En résumé, le corpus anglais compte 58 patrons compositionnels tels que définis. Le pourcentage moyen de composés par patron est 1,72%. Si nous considérons le nombre total de composés et surcomposés, il apparaît que :

- Les composés les plus nombreux, N + N (36,35 %), PS + N (17,36%), A + N (13,02%), N + N + N (6,15 %), A + N + N (4,70 %), PS + N + N (4,52%) ont produit 82,10% des termes. Les 52 autres patrons ont donc généré 17,89 % des composés seulement.
- La longueur des surcomposés a un lien avec leur fréquence :

Nombre d'éléments des composés	Fréquence « type »	%
2	398	71,97
3	126	22,78
4	21	3,80
5, 6	8	1,45
Total	553	100

Outre la fonction de communication – qui favorise la brièveté des termes –, une explication de la baisse de fréquence des termes longs peut être la siglaison. Les surcomposés donnent plus d'indications sur le concept que les termes courts mais l'économie linguistique fait qu'ils sont abrégés : *FAQ* est employé à la place de *Frequently Asked Questions*, *ASP* est employé pour *Application Service Provider*, *XML* pour *eXtensible Markup Language*, *SMTP* pour *Simple Mail Transfer Protocol* ou *TCP/IP* pour *Transmission Control Protocol/Internet Protocol*, etc.

Les pseudo-confixes ont une fréquence élevée surtout chez les composés (104 occurrences). Toutes catégories confondues, nous en avons dénombré 153, soit 27,67% du corpus anglais.

1.2. Les composés pseudo-confixés

Les premiers termes pseudo-confixés du domaine d'Internet ont été créés en anglais (*cyberspace*, *e-mail*, *hypertext*, *hyperlink*). Cette étude a pour but de mettre en lumière la signification de chaque pseudo-confixe du corpus trilingue et de calculer leur fréquence. Les travaux de Mopoho (1998), Raus (2001) et Brolles (2001) serviront de base pour atteindre ce double objectif :

- Nous vérifierons si les propriétés sémantiques (nouveau signifié pour les éléments dont le signifiant existait déjà) et morphosyntaxiques (position préfixale/suffixale, soudure des éléments, etc.) des pseudo-confixes français du lexique d'Internet dégagées par Mopoho s'appliquent aux termes anglais, français et espagnols de notre corpus.
- Nous reprendrons les remarques diachroniques de Raus concernant *cyber-* et *hyper-* dans le lexique français d'Internet et nous procéderons à l'analyse diachronique des autres pseudo-confixes du corpus.
- Nous chercherons à confirmer ou infirmer les remarques de Brolles (2001) concernant l'utilisation des pseudo-confixes dans le lexique d'Internet.

En français, 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 paraissent contestables. L'auteur démontre que FR *micro-* est l'apocope de *micro-ordinateur* ou *micro-informatique* dans *microédition*, que *macro-* est l'apocope de *macrocommande* dans *macro-instruction*. En quoi *anti-* et *nano-*, par exemple, 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' ?⁸⁸ Mopoho ne donne aucune explication, ni aucun exemple avec *nano-* (idem dans son tableau des équivalents français-anglais). Nous inclurons 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-*).

Compte tenu des faits ci-dessus, notre corpus anglais contient sept pseudo-confixes.

⁸⁸ 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 *antifièvre* ou *antibactérien*). Dans ces deux domaines, *anti-* exprime une action bienfaisante et sa fréquence est élevée.

Nous commencerons par étudier *cyber*-, *hyper*-, *meta*-, les pseudo-confixes antérieurs, homographes de confixes grecs, puis *e*-, *net*-, *web*-, *-ware*, les pseudo-confixes antérieurs et postérieurs, troncations de lexies anglaises.

1.2.1. **cyber**-

D'après une étude diachronique de *cyber*- (Raus 2001), ce pseudo-confixe a été utilisé pour la première fois en 1834 dans le néologisme français *cybernétique* signifiant ‘science du gouvernement’. Au XX^e siècle, ce néologisme est introduit aux États-Unis où il acquiert un nouveau sens, celui de ‘branche de la robotique relative aux théories de la communication et à la régulation de l’être vivant et de la machine’ (*ibid.* : 74). En 1984, le romancier américain Gibson crée le néologisme *cyberspace* après *cyber*- employé en robotique (*ibid.* : 75). Le pseudo-confixe retient de son emploi dans *cybernetics* le sens ‘d'espace d'interaction entre l'homme et la machine’. *Cyberspace* renvoie au « monde des ordinateurs connectés et la société fonctionnant autour » (Newton, 2000 :183). Cette dernière acception a fait le succès du mot et du pseudo-confixe. Gibson a produit la lexie *cyberpunk* qui caractérise les sociétés « hautement technicisées, déshumanisées et déclinantes, peuplées de personnages anticonformistes utilisant l'informatique pour mener de véritables guerres de rébellion » (Colombain, 1998 : 64). Mais selon Newton (2000 : 183), la lexie *cyberpunk* a été créée par Katie Hafner et John Markoff dans le livre *Cyberpunk : Outlaws and Hackers on the Computer Frontier* publié en 1992. Newton semble faire référence au sens utilisé dans le cadre d'Internet puisque le livre d'Hafner et Markoff présente le *cyberpunk* comme un hacker (une personne qui pénètre dans les systèmes informatiques d'autrui).

Depuis la création de *cyberspace*, le pseudo-confixe *cyber*- a pris le sens de ‘réalité virtuelle’. Dans le domaine d'Internet, le sens de *cyber*- évolue encore. Une activité du monde réel est possible grâce à ce nouveau support média qu'est Internet (seule exception : EN *cybercafé* qui désigne un lieu réel).

Tableau 16 : Composés anglais du corpus avec le pseudo-confixe *cyber*-.⁸⁹

⁸⁹ Les termes des tableaux écrits en majuscules **et** caractères gras sont les plus fréquents (nombre d'occurrences supérieur à 5 000 000 sur Google.com au 14-01-04). Les termes en majuscules sont très fréquents (entre 1 000 000 et 4 999 999 occurrences), ceux en caractères gras sont fréquents (entre 100 000 et 999 999) et ceux en italique ont un nombre d'occurrences compris entre 10 000 et 99 999. Les trois orthographies (en un bloc, en deux mots ou reliés par un trait d'union) sont comptabilisées.

Pseudo-confixe	Base	Champ sémantique	Base	Champ sémantique
cyber	activity	Sé	merchant	C
	album	Sé	monitoring	S
	attack	S	<i>protection</i>	S
	brand	C	punk	S
	café	Sé	recruitment	Sé
	conference	Sé	referendum	Sé
	conflict	Sé	sex	Sé
	<i>cop</i>	S	<i>shop</i>	C
	crime	S	soldier	S
	criminal	S	SPACE	Sé
	criminality	S	<i>squatter</i>	S
	debate	Sé	<i>squatting</i>	S
	delinquency	S	STRESS	Sé
	delinquent	S	supermarket	C
	directory	N/A	<i>terrorism</i>	S
	editor	Sé	<i>terrorist</i>	S
	ethics	Sé	univers	Sé
	flirt	Sé	visitor	N
	frontier	Sé	world	Sé
	grocer	C		
	hacking	S		
	language	Sé		
	law	S		
	<i>love</i>	Sé		
	market	C		

Légende : C = commerce, N = navigation, S = sécurité relative au cyberspace, Sé = société

Quarante-quatre composés du corpus anglais contiennent *cyber-*. En ce qui concerne leur morphosyntaxe, le tableau 16 montre que *cyber-* a toujours une position préfixale et que l'ordre des éléments des composés est toujours déterminant/déterminé. Les composés avec *cyber-* respectent donc les règles de construction des composés anglais.

Sémantiquement, la réduction *cyber-* sert principalement à modifier le sens traditionnel de la lexie à laquelle il est accolé pour que l'assemblage s'applique au Réseau. Tout comme un *address book* est un type de *book*, un *cyberalbum* est un *album* ; ces deux composés sont endocentriques. Des termes comme *cyber-crime*, *cyber-grocer* montrent bien que *cyber-* signifie ‘relatif à Internet’. Le domaine d’Internet s’est approprié ce formant qui a un sens autre que celui du suffixe *cyber-* (voir *supra* pp. 156-157). Les ouvrages lexicographiques attestent de la nouvelle vie de *cyber-*. Dans le *Merriam Webster Online Dictionary* (MWOD) et le *Cambridge Advanced Learner’s Dictionary online* (CALDO), *cyber-* est défini de la manière suivante :

MWOD : **cyber-**

Function: *combining form*

Etymology: *cybernetic:*

computer : computer network <*cyberspace*>

CALDO : **cyber-**

Prefix

involving, using or relating to computers, especially the Internet:
cybercrime, cyberculture

À la lecture de ces définitions deux remarques s'imposent. Le MWOD définit *cyber-* comme une « combining form » et le CALDO comme un préfixe. Les lexicographes anglophones utilisent le terme *combining form* pour désigner des éléments du lexique qui se combinent avec d'autres (Tournier, 1991b : 151-152). Ils diffèrent des affixes car, contrairement à ces derniers, ils se combinent généralement entre eux et peuvent se placer à droite ou à gauche. Ne figure dans les deux ouvrages que l'acception informatique/Internet.

Le tableau 16 montre que les termes construits avec *cyber-* se regroupent en quatre thèmes (ordre décroissant de fréquence) :

- la société. La structure *cyber* + N désigne les actes de la vie courante possibles sur le Réseau comme les divertissements offerts par le Net (*cyber-album*, etc.) et ce qui a trait aux 'relations amoureuses' (*cybersex, cyber-flirt, cyberlove*, etc.). Elle sert aussi à désigner des lieux du Web ou des lieux réels : le *cyberspace* (*cyberworld, cyberunivers*) est l'univers Internet dans lequel on se promène tout en étant assis sur son siège ; un *cybercafé* est un café public avec des ordinateurs qui permettent aux clients de surfer sur le Net.
- la sécurité. Les termes désignant les agissements des pirates contiennent souvent *cyber-*. Par exemple, le *cyber-terrorism* qualifie les agressions des serveurs perpétrées par des pirates et un *cybercrime* est un acte illicite perpétré par le moyen de l'informatique, ou ayant pour cible le système informatique ou l'un de ses éléments. La *cyber-delinquence* est la délinquance qui se pratique au moyen de réseaux informatiques. *Cyber-* forme aussi des termes permettant de combattre les nombreuses fraudes possibles en utilisant Internet (vols de numéros de cartes bancaires, casinos illégaux, etc.) comme la *cyber-protection*.
- le commerce. La structure *cyber* + N génère de nombreux termes relatifs à l'activité commerciale sur Internet. Par exemple, un *cybermarket* et un *cybershop* sont des magasins du Net, ouverts 24 heures sur 24, qui permettent aux cyberclients de faire des achats depuis leur domicile.
- la navigation. Quelques lexies comme *cyber-directory* et *cyber-visitor* ont trait au thème de la navigation.

Notre classification confirme l'observation de Brolles (2001) : *cyber-* fait référence à des activités, des choses et des personnes existant dans le monde réel et reproduites sur Internet. Colombain (1998 : 60) va même jusqu'à dire que *cyber-* « s'utilise à toutes les sauces ». Aucun des termes du tableau 16 ne désigne un intervenant direct du Réseau ou un utilisateur (excepté *cybernaut*, terme absent de notre corpus, dont la fréquence est bien inférieure à celle de *internaut*). Certains termes comme *cyberpunk*, *cybersquatter*, *cybermerchant*, *cyberterrorist*, *cybervisitor*, *cyberdelinquent* désignent des personnes qui utilisent le Net à des fins plus ou moins légales et bien intentionnées mais aucun de ces personnages ne participe au développement du Net (contrairement à un *webdesigner*). Il faut également noter que *cyber-* a une telle fréquence que certaines lexies resteront des hapax ou quasi-hapax : *cyberrestaurant*, *cybersandwicherie* (lexies exclues de notre corpus car non attestées).

1.2.2. *hyper-*

En 1960, Ted Nelson a introduit le terme *hypertext*, créé à partir du grec *hyper-* et du substantif anglais *text*, pour faire référence à sa vision de l'information représentée et accessible à partir de liens actifs intégrés dans des documents. Colombain (1998 : 101) décrit l'hypertexte comme un « procédé permettant de consulter un document de manière non linéaire en cliquant sur des images ou des mots clés, appelés liens hypertextes (ou hyperlien), menant à d'autres documents ». Newton (2000 : 404) précise que cette technologie n'est pas utilisée que dans le Web mais aussi dans les CD-Rom interactifs. La définition de Colombain montre bien que le sens de *hyper-* est très différent de celui présenté dans le MWOD :

Main Entry: **hyper**
Function: *adj*
Etymology: Middle English *iper*, from Latin *hyper*, from Greek, from *hyper* – more at OVER
1 : above; beyond; **SUPER** <*hyperspace*>
2 a : excessively <*hypersensitive*> b : excessive <*hyperemia*>
3 : that is or exists in a space of more than three dimensions <*hypercube*>

La valeur intensive a disparu. L'acception d'«au-delà» n'est plus prise dans le sens de dépassement d'une mesure. Un *hypertext* est un renvoi obtenu par un *link* à une autre zone du *text* dans le *cyberspace*. Par métaphore, *hypertext* désigne aussi le texte qui possède des *hyperlinks*.

Tableau 17 : Composés anglais du corpus avec le pseudo-confixe *hyper*

Pseudo-confixe	Base	Champ sémantique
hyper	TEXT	P
	LINK	P
Surcomposés avec : <i>hypertext</i>	Dynamic HyperText Markup Language	P
	HyperText Markup Language	P
	HyperText Transfer Protocol	P
hyperlink	hyperlink video	P
Légende : P = procédé		

Hyper- a une fonction préfixale. Dans le composé, il détermine la lexie tête. Sur le plan sémantique, ces composés sont endocentriques : un *hypertext* est un *text* et un *hyperlink* est un *link*. Tout comme *cyber-*, *hyper-* permet de spécifier le champ d'application de l'élément déterminé. Raus (2001 : 84) fait deux remarques :

- *hyper-* n'est plus antonyme de *hypo-* puisqu'il n'a plus une valeur intensive,
- *hyper-* a une connotation positive. Les termes créés avec *hyper-* ont une valeur positive rattachée au cyberspace, un espace en puissance.

Hyper- semble avoir une double casquette. Dans *hypertext*, l'acception d' 'au-delà' est prise dans le sens de glissement spatial vers un au-delà. Et il semblerait que dans *hyperlink*, par exemple, *hyper-* soit l'apocope d'*hypertext*. Le succès du procédé technologique fait que le patron syntaxique est utilisé pour créer d'autres néologismes. *Hyper* + déterminé devient fréquent grâce à la connotation positive attribuée à l'hypertexte. *Hyper-* sert à désigner un procédé, un moyen de navigation et tout ce qui s'y rapporte. En témoignent les autres lexies relatives à l'hypertexte créées avec *hyper-* citées dans le NND (2000), *hypermedia*, ⁹⁰ *hypertext document*, *hypertext link*, *hypertextual*, *hyperword*, ou dans d'autres glossaires *hyperdocument*, *hyperimage*.

1.2.3. meta-

Meta-, absent du CALDO, est un préfixe d'étymologie classique pour le MWOD. Il signifie :

⁹⁰ « Mode hypertexte étendu qui permet à l'utilisateur de naviguer sans parcours préétabli à travers les données textuelles, sonores ou visuelles, les images animées ou les séquences vidéo » (Newton, 2000 : 404). *Hypermedia* = *hyper-* (apocope d'*hypertext*) > l'hypermédia est un mode hypertexte) + *media* (mot hyperonyme désignant la nature des données).

Main Entry: meta-
 Variant(s): or met
 Function: prefix
 [...] 1 a : occurring later than or in succession to : after <*metameric*> b : situated behind or beyond <*metencephalon*> <*metacampus*> c . later or more highly organized or specialized form of <*metaxylem*>
 2 : change : transformation
 3 [*metaphysics*] : more comprehensive : transcending <*metapsychology*> -- used with the name of a discipline to designate a new but related discipline designed to deal critically with the original one <*metamathematics*>
 4 a : involving substitution at or characterized by two positions in the benzene ring that are separated by one carbon atom <*meta-xylene*> b : derived from by loss of water <*metaphosphoric acid*>

Tableau 18 : Composés anglais du corpus avec le pseudo-prefixe meta-

Tableau 18 : Composés anglais du corpus avec le pseudo-prefixe meta-

Pseudo-suffixe	Base	Champ sémantique
meta	tag	C
	portal	A
	search engine	A

Légende : A = application, C= commande

Meta-, qui a une fonction préfixale, retrouve une nouvelle vie grâce à Internet. En effet, *meta-* est un pseudo-suffixe et non un suffixe car il ne correspond plus aux acceptations de la définition du MWOD. Un *meta tag* est un marqueur HTML, utilisé dans un en-tête de page Web, qui encadre des informations servant de mots-clés pour les outils de recherche (Newton, 2000 : 499). Dans *meta tag*, *meta-* est probablement la troncation de *metaelement* qui est l'information ajoutée dans l'en-tête de la page Web au moyen des marqueurs HTML. Dans *metaelement*, *meta-* est certainement la réduction de *metalanguage* puisque le langage HTML est un métalangage. C'est donc dans cette dernière lexie, non retenue dans notre corpus, que *meta-* a le sens ancien de « langage qui permet de définir un autre langage ».⁹¹

Un *meta search engine* est un logiciel permettant à un internaute de lancer une requête dans plusieurs moteurs de recherche en même temps et un *metaportal* est un portail qui interroge d'autres portails à la place de l'utilisateur et synthétise les résultats en utilisant les *meta tags*. Dans les dénominations *metaportal* et *meta search engine*, *meta-* fait référence au principe de fonctionnement de ces outils (les *meta tags*). Newton recense d'autres termes dans lesquels *meta-* a une signification identique : *meta data*, *meta directory*, *meta element* et *meta search*.

1.2.4. e-

⁹¹ a metalanguage is a language that describes other languages (<http://www.netlingo.com/inframes.cfm>)

Le pseudo-confixe *e-* est né de la troncation postérieure de *electronic*. Il a été utilisé pour la première fois en 1971 dans *e-mail* (Newton, 2000 : 326). Depuis l'avènement d'Internet, le nouveau formant *e-* est devenu très productif, à tel point qu'il possède sa propre entrée dans le MWODet dans le CALDO :

MWOD : Main Entry: *e-*
 Function: *combining form*
 Etymology: *e-mail*
 : electronic <*e-commerce*>

CALDO : *e-*
 Prefix
 ABBREVIATION FOR electronic (ELECTRICAL):
e-commerce, email

Tableau 19 : Composés anglais du corpus avec le pseudo-confixe *e-*

Tableau 19 : Composés anglais du corpus avec le pseudo-confixe *e-*

Pseudo-confixe	Base	Champ sémantique	Base	Champ sémantique
<i>e-</i>	BIZ	C	ACCOUNT	M
	BOOK	C	ADDRESS	M
	BUSINESS	C	ALERT	M
	CARD	C	attachment	M
	COMMERCE	C	client	A/C
	<i>customer</i>	C	encryptor	A/S
	<i>employee</i>	T	filtering	S
	<i>job</i>	T	<i>header</i>	M
	LEARNING	T	LIST	M
	MAIL	M	<i>manager</i>	A
	market place	C	NEWSLETTER	M
	<i>purchase</i>	C	provider	M/C
	tourism	C	SERVER	M
	trade	C	software	A
	<i>trader</i>	C	virus	S
	ZINE	Sé	worm	S
<i>Surcomposés avec :</i>			<i>encrypted</i>	S
<i>e-business</i>	integrator	A	junk	S
<i>e-commerce</i>	site	C	POP3 email box	M
	<i>package</i>	A	voice	M
<i>email</i>	bulk	S		

Légende : A = application, C = commerce, M = courrier, S = sécurité, T = travail

Le nouveau formant e-, très présent dans notre corpus anglais, est utilisé dans la formation de 16 composés et de 24 surcomposés (principalement à partir de *e-mail*), soit un total de 40 termes. Il a une fonction préfixale. L'ordre des éléments des composés respecte la règle de formation des composés anglais. Dans les surcomposés, l'élément qui contient le pseudo-confixe est habituellement en position de déterminant (21 cas sur 24). *E-mail virus*, *email software* sont des types de *virus* et de *software*.

En ce qui concerne sa signification, contrairement à *cyber-* et *meta-*, *e-* conserve le sens de sa forme développée *electronic* et pourrait même être remplacé par *electronic* en discours. *E-zine*, *e-book* ou *e-learning* montrent que, dans la terminologie d'Internet, *e-* et *electronic* signifient ‘consultable sur le réseau Internet’ ou ‘réalisable sur le réseau Internet sans déplacement’.

Certains des thèmes relevés dans le corpus sont identiques à ceux relevés pour *cyber-* (les applications, le commerce, la sécurité), mais on trouve d'autres thèmes : le courrier, le travail. Les thèmes de ces deux pseudo-prefixes sont identiques. Comme dans le corpus de Brolles (2001), *e-* est utilisé pour construire des termes qui désignent des occupations, des pratiques et des acteurs du monde réel reproduits sur le Net sans que des connaissances informatiques soient nécessaires. Nous remarquons toutefois une différence entre l'emploi de *e-* dans les composés et dans les surcomposés. Ces derniers sont plus spécifiques et font très souvent référence à des acteurs (logiciels ou humains) qui accomplissent des tâches à la place de l'utilisateur : *email encryptor*, *email manager*, *email provider*, *e-business integrator*.

1.2.5. net-

Selon Mopoho (1998 : 553), le pseudo-confixen⁹²et est la troncation antérieure de l'amalgame *Internet*, apparu pour la première fois en 1982.

Comment le MWOD et le CALDO définissent-ils *net-* ? Dans le MWOD, *net-* n'est pas un mot-vénette. Il est inclus dans l'entrée « *Internet* ». Le MWOD ne présente pas *net-* comme une « combining form » ou un préfixe (ce qui illustre combien il est difficile d'établir des critères dictionnaires). *Net-* est sémantiquement proche de *e-* et *cyber-* et il est apparu à la même période. Dans le CALDO, c'est le substantif *Net* qui possède sa propre entrée. Il est présenté comme étant la réduction d'*Internet*. Contrairement à *e-* et *cyber-*, l'emploi de *net-* comme élément de composition n'est pas reconnu.

Main Entry: net Pronunciation: 'net Function: noun [...] 4 a : a group of communications stations operating under unified control b : NETWORK 5 : INTERNET the Net noun [S] ABBREVIATION FOR the Internet: I've found a really useful website about allergies on the Net.

Il faut préciser que, dès 1969, la réduction *net* sert à former des composés. À cette époque, il ne s'agit évidemment pas de l'abréviation d'*Internet* mais de celle de *network*. *Net* entre dans la composition de noms propres comme Arpanet (1969), NetBEUI,

⁹² <http://www.infoplease.com/ipa/A0193167.html> / ‘Internet timeline’ *Internet* n'est pas une vedette de l'OED 1992. Il figure en exemple dans la définition du préfixe *inter-* : 1 b. denoting « Between or among themselves or one another ; with each other ; mutually, reciprocally, alternately (in a vbl. sb. , mutual, reciprocal), together : as in [...] internet »

Les noms composés anglais français et espagnols du domaine d'Internet :

NetBIOS, NetWare (1983)⁹³ dans lesquels il signifie *network*. Mais avec l'ouverture d'Internet au grand public en 1994 (Colombain, 1996 : 112), *net-* gagne en productivité et entre dans la composition de termes qui ont une fréquence très élevée comme *Net access*, *net economy*.

Tableau 20 : Composés anglais du corpus avec le pseudo-confixenet

Pseudo-confixe	Base	Champ sémantique
net	access	C
	congestion message	C
	connection	C
	economy	C
	<i>izen</i>	N
	marketer	T
	<i>marketing</i>	T
	user	N
Surcomposé : net	connection speed	C

Légende : C = connexion, N = navigation, T = travail

Dans les 9 composés ci-dessus, *net-* a une fonction préfixale (il a quelquefois une fonction suffixale comme dans *EN supernet*). Comme pour la majorité des pseudo-prefixes précédents, l'ordre des éléments est déterminant/déterminé (*net access*, *net economy*).

La réduction *net-* conserve le sens d'*Internet*. Elle peut être remplacée par sa forme développée (sauf dans le cas de l'amalgame *netizen*) : *Internet access*, *Internet congestion message*, *Internet economy*, *Internet user*, etc. Sémantiquement, *net-* a les mêmes caractéristiques que *e-*. Les termes construits avec *net-* désignent des acteurs, occupations et pratiques ne nécessitant aucune formation en informatique. Les thèmes abordés, le commerce, la navigation et le travail, sont moins nombreux que pour *e-*. *Net-* est moins fréquent.

1.2.6. web-

Le pseudo-prefixe *web-* est la troncation antérieure de *World Wide Web*. Le terme *World Wide Web* a été créé vers 1989, par Tim Berners Lee au CERN, pour désigner la partie multimédia d'Internet, la partie la plus populaire (Colombain, 1998 : 220-221).

Le MWOD et le CALDO présentent sommairement *web* :

CALDO : the Web(INTERNET)noun[S] the system of connected documents on the Internet, which often contain colour pictures, video and sound, and which can be

⁹³ « NetBIOS a été développé par IBM pour la communication de correspondant à correspondant entre les PC » et NetBEUI est le « protocole de transport de réseau utilisé par tous les systèmes de réseaux de Microsoft et les systèmes LANserver d'IBM ». (Newton, 2000 : 521-522)

searched for information about a particular subject. MWOD : 9 capitalized : WORLD WIDE WEB

La troncation a très vraisemblablement été utilisée seule pour la première fois en remplacement de la forme développée *World Wide Web*, c'est-à-dire pour réduire le coût linguistique. Il existe bien des sigles, comme *WWW* et *W3*, qui reprennent le terme dans sa totalité mais leur prononciation est plus longue et moins aisée que celle de *web*. *Web* a ensuite été utilisé en composition dans le terme *webmaster*, par exemple, créé en 1994.⁹⁴

Tableau 21 : Composés anglais du corpus avec le pseudo-prefixe web-

Pseudo-composé	Base	Champ sémantique	Base	Champ sémantique
web	<i>accelerator</i>	A	MAIL	Co
	ACCESS	C	manager	A
	ADDRESS	N	marketer	M
	agency	M	MASTER	M
	attack	S	PAGE	N
	BROWSER	A	radio	Co
	<i>bug</i>	S	remix	Co
	CAM	Co	remixer	Co
	center	M	SERVER	C
	cobrowsing	M	SITE	N
	collaborating	M	software	A
	community	Sé	SPACE	N
	DESIGN	M	<i>surfer</i>	N
	DESIGNER	M	surfing habit	N
	DEVELOPER	M	user interface	A
	DIRECTORY	N	high-speed Web access	C
	HOSTING	N	hosting company	N
	LOG	Co		
	logger	Co		
	<i>logging</i>	Co		

Légende : A = application, C = connexion, Co = communication, M = métier, N = navigation, S = sécurité, Sé = société

Web- a lui aussi une fonction préfixale. Tout comme les autres pseudo-prefixes d'origine anglaise, il respecte la syntaxe anglaise. L'ordre des éléments des composés contenant *web-* est déterminant/déterminé. *Web-* peut être remplacé par sa forme complète (*to launch a World Wide Web attack*).

Sept thèmes différents sont recensés dans notre corpus : les applications, la connexion, les métiers, la communication, la navigation, la société et la sécurité. Comme l'observe Brolles (2001 : 75), *web* désigne des intervenants avec ou sans compétence informatique (*webmaster*, *webdesigner* / *web surfer*) mais qui œuvrent au développement

⁹⁴ Source : *Merriam Webster's Collegiate Dictionary*, 10th edition.

d'Internet. Au-delà, *web* identifie le réseau sur lequel se trouve un objet, par exemple, parmi l'ensemble de réseaux qu'Internet constitue. Il sert à désigner tout ce qui réfère aux sites (*web address*, *web directory*). On pourrait même envisager que *web* soit, dans certains termes, la troncation de *website*. *Web* peut alors être remplacé par *site*. Il serait préférable de parler de *site master* (pour *website master*) plutôt que de **World Wide Web master*.

1.2.7. -ware

Ware était à l'origine une lexie autonome désignant des produits souvent utilisée en combinaison (*kitchenware*, *tableware*). Dans le domaine qui nous intéresse, cette lexie a été utilisée pour la première fois en combinaison pour former le composé *hardware* créé lors de l'avènement de l'informatique (Newton, 2000 : 389). L'OED 1992 donne 1953 comme date de première apparition pour *hardware* et 1960 pour *software*.

Le MWOD présente *ware* comme un nom sanscrit, qui signifiait 'prix', souvent utilisé dans des combinaisons alors que le CALDO lui attribue une fonction suffixale.

**MWOD : 1 a : manufactured articles, products of art or craft, or farm produce :
GOODS - often used in combination <tinware> b : an article of merchandise 2 :
articles (as pottery or dishes) of fired clay <earthenware> 3 : an intangible item
(as a service or ability) that is a marketable commodity. CALDO : -ware suffix
used, often in shops, to refer to items of the same material or type, especially
items used in cooking and serving food: tableware, the kitchenware department**

Aucune des deux définitions ne cite l'exemple de *software* ou *hardware*, deux termes pourtant très diffusés.

Tableau 22 : Composés anglais du corpus avec le pseudo-confixe -ware

Base	Pseudo-confixe	Champ sémantique
free	ware	A
group		A
hard		O
open		A
share		A
soft		A
spy		A
chat	Surcomposés avec : software	
Internet connection		A
mailing list management		A
newsgroup		A
software firewall		S
hardware firewall	hardware	S

Légende : A = application, O = ordinateur, S = sécurité

-Ware est le seul pseudo-confixe de notre corpus qui a une fonction suffixale. Il a produit sept néologismes et six termes surcomposés majoritairement formés à partir de *software*.

L'appellation *pseudo-confixe* est justifiée puisque dans les termes qui ont été créés pour désigner des concepts d'Internet, -ware n'a plus le sens de 'produit' comme dans *hardware* et *software* mais le sens de 'software'. Il pourrait même être perçu comme étant la troncation de *software*, ce qui explique peut-être que dans les sept composés et dans quatre des six surcomposés, -ware est l'élément déterminé (alors que la tête d'un composé n'est généralement pas un suffixe). Dans les définitions de *freeware* et *spyware*, l'hyperonyme est *software*.⁹⁵

-Ware pourrait être remplacé par *software* : *freeware* > *free software*, *openware* > *open software*, *spyware* > *spy software*, etc. Par ailleurs, *chat software* pourrait être réduit à *chatware* qui répondrait à la loi du moindre effort. Neuf des douze termes du tableau 22 appartiennent au même champ notionnel, les applications.

1.2.8. Conclusion

Les sept pseudo-confixes de notre corpus anglais peuvent être regroupés dans trois catégories (*supra* p. 89). *Cyber-*, *meta-*, *hyper-* sont des pseudo-confixes sémantiques. L'utilisation dans le domaine d'Internet de troncations de lexies homographes de confixes classiques confère, en quelque sorte, une nouvelle vie à d'anciens confixes. *Web-*, *net-*, *e-*, -ware sont des pseudo-confixes antérieurs (apocopes) et postérieurs (aphérèses). *Net-* et *e-* entraient dans la composition de termes avant l'avènement d'Internet. *Net-* était alors la troncation de *network* et non de *Internet*. *E-* était déjà la troncation de *electronic* utilisée dans le cadre de Arpanet, l'ancêtre d'Internet.

Les fréquences des pseudo-confixes sont également diverses :

⁹⁵ Freeware : free software available on the Internet for downloading . It is used and redistributed at no cost to the user. [...]. Spyware: **software** that gathers information about a user as he or she navigates around the Web, it is intended to track surfing habits in order to build marketing profiles[...]. Source : <http://www.netlingo.com>

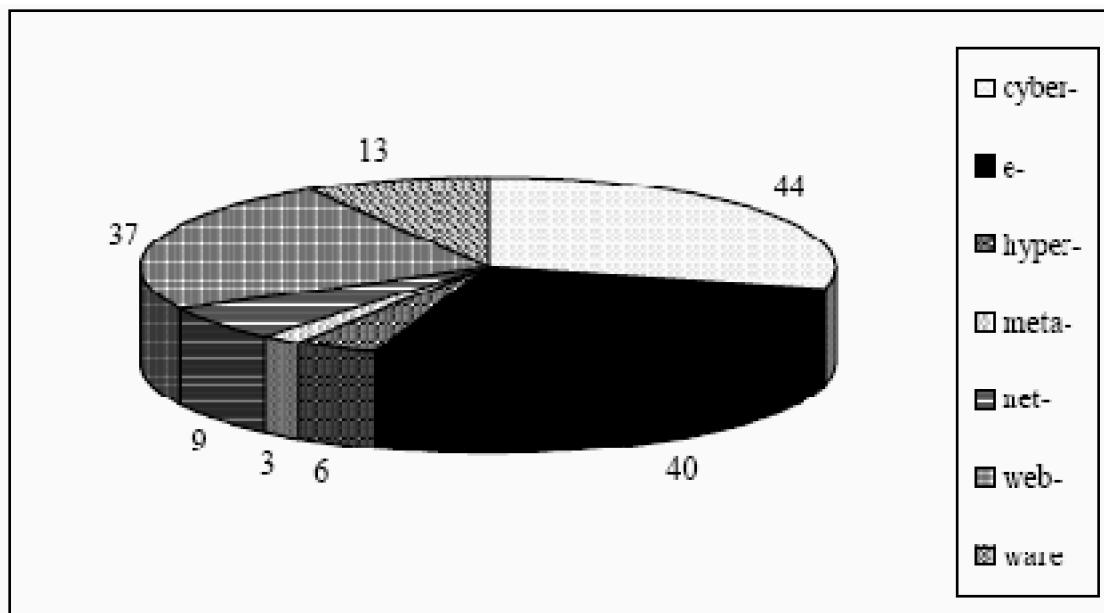


Figure 1 : Fréquence des pseudo-confixes anglais du corpus

Cyber-, rencontré dans 44 termes, est le pseudo-prefixe le plus fréquent. Il est suivi de *e-* qui a formé 40 termes et de *web-* retrouvé dans 37 termes. Les pseudo-prefixes *-ware* (13 termes), *net-*, *hyper-*, et *meta-* ont des fréquences inférieures. Toutefois, la fréquence des lexies produites avec un même pseudo-prefixe est extrêmement variée comme le montrent les chiffres du nombre d'occurrences donnés par Google.com. *Cyberspace*, *e-business*, *weblog* comptent plus de 5 000 000 occurrences, *cyberstress*, *hypertext*, *e-learning*, *web address* sont répertoriés plus de 1 000 000 de fois alors que *cyber-directory*, *e-tourism*, *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 usité que *cyber-recruitment*, par exemple, et il est souvent employé comme synonyme d'*Internet*.

Ces chiffres soulignent la fréquence variée des pseudo-prefixes. Les termes avec *cyber-* ou *e-* sont généralement bien plus fréquents que ceux contenant *meta-* ou *hyper-*. Comme le montre notre analyse des termes collectés, les pseudo-prefixes sont employés dans des champs sémantiques précis et leur différence de fréquence provient de l'importance des champs sémantiques. Certains pseudo-prefixes, comme *cyber-*, entrent dans des champs sémantiques vastes et d'autres, comme *meta-*, dans des champs sémantiques plus restreints. Des pseudo-prefixes désignant des notions similaires, comme *cyber-* et *e-*, sont également prisés par les journalistes (avec respectivement 44 et 40 termes créés).

Enfin, le traitement des pseudo-prefixes dans les ouvrages de référence et à l'intérieur d'un même ouvrage est inégal. Dans le MWOD, tous n'ont pas le même statut. *Cyber-* et *e-* sont des « combining forms », *hyper-* et *meta-* sont des préfixes classiques, *net-* n'a pas de statut (seul figure un renvoi à l'article sur Internet), *web-* et *-ware* sont des noms, mais dans le cas de *web-* aucune mention n'est faite de sa faculté de former des composés. Pour *cyber-*, le domaine d'application mentionné est « computer », pour *e-* le domaine spécifié est « electronic », pour *hyper-* et *meta-* aucune mention de leur

utilisation dans le domaine d'Internet et de leur nouvelle signification. Dans le CALDO, *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 »).

2. Le profil sémantique des noms composés collectés

Les analyses de cette deuxième section répondent à un double objectif. Elles permettront, d'une part, de décrire le profil sémantique des composés du domaine d'Internet et, d'autre part, d'expliquer les exceptions au système d'équivalences types par un croisement de typologies morphosyntaxiques et sémantiques, si nécessaire.

2.1. La centricité

Dans cette première analyse sémantique, les composés du corpus seront classés en fonction de leur degré de transparence par rapport à la classe d'objets qu'ils désignent. Recourir à des définitions simples et claires⁹⁶ ne suffit pas à éliminer tous les obstacles. En effet, nous nous heurtons aux composés dont au moins un des éléments est métaphorique. Par exemple, *address book* a pour déterminé (ou tête) *book*. Peut-on considérer que la classe d'objets à laquelle ce composé renvoie /fonction d'un logiciel de courrier électronique/⁹⁷ est la même que celle à laquelle correspond le déterminé *book* /assemblage de feuilles, destinées à être lues ou écrites, maintenu par une couverture/ ?⁹⁸ Nous devrons définir jusqu'à quel degré de ressemblance entre deux concepts un composé métaphorique est endocentrique. Les composés métaphoriques associant les deux conditions suivantes seront ici considérés comme endocentriques :⁹⁹

- Ils ont une tête identifiable, conformément à la définition de Tournier vue précédemment.
- L'analogie à l'origine de la métaphore est perceptible : analogie de fonction (*mailbox*), analogie d'une propriété perceptive (comme le côté peu attristant et inélégant véhiculé par *nerd* dans *techno-nerd*), analogie de forme (*Web* dans *World Wide Web*), etc.

Appliquons ces conditions à deux unités lexicales. Le composé *address book* a une lexie tête (*book*) et présente une analogie (fonction : répertorier des adresses). Le terme *address book* réunit les deux critères d'endocentricité. Le composé *firewall* n'a pas de lexie tête, un *firewall* n'est pas un type de *wall*. *Firewall* est exocentrique dans le domaine d'Internet mais endocentrique dans sa première acceptation, celle de /mur construit pour empêcher la propagation d'un feu/. Une analogie de fonction (protection) est sentie mais ce seul critère est insuffisant puisqu'un composé correspondant à une classe différente de

⁹⁶ Voir Part. I, Ch. 1, Sec. 1.1. pour les définitions d'endocentrisme et d'exocentrisme.

⁹⁷ Et, plus précisément, fonction servant à établir et gérer une liste d'adresses électroniques.

⁹⁸ L'ambiguïté provoquée les polysémies (Cruse 1986) a déjà été évoquée (Part. I, Ch. 1, Sec. 1.2.5.).

⁹⁹ Le premier critère suffira pour déterminer la classe des composés non métaphoriques.

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, les termes *proxy firewall*, *network firewall*, *hardware firewall* et *software firewall* sont endocentriques. Ces composés ont tous *firewall* pour lexie tête, ce sont tous des hyponymes de *firewall*. Une analogie de fonction (protection) est identifiable. Un *software firewall*, par exemple, est un dispositif de protection comme l'est un *firewall* dans le domaine de la protection incendie.

La méthode d'identification des noms composés trouvée, il faut aménager la procédure de catégorisation. Deux étapes préexistent au classement des noms composés du corpus :

- À chaque composé correspondra une paraphrase. Par exemple, *Internet protocol* est un type de *protocol* ou *black hat* n'est pas un type de *hat*. Pour uniformiser les classes, nous formulerais des paraphrases types dans lesquelles N1 représentera le déterminé et N2 le déterminant.
- Les composés seront ensuite répartis dans les quatre catégories créées à l'aide des paraphrases : N2N1 est un N1, N1N2 est un N1, N2N1 est un N1 et un N2, *N2N1 est N1 et un N2.

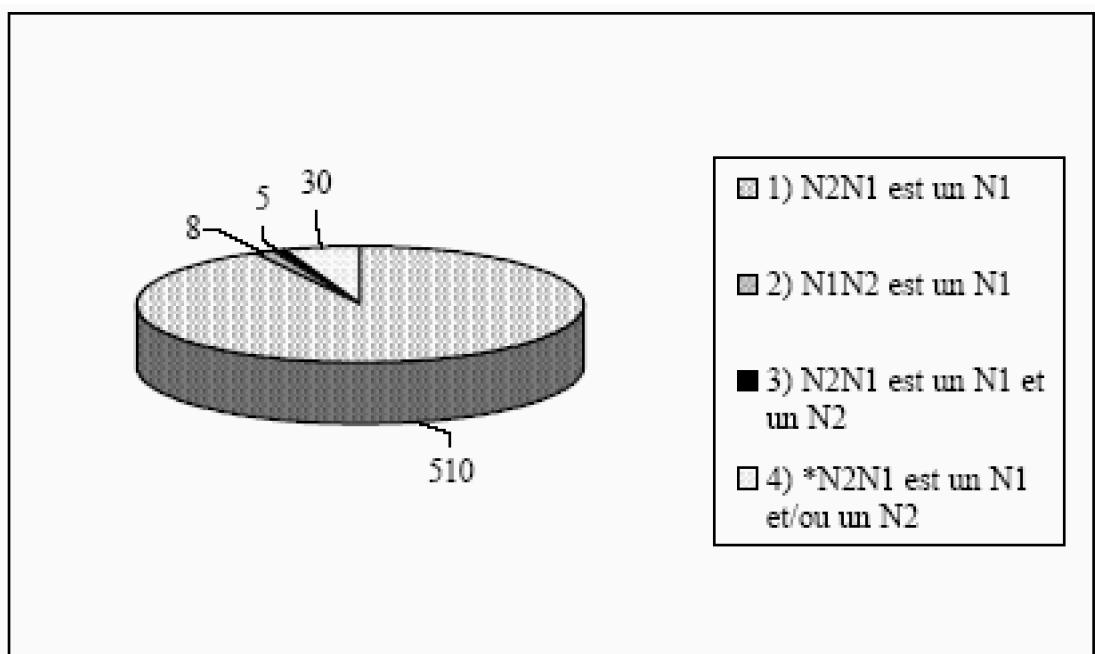


Figure 2 : Centricité des composés anglais du corpus¹⁰⁰

Les composés endocentriques dominent. Les catégories 1, 2 et 3 représentent 94,57% du total des termes. Ce fait n'est pas surprenant puisque dans les vocabulaires anglais des sciences et des techniques, les composés sont majoritairement endocentriques (Depierre, 2000 : 54). Pourquoi une telle prépondérance de l'endocentricité ?

¹⁰⁰ Le classement complet est en Annexe (Annexe 9).

L'endocentricité facilite la compréhension (surtout entre spécialistes et non spécialistes). Les combinaisons endocentriques présentent un certain degré de motivation.¹⁰¹ Les composés endocentriques « exposent dans leur signifiant un rapport non arbitraire, mais non iconique (du genre de l'onomatopée), avec leur signifié » (Arnaud, 2003 : 35). La composition est motivée morphologiquement et la métaphore – une motivation sémantique – étant présente dans la plupart des composés recueillis, de nombreux termes ont une double motivation morphologique et sémantique.

Respectant l'ordre déterminant/déterminé des composés anglais, les composés de la catégorie 'N2N1 est un N1' sont les plus nombreux (92,22% du corpus anglais). 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, 'N1N2 est un N1', et ceux dont les deux éléments sont sur le même plan, 'N2N1 est un N1 et un N2', sont peu nombreux (respectivement 1,45% et 0,90%). Les composés de la catégorie 2 sont en majorité de type roman comme *network of networks*.

La catégorie des composés exocentriques occupe la deuxième place mais les critères d'exocentrisme utilisés font que son pourcentage est bien inférieur à celui des endocentriques respectant l'ordre déterminant/déterminé (5,43% contre 92,22%). Elle est principalement constituée de composés empruntés à d'autres domaines : *backdoor* (construction), *firewall* (protection incendie), *bottleneck* (gestion/circulation routière), *Trojan Horse* (mythologie), *thumbnail* (anatomie, publicité), *guestbook* (hôtelierie), *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).

En raison de leur particularité sémantique (non-compositionnalité due à la métonymie ou à la métaphore) et de leur origine (emprunts internes à d'autres domaines), il est évident que les composés exocentriques seront plus problématiques pour la mise en place du système d'équivalences types que les composés endocentriques. Comme ce sont des emprunts internes utilisés pour dénommer un nouveau concept dans la langue source, il y a une forte probabilité pour que des termes français et espagnols désignant ces concepts existent. Par exemple, EN *firewall* et *bottleneck* avaient des équivalents avant Internet (FR *mur coupe feu*, *goulot d'étranglement* ; ES *cortafuegos*, *cuello de botella*). Les traductions constatées dans les corpus sont l'ajout d'un nouveau sens au terme français ou espagnol existant (FR *porte de derrière* ; ES *puerta trasera*, *cortafuegos*) ou l'emprunt (FR *startup*, *dotcom* ; ES *ciberpunk*, *firewall*). Ces deux solutions seront préconisées (Part. III, Ch. 3, Sec. 2.2.) car la non-compositionnalité affaiblit la possibilité de recourir au calque (*bottleneck* > **goulot de bouteille*, *click and mortar* > **clic et ciment*, *black hat* > **chapeau noir*).

2.2. Les relations sémantiques

Arnaud (2003 : 37 seq.) a récemment fait l'inventaire de la littérature sur les relations de

¹⁰¹ Voir Part I, Ch. 1, Sec. 2. pour les détails sur cette notion.

détermination entre les éléments des noms composés (relations N1-N2). Il en tire quelques leçons (*ibid.* : 61) : 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, d) il est nécessaire de s'interroger sur la forme des étiquettes des catégories.

Les noms composés du corpus anglais seront classés en fonction de leur sens connu. Le nombre de catégories du classement ne sera pas limité, ce qui a l'avantage de ne pas regrouper des composés dont les relations ne sont pas tout à fait identiques. Imposer un nombre limite pourrait occulter des particularités sémantiques utiles pour notre deuxième objectif (justification des exceptions du système d'équivalences). Ensuite, il faudra s'interroger sur le degré de formalisme à adopter. Les relations sémantiques entre les éléments des composés seront exprimées par une paraphrase. Comme pour l'analyse de la centricité, dans toutes les formulations N1 désignera le déterminé et N2 le déterminant. Enfin, la méthode de classification des composés sera semblable à celle utilisée par Arnaud (2003 : 61 sq.) :

- Nous partirons des données du corpus et formulerons pour chaque composé une interprétation avec N1 et N2. Par exemple, *domain name* se paraphrase en 'N1 est attribué à N2' et *web address* en 'N1 permet de situer N2'.
- Nous procéderons à l'uniformisation des paraphrases. La relation entre les éléments de *domain name* et *web address* est une relation d'identification. Un nom identifie un domaine tout comme une adresse identifie un site web. Les paraphrases 'N1 est attribué à N2' et 'N1 permet de situer N2' seront réunies sous l'étiquette 'N1 identifie, situe N2'.
- Dans le cas d'ambiguités catégorielles, le composé sera inséré dans une nouvelle catégorie regroupant les deux catégories possibles.

Tableau 23 : Synthèse des relations sémantiques des composés anglais du corpus

Code	Description	Exemples	Fréquence « type »
1	N1 est désigné par analogie à une propriété de N2	<i>snail mail, mirror site</i>	5
2	N1 s'adresse à N2 (e-commerce, échange)	<i>peer-to-peer, business to consumer</i>	4
3a	N1 (logiciel, matériel, technique) a pour activité, fonction N2	<i>archive manager, download site</i>	82
3b	N1 (humain, entreprise) a pour object d'activité, fonction N2	<i>space provider, bug hunter</i>	32
3c	N1 (virus, attaque, pirate) a pour fonction N2	<i>DoS attack, buffer overflow attack</i>	15
4	N1 a un fonctionnement dont le procédé, le mode est N2	<i>ADSL modem, ebook</i>	104
5	*N1N2 est un N1 et/ou un N2	<i>firewall, backdoor</i>	24
6	N1 se distingue des autres N1 car il est N2 (nature)	<i>freeware, general portal</i>	78
7	N1 a N2 comme localisation, environnement	<i>email virus, cyber-shop</i>	74
8	N1 est combiné à N2	<i>hacktivist, modem</i>	5
9	N1 est engendré, provoqué par N2	<i>browsing habit, virus alert</i>	16
10	N1 est constitué, formé par N2	<i>hacker community, mail bombing</i>	6
11	N1 évalue, mesure N2 (durée, vitesse, nombre)	<i>chat session, connection time</i>	8
12	N1 est transféré en direction N2	<i>download, upload</i>	3
13a	N1 est un instrument, mode, moyen utilisé par N2	<i>client interface, user account</i>	5
13b	N1 est un instrument, mode, moyen utilisé pour N2	<i>access control, Internet protocol</i>	92
TOTAL			553

Le tableau 23 reproduit treize relations sémantiques principales. Si l'on détaille les descriptions 3a, 3b, 3c et 13a, 13b, le nombre de relations s'élève à seize.

Aucune catégorie ne regroupe plusieurs relations sémantiques, ce qui ne signifie pas qu'aucun cas d'ambiguïté catégorielle ne s'est présenté. La relation N1-N2 de certains noms composés a été très complexe à paraphraser. Comme le constate Jouneau :

« 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. » (1998 : 59)

Par exemple, la relation qui unit *address* (N2) et *book* (N1) est-elle 10 'N1 est constitué, formé par N2' ou 13b 'N1 est l'instrument, mode, moyen utilisé pour N2' ? 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 :

- 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. En effet, les descriptions 3a, 3b et 3c indiquent toutes une fonction. Seule la nature de N1 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, N1 est actif et concret, il est un objet, une personne ou une entreprise dont la fonction principale est d'«accomplir» N2 (*ad blocker*, *ADSL provider*). Dans les paraphrases 13, N1 est passif et immatériel, il est un moyen, une technique, utilisé pour 'agir sur' N2 ou qui est utilisé par N2 (*distribution list*, *encryption key*).
- Comme on pouvait s'y attendre, la ressemblance entre certains composés ne doit pas être prise en compte. *Address book* et *e-book* possèdent le même déterminé mais dans les deux composés, le déterminant et le déterminé ne sont pas unis par la même relation sémantique : 13b 'N1 est l'instrument, mode, moyen utilisé pour N2' et 4 'N1 a un fonctionnement dont le procédé, le mode est N2'.
- Autre fait évident, plusieurs relations sémantiques correspondent à un type morphosyntaxique. Par exemple, *address book*, *boot virus* et *bottleneck* sont des composés N + N et leurs catégories sémantiques diffèrent : 13b 'N1 est l'instrument, mode, moyen utilisé pour N2', 3c 'N1 (force malveillante) a N2 pour cible' et 5 '*N2N1 est un N1 et/ou un N2'.
- La catégorie 5 réunit les composés dont la relation sémantique est complexe. Elle contient majoritairement des exocentriques. Pour éviter de multiplier les catégories hapax, les exocentriques ont été regroupés bien que leurs paraphrases soient parfois radicalement différentes : *firewall* > 'N2N1 est un dispositif de protection', *intelligent agent* > 'N2N1 est un programme, un procédé automatique', *mailto* > 'N1N2 est un mot clé', etc. Cependant, sur les 30 composés estimés exocentriques (Part. II, Ch. 2, Sec. 2.1.), 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*¹⁰² 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 lexie tête mais l'analogie est, d'après nous, imperceptible (les termes sont opaques).

Les relations sémantiques les plus fréquentes sont :

- La fonction avec N1 actif (3a, 3b et 3c). Cette relation est présente dans 129

¹⁰² *Shopbot* est un amalgame construit à partir de *shopping* et *robot*.

- Avec 104 composés, le principe de fonctionnement (4) arrive en deuxième position (18,81% du corpus).
- La nature (6) est la troisième relation la plus fréquente avec 78 composés (14,10% du corpus).
- La fonction avec N1 passif (13a et 13b). Le déterminant et le déterminé sont unis par cette relation dans 97 composés, soit 17,54% du corpus.

En conclusion, une ou plusieurs relations sémantiques seront-elles à l'origine des exceptions du système d'équivalences types ? Les composés des catégories minoritaires, et surtout ceux de la catégorie 5, seront-ils plus des obstacles pour le système d'équivalences types que les composés des types majoritaires ? C'est ce que nous tenterons de découvrir dans la troisième partie.

3. La combinatoire intermatricielle

Dans notre corpus trilingue, il est fréquent que plusieurs matrices lexicogéniques interviennent simultanément dans la formation d'un terme du domaine d'Internet. Selon nos critères (Part. II, Ch. 2, Sec. 3.2.1. et 3.2.4.), parmi les 553 composés du corpus anglais, seulement 54 (9,7%) sont formés par un seul processus. Deux ou trois matrices se combinent souvent pour produire une nouveauté lexicale. Notre but est de savoir quelles matrices présentent des affinités avec la composition et entre elles afin de tenter de prévoir les difficultés du système d'équivalences types (Part. III).

3.1. Les matrices lexicogéniques

Pour étudier la combinatoire, la connaissance de toutes les matrices lexicogéniques s'impose. C'est pour cette raison que nous fonderons notre analyse sur les définitions de Tournier (1991b). Tournier (1985: 51) dénombre treize matrices qu'il classe en trois catégories :

- Les matrices morphosémantiques. Les lexies créées sont nouvelles tant sur le plan du signifiant que du signifié. Il s'agit de la composition (par juxtaposition), la préfixation, la suffixation, la dérivation inverse et l'amalgame.
- Les matrices sémantiques. Les lexies sont une nouveauté uniquement sur le plan du signifié. Il s'agit de la conversion, la métaphore, la métonymie et l'onomatopée.
- les matrices morphologiques. Seul le signifiant constitue une nouveauté. Ces lexies sont issues de la siglaison, l'apocope et l'aphérèse.

Tournier n'inclut pas le treizième processus, l'emprunt, dans cette classification. Contrairement aux douze autres matrices, c'est un processus externe ce qui signifie que les ressources proviennent de systèmes linguistiques extérieurs et non de la dynamique interne. Les matrices peuvent se combiner pour créer une nouveauté lexicale. Tournier parle de combinatoire matricielle. Il distingue deux types de combinatoire qui peuvent créer un néologisme :

- La combinatoire intramatricielle. Une même matrice est utilisée plusieurs fois dans la formation d'un terme. On parle alors de récursivité. On l'a vu, la composition peut être récurrente : *download* est réutilisé dans *download manager*, *download site*, *download speed*. *Digital Subscriber Line* est un composé réutilisé dans la formation d'autres composés dont *Asymmetric Digital Subscriber Line*, *High-Speed Digital Subscriber Line* et *Symetric Digital Subscriber Line*, *Rate-Adaptative Digital Suscriber Line*.
- La combinatoire intermatricielle. Une matrice lexicogénique est utilisée conjointement avec d'autres matrices dans la formation d'un terme. *Hacker attack* associe la composition et la métaphore (dans *hack/hacker* et dans *attack*) et *dynamic IP protocol* combine la composition qui est récurrente (*IP protocol*), la siglaison (*IP*) et la métaphore (*protocol*).

3.2. Méthode d'analyse du corpus anglais

Étant donné la nature particulière du lexique d'Internet (emprunts à d'autres domaines de spécialité, à la langue générale), la méthode de calcul de la combinatoire utilisée ici ne comptabilise pas toutes les matrices apparentes.

3.2.1. Contexte¹⁰³ considéré

Trois exemples (*ad-blocker*, *bookmark* et *search engine*) vont nous permettre d'expliquer notre démarche. Le découpage hors contexte du terme *ad-blocker* fait apparaître les matrices suivantes : 1) double troncation postérieure (*ad < advert < advertisement*), 2) suffixation (-er dans *blocker*) et 3) composition (*ad + blocker*). Dans le contexte d'Internet, seule la matrice de la composition est retenue. En effet, les éléments *blocker* et *ad* existaient déjà indépendamment. Les matrices de la troncation et de la suffixation sont antérieures à l'utilisation du terme pour désigner une réalité d'Internet. Hors contexte, *bookmark* est considéré comme un composé (*book + mark*). Dans le lexique d'Internet, la composition est exclue car la combinaison *book + mark* existait avant l'utilisation du terme dans ce domaine. Seule la matrice de la métaphore est prise en compte (métaphore de fonction). Le terme *search engine* est formé par les matrices de la métaphore (*engine*) et de la composition (*search + engine*). C'est un composé partiellement métaphorique (un seul formant comporte une métaphore).

Comme l'indiquent ces exemples, seules les matrices utilisées pour le domaine d'Internet seront comptabilisées. On ne remontera pas jusqu'à la création de chacun des éléments constitutifs des composés car certains éléments ne sont pas pensés au moment de la nouvelle création mais simplement réutilisés pour dénommer une réalité d'Internet. Cette démarche a pour but d'éliminer le bruit, c'est-à-dire d'occuper les matrices dont les utilisateurs n'ont pas conscience lorsqu'ils commencent à employer un terme existant pour désigner une nouvelle réalité. Cependant, dans les termes incluant un élément issu du domaine de l'informatique, l'élément tiré du lexique informatique sera considéré. Par exemple, *chat software* est formé par les matrices suivantes : métaphore (*chat*),

¹⁰³

Par contexte, nous entendons ici domaine ou ensemble des domaines dans lequel le signifiant étudié apparaît.

composition (*software*) et surcomposition (*chat + software*). Nous opérerons ainsi pour les motifs suivants :

- La frontière entre les deux disciplines est floue. Internet est en grande partie basé sur les ressources informatiques. Certains termes sont communs aux deux disciplines.
- Il y a un changement de dimension. Le contexte n'est plus le même et la fréquence des termes est différente dans chacun des deux domaines. *Software* a vu sa fréquence augmenter parce qu'Internet nécessite le développement de nouveaux types de softwares (*chat software*, par exemple).

3.2.2. Pseudo-confixes et troncations

Une nouvelle matrice sera ajoutée aux matrices de Tournier. Il s'agit de la pseudo-confixation. Il semble indispensable de consacrer une catégorie à la pseudo-confixation qui, comme nous l'avons constaté, est très fréquente (Part. II, Ch. 2, Sec. 1.2.).

Deux exemples permettront d'y voir plus clair. *E-mail* est un terme qui combine 1) la métaphore (*mail*), 2) la pseudo-confixation (*e-* < *electronic*) et 3) la composition (*e- + mail*). Dans *webzine*, sont présentes les matrices de la métaphore (*web*), de la pseudo-confixation (*web*), de la troncation (*zine > magazine*) et de la composition (*web + -zine*). La plupart des pseudo-confixes sont des troncations de lexies fréquemment utilisées et sont semblables à des confixes (*e-*, *meta-*, *net*, etc.). Les troncations peu fréquentes et éloignées des confixes seront répertoriées en tant qu'apocopes ou aphérèses, comme chez Tournier, et non en tant que pseudo-confixes. Par exemple, *open source app* est formé de : 1) un composé (*open source*), 2) une apocope (*app<application*), et il est 3) un surcomposé (*open source + app*). *App* n'est pas un pseudo-confixe (fréquence « type » très faible dans le corpus).

3.2.3. Composés et amalgames

Les amalgames étant un type de composés, ils seront détaillés au même titre que les composés. Si *search engine* est à fois un composé et une métaphore alors *Internet* est un amalgame (*inter national + net work*) et une métaphore (*network*). *Entreprenaute* est un amalgame formé d'une troncation postérieure (*entrepre neur*) et d'un pseudo-confixe (*-naute*).

3.2.4. Extraction des matrices

Quelques exemples vont exposer la démarche d'extraction des matrices utilisées dans les termes du corpus comparable trilingue (voir Part. II, Ch. 3 et 4 pour le français et l'espagnol) :

- *electronic mail* est formé par : 1) la métaphore (*mail*), et 2) la composition (*electronic + mail*).
- *email* est un terme créé par les matrices de 1) la métaphore *mail*, 2) la

pseudo-confixation (*e*), et 3) la composition (*e-* + *mail*).

- dans *e-mail address* les matrices recensées sont : 1) la métaphore (*address*), 2) la composition (*e-mail*) et 3) la composition (*e-mail* + *address*).
- *peer-to-peer* est un composé (*peer* + *to* + *peer*) et une métaphore (*peer*).
- *P2P network* est un composé (*P2P* + *network*) formé par les matrices de la siglaison (*P2P*) et de la métaphore (*network*).

3.3. Résultats¹⁰⁴

3.3.1. Analyse

Tableau 24 : Combinaisons des matrices lexicogéniques du corpus anglais

¹⁰⁴

Abréviations utilisées pour l'analyse de la combinatoire : A = amalgame, Ah = aphérèse, Ap = apocope, C = composition, M = métaphore, Me = métonymie, P = préfixation, PS = pseudo-confixe, Si = sigle.

Rang	Combinaisons	Fréquence « type »	%	Exemples
1	C M	104	18,8%	<i>massive attack</i>
2	C M M	92	16,6%	<i>archive manager</i>
3	C P S	60	10,8%	<i>cyber activity</i>
4	C C M	55	9,9%	<i>Denial of Service attack</i>
5	C P S M	48	8,7%	<i>cyberworld</i>
6	C C	32	5,8%	<i>encrypted email</i>
	C Si M			<i>cookie ID</i>
7	M	30	5,4%	<i>address book</i>
8	C	19	3,4%	<i>ad blocker</i>
9	C A	12	2,2%	<i>analog modem</i>
10	C C C	11	2,0%	<i>dotcom startup</i>
11	C A M	10	1,8%	<i>Internet access</i>
12	C Me M	6	1,1%	<i>antivirus provider</i>
	C P S C			<i>meta search engine</i>
13	A	5	0,9%	<i>Internet</i>
	C A C			<i>Internet Service Provider</i>
14	C Me	4	0,7%	<i>Boolean search</i>
	C Si			<i>anonymous FTP</i>
15	C Me Ap	2	0,4%	<i>dot com</i>
	C Si Me M			<i>MP3 site</i>
16	A M Ah	1	0,2%	<i>shopbot</i>
	C Si C			<i>POP3 email box</i>
	C C C M			<i>always-on broadband connection</i>
	C Si Me			<i>voice IP</i>
	A PS			<i>eiz</i>
	A PS M			<i>netizen</i>
	C Ap M			<i>techno-nerd</i>
	C C Ap			<i>open source app</i>
	C M M M			<i>Challenge Handshake Authentication Protocol</i>
	C Me C			<i>voice email</i>
	C Me Si A			<i>ADSL modem</i>
	C P S P			<i>web cobrowsing</i>
	C S Me M			<i>Napsterized file</i>
	C M M C			<i>mailing-list management software</i>

Les noms composés du corpus anglais ont été formés d'après trente-quatre schémas combinatoires. De une matrice (*ad blocker*, *address book*, etc) à quatre matrices différentes (*MP3 site*, *Napsterized file*, etc.) sont utilisées dans la création des termes recensés. La moyenne est de 16,3 composés par schéma combinatoire. Huit combinaisons (CM, CMM, CPs, CCM, CPsM, CC, CSiM, M, C), représentant 85,4% des

noms composés, se situent au-dessus de cette moyenne.

Selon les critères mentionnés (Sec. 3.2.4.), la composition est présente dans trente-trois combinaisons sur trente-quatre. Elle représente donc 94,6% des termes du tableau 24. Cette matrice terminogénique est récurrente dans treize combinaisons (C C M, C C C, C A, C C C, C A M, C Ps C, C A C, C Si C, C C C M, C C Ap, C Me C, C Me Si A, C M M C), soit 24,9% des termes.

La composition se combine le plus fréquemment avec la métaphore (quinze schémas sur trente-quatre). Dans l'ensemble du corpus, 64,5% des termes sont formés à partir de la composition et la métaphore. Nombreux sont les composés contenant un élément métaphorique (47,5%), voire deux (17%). Le pourcentage de composés avec trois éléments métaphoriques est très faible (0,2%). En ajoutant le cas où la métaphore n'est pas associée à la composition, cette matrice sémantique est présente dans 69,9% des termes (seize schémas combinatoires sur trente-quatre). Hormis la composition et la métaphore, aucune autre matrice n'a une récursivité attestée dans notre corpus. Avec la métaphore, la métonymie est la seule matrice sémantique présente dans notre échantillon. Elle entre dans la formation de seulement 3,4% des composés (huit combinaisons).

La deuxième combinaison la plus fréquente est composition-pseudo-confixation. Ces deux matrices ont produit 21,2% des termes du corpus (10,8% C Ps, 8,7% C Ps M, 1,1% C Ps C, 0,2% pour A Ps, A Ps M et C Ps P).

En regroupant les résultats du tableau 24, c'est-à-dire en calculant combien de fois une matrice est utilisée en même temps que la composition – en ignorant la récursivité des matrices – on obtient la synthèse suivante (tableau 25).

3.3.2. Bilan

Tableau 25 : Fréquence des combinaisons de matrices dans le corpus anglais

Combinaisons	%	Fréquence « type »
Composition + métaphore	64,5	15
Composition + pseudo-confixation	21,2	6
Composition + siglaison	7,5	6
Composition + métonymie	3,4	8
Composition + apocope	0,6	2
Composition + aphérèse	0,2	1
Composition + préfixation	0,2	1
Composition + suffixation	0,2	1

Le regroupement des combinaisons fait apparaître que la métaphore est de loin la matrice la plus associée à la composition (64,5%). Dans le lexique de la science fiction, une seule matrice (la composition) est à l'origine de la plupart des néologismes¹⁰⁵ et la

¹⁰⁵ Les deux tiers sont créés par la composition (puis métonymie et préfixation).

combinaison composition-métaphore implique au moins quinze lexies, au même titre que composition-emprunt, composition-siglaison et composition-métonymie (Gindre, 1998 : 53). Pourquoi une telle affinité entre la composition et la métaphore dans le domaine d'Internet ? Pourquoi cette affinité se démarque-t-elle autant des autres combinaisons et notamment de la métonymie ?

Tout d'abord, il est utile de préciser que la forte proportion de termes formés à partir de ces deux matrices n'est pas inhabituelle puisque, selon Ayto (1996 : 184), la grande majorité des néologismes anglais de la fin du XX^e siècle provient de la composition, de l'affixation et des changements sémantiques.¹⁰⁶ Cependant, la métaphore occupe une place beaucoup plus importante dans le lexique d'Internet que dans l'anglais général et dans d'autres langues de spécialité. Par exemple, selon Brocard (1998 : 17), les termes du lexique du cinéma américain sont souvent formés par métaphore (sixième matrice la plus fréquente avec 39 termes sur 985). Pour répondre aux questions précédentes, nous allons étudier succinctement quelques fonctions de la métaphore. La littérature sur la métaphore étant abondante, nous reprendrons les études de Brolles (2001 ; thèse en cours)¹⁰⁷ qui répertorient les fonctions de la métaphore utiles dans le cadre d'études néologiques :

- La fonction persuasive. Un locuteur met tous les atouts de son côté pour convaincre son interlocuteur en minimisant ou supprimant des traits au profit d'autres. D'après Brolles, cette fonction est surtout utile aux éditeurs et concepteurs de logiciels. Par exemple, le nom *Net nanny* illustre la volonté des concepteurs de donner une image positive au concept. Le logiciel de contrôle parental acquiert une relation affective de protection. La fonction persuasive semble moins adaptée au lexique d'Internet qu'aux noms de produits.
- La fonction de séduction. Elle inclut une fonction esthétique et une fonction ludique. Le locuteur éviterait les formules banales grâce à la métaphore, c'est la fonction esthétique. L'aspect inventif d'une métaphore est primordial pour qu'elle soit adoptée et diffusée. La métaphore crée une complicité entre le concepteur et l'utilisateur en évitant le style formel des discours scientifiques (Brolles, 2001 : 68). Comme l'ont fait remarquer Meyer et al. (1997), la nouvelle génération d'adeptes d'Internet apprécie les formations ludiques. Les noms *Nero* et *Sharlock* illustrent cette tendance. Certaines lexies seraient créées et utilisées par jeu, comme *FR souris*.
- La fonction heuristique. Elle englobe une fonction conceptualisatrice¹⁰⁸ et une fonction didactique. L'homme appréhende l'univers de façon métaphorique (Lakoff et Johnson, 1980 : 13). La métaphore serait un outil permettant de saisir l'abstrait ou d'en donner une représentation. Brolles cite l'exemple du FR *bureau* et explique que les programmeurs de Xerox et d'Apple voulaient créer une interface graphique dont

¹⁰⁶ Les données de l'auteur proviennent de l'analyse de dictionnaires de néologismes.

¹⁰⁷ Thèse de doctorat sur les métaphores en cours. Informations tirées du chapitre 3 : Les utilisations de la métaphore.

¹⁰⁸ Parmi toutes les fonctions citées, celle-ci est la moins bien admise.

le principe était le même que celui d'un bureau. C'est la fonction conceptualisatrice. En expliquant l'inconnu par le connu, la métaphore est un instrument didactique. Un novice peut apprécier un nouvel objet ou un nouveau concept à partir d'une expérience passée. Le terme FR *cheval de Troie* permet à un novice de comprendre le mode de fonctionnement du programme : un programme malveillant est caché à l'intérieur d'un autre programme. Non seulement la métaphore offre des repères mais elle contribue également à la diffusion de concepts complexes et à la communication.

La fonction cryptique. Brolles prend l'exemple de la musique Blues pour décrire cette fonction. Les métaphores du Blues ont une fonction dissimulatrice car elles forment un code compris de la seule population noire. Par exemple, les propriétaires des plantations ne comprenaient pas les Negro-Spirituals chantés par les esclaves.

Les propriétés de la métaphore et de la composition expliquent leur succès en tant que procédés de création terminologique en anglais. La combinaison composition-métaphore confère aux termes des qualités facilitant la compréhension des termes créés. Le fait que la métaphore interfère avec la centricité n'est pas un obstacle à la compréhension des concepts par les locuteurs puisque 94,20% des composés anglais sont endocentriques et que des composés exocentriques peuvent contenir une métaphore didactique (EN *Trojan Horse*, *backdoor*, *bottleneck*). La grande majorité des analogies est perceptible. Dans le domaine d'Internet, les métaphores sont le plus souvent didactiques (EN *address bar*, *chat server*, *cybercrime*, *virus attack*, etc.). Par ailleurs, plus une dénomination contient d'éléments de nomination, et par là de traits conceptuels, plus elle tend à être transparente, à dévoiler le concept désigné. Grâce à une analogie de forme (EN *World Wide Web*), de fonction (EN *virus writer*) ou à son caractère ludique (EN *snail mail*, *hacktivism*), un composé métaphorique est peut-être plus facile à mémoriser qu'un composé non métaphorique.

Pour répondre à la deuxième question – pourquoi une si faible représentation de la métonymie (3,4%) par rapport à la métaphore (64,5%) ? – nous nous appuierons sur l'hypothèse de Brolles (2001 : 34). La métonymie, construite à partir d'une association d'idées, serait un processus de nomination moins spontané que la métaphore qui, elle, est basée sur une ressemblance. Un concept a déjà une dénomination avant qu'un rapprochement métonymique ne soit effectué. Brolles (2001 : 35) cite la métonymie *dotcom*, désignation apparue après *startup*. Le signifiant renvoie à un concept possédant déjà un signifiant (*startup*). Les métonymies n'ont peut-être pas eu le temps de se développer parce que le domaine d'Internet est récent.

Quant à la dérivation (préfixation et suffixation), elle figure souvent parmi les processus lexicogéniques les plus fréquents. C'est le cas par exemple dans la terminologie des systèmes experts (Cormier et Rioux, 1991 : 248). Dans notre étude, la très faible fréquence de la préfixation et de la suffixation s'explique par les limites fixées (Sec. 3.2.4.). Cormier et Rioux (*ibid.* : 261) citent entre autres *-ion* et *-eur* comme suffixes fréquents (*acquisition de connaissances*, *capteur d'acquisition des données*). Or, pour nous, *acquisition* et *capteur* sont des emprunts sémantiques à d'autres domaines puisque les formes existaient avant leur utilisation dans la terminologie de l'intelligence artificielle. La suffixation, n'ayant pas été pensée au moment de l'introduction de ces lexies dans une

nouvelle terminologie, n'est pas comptabilisée.

Dans notre classement, le processus lexicogénique de la siglaison occupe la troisième place. Au regard de sa principale caractéristique, la brièveté, l'importance de la siglaison n'est pas inattendue. Mais l'apocope et l'aphérèse, elles aussi des réductions, sont très faiblement représentées (0,8%). Ce chiffre concorde avec ceux de Ayto (*ibid.* : 184). Les « shortenings » occupent 1,7% des entrées du Barnhart *Dictionary of New English 1963-1972* (1973), 2,1% de celles du *Second Barnhart Dictionary of New English* (1980) et 2,2% de celles du *Longman Register of New Words* (1989). Nos critères excluent certains raccourcissements : *ad* (sept occurrences), *biz* (une occurrence).

Le lexique anglais d'Internet se caractérise par un recours massif à la combinaison composition-métaphore, ce qui le rapproche de la langue parlée – plutôt que des terminologies techniques soutenues. En effet, administrateurs de sites et internautes occasionnels utilisent les mêmes termes. On ne peut pas opérer de distinction langue scientifique - langue populaire dans ce domaine.

Chapitre 3. Les noms composés français

L'analyse de la formation morphosyntaxique et sémantique des composés d'Internet continue avec l'étude des composés du lexique français. La méthode d'analyse de la morphosyntaxe est identique à celle du chapitre précédent : les noms composés sont classés selon leurs patrons syntaxiques. Des comparaisons interdisciplinaires mettront en valeur les particularités des composés d'Internet. Confronter les données du français avec les résultats obtenus pour l'anglais permettra d'évaluer l'influence de l'anglais sur le français.

1. Le profil morphosyntaxique des noms composés collectés

Le profil morphosyntaxique sera dressé grâce à une double typologie car une typologie unique mélangerait des critères de nature différente, c'est-à-dire l'origine (en incluant les emprunts à l'anglais) et la fonction (en détaillant les patrons syntaxiques), et obscurcirait les grandes tendances. Un premier classement, basé sur l'origine des termes (composés endogènes vs. composés exogènes), permettra d'isoler les emprunts à l'anglais. Un deuxième classement, reposant sur les patrons syntaxiques des composés et surcomposés endogènes uniquement, permettra de savoir si les patrons syntaxiques observés dans le lexique français d'Internet sont ceux habituellement employés pour créer des composés en français et/ou si ce sont des calques des structures anglaises.

1.1. Les structures compositionnelles et leur fréquence « type »

1.1.1. Les composés binaires

Tableau 26 : Structures compositionnelles endogènes et exogènes des composés du corpus français

Les noms composés anglais français et espagnols du domaine d'Internet :

	Structures	Fréquence « type »	%
1	<u>Composés endogènes :</u>	<u>461</u>	<u>85,37%</u>
1.1	purs	225	41,11%
1.1.1	N + de + N	99	18,33%
1.1.2	N + A	85	15,74%
1.1.3	N + N	24	4,44%
1.1.4	PS + N	5	0,93%
1.1.5	N + à + N	3	0,56%
1.1.6	A + N	2	0,37%
1.1.7	N + PS	2	0,37%
1.1.8	non + N	1	0,19%
1.1.9	V + N	1	0,19%
1.2	hybrides	236	44,26%
1.2.1	N + N	93	17,22%
1.2.2	PS + N	63	11,67%
1.2.3	N + de + N	32	5,93%
1.2.4	N + PS	19	3,52%
1.2.5	N + A	17	3,15%
1.2.6	Flud	10	1,85%
1.2.7	PS + A	3	0,56%
1.2.8	N + de + PS	2	0,37%
2	<u>Composés exogènes</u>	<u>79</u>	<u>14,63%</u>
	TOTAL	540	100,00%

Dans le tableau 26, les composés endogènes sont subdivisés en composés endogènes purs et composés endogènes hybrides. Les éléments formant les composés de la première sous-catégorie sont tous d'origine française (*FR attaque de ver, groupe de discussion*).¹⁰⁹ Les noms composés de la deuxième sous-catégorie contiennent à la fois un ou des éléments d'origine française et un ou des éléments du xénolexique, ici l'anglais, (*FR canal de chat, chiffrement d'email*). Ils respectent la syntaxe du français. Ils sont à mi-chemin entre les véritables composés endogènes et les emprunts intégraux.

Les composés du lexique français d'Internet sont massivement endogènes (85,37%). La proportion des termes exogènes est faible (14,63%), 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 ?

Premièrement, la nature même des unités lexicales étudiées explique le faible taux

¹⁰⁹ C'est l'étymologie immédiate du terme qui nous intéresse et non son étymologie « ultime ». Identifier le créateur et la date de création d'un terme permettent d'établir son étymologie. Humbley (1994) développe une méthode de datation des termes techniques. À ce stade, nous ne prenons pas en compte l'ajout d'un nouveau signifié à un signifiant existant sous une influence étrangère.

d'emprunts. Notre étude est consacrée à des unités lexicales longues (entre deux et cinq éléments). Or, dans la langue générale ou dans les langues de spécialité, les lexies empruntées sont généralement brèves. EN *bug*, *chat* sont monosyllabiques et *spyware*, *adware* ou *popup* sont dissyllabiques.

Autre fait non négligeable, les composés que nous analysons ont été recueillis dans un corpus écrit, et les emprunts sont d'ordinaire plus utilisés dans la communication orale que dans la communication écrite.

Ensuite, dans la présente étude (contrairement à Ahronian 2001), 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*, FR *moteur* < EN *engine*) 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 grand public traitant de l'actualité d'Internet qui intéresse tous les internautes. Certaines sont destinées à des profanes (*Web magazine*) et les autres à un public plus expérimenté (*Netsurf*, *Officiel du Net*, *.Net*) comme des webmasters ou des webdesigners. D'après Brolles (2001), ces deux types de magazines présentent les mêmes termes. Le corpus de Brolles (travaux en cours) est constitué de revues grand public et de revues underground traitant de l'actualité du piratage informatique.¹¹⁰ Ces dernières sont moins accessibles aux profanes et visent un public revendiquant plus l'appartenance à une communauté.¹¹¹ Elles regorgent d'emprunts et de sigles (notamment les termes relatifs aux attaques). C'est ce qui explique que les chiffres de l'auteur diffèrent des nôtres (64% d'emprunts à l'anglais, 31% de calques et 5% de transpositions). La différence de pourcentages provient de la nature de nos corpus. Le taux d'emprunts est plus élevé dans les publications spécialisées que dans les ouvrages de vulgarisation. Il n'est pas nécessaire de faciliter la compréhension de lecteurs aguerris maîtrisant les mêmes termes spécialisés.

Il convient maintenant de s'attarder sur les composés hybrides. 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, la performance et l'influence des moyens de communication actuels font que les néologismes anglais sont rapidement introduits et diffusés en français, favorisant ainsi leur implantation (absence de termes français concurrents). Dans une situation de néologie d'urgence, les spécialistes français d'Internet et les néologistes n'ont peut-être pas toujours le temps matériel pour 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*).

¹¹⁰ L'auteur n'a pas retenu les magazines présentant les techniques et algorithmes de piratage électronique.

¹¹¹ De très nombreuses entrées du dictionnaire de la revue *Zataz* sont des emprunts (www.zataz.com/dictionnaire-informatique/).

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*). Les termes pseudo-confixés, très fréquents en français, sont des créations hybrides puisque les pseudo-confixes ont d'abord été utilisés en anglais.

Créer de nombreux termes hybrides permet aux internautes de démarquer leur lexique des lexiques spécialisés soutenus. Cela confirme l'hypothèse avancée précédemment (Ch. 2, Sec. 3.2.2.) : il n'y a pas de frontière langue populaire/langue scientifique dans le domaine d'Internet.

Tableau 27 : Structures compositionnelles endogènes des composés du corpus français

	Structures	Fréquence « type »	%
1	N + de + N	131	28,42%
2	N + N	117	25,38%
3	N + A	102	22,13%
4	PS + N	68	14,75%
5	8 Structures mineures 112	43	9,32%
	Total	461	100,00%

Les composés endogènes sont issus de douze procédés inégalement fréquents, dont quatre sont quantitativement plus importants. Les composés les plus représentés sont : ceux comportant deux noms reliés entre eux, N + de + N (type *carnet d'adresses, notificateur de mail*), ceux constitués de deux substantifs, N + N (type *code source, adresse IP, accès Internet*), les substantifs plus adj ectifs, N + A (type *adresse électronique*), et les composés formés d'un pseudo-confixe et d'un substantif, PS + N (type *cybercafé, télétravail, nétiquette*). Ces quatre patrons représentent 90,68% du total des endogènes. Les huit autres structures ont généré les 9,32% restants.

Comme l'illustre le tableau 27, le français a le plus souvent recours aux formations syntagmatiques : 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 formations syntagmatiques concernent 47,75% des composés (106 noms sur 222) contre 16% dans la langue générale (Béciri, 1999 : 137). Leur présence est bien plus significative que dans le lexique d'Internet. L'hypothèse la plus plausible pour expliquer la fréquence de cette structure est le calque. Comme nous l'avons constaté, l'anglais crée beaucoup par juxtaposition (50,50% de N + N, 18,09% de A + N). Le calque permet de traduire ces créations anglaises au moyen d'équivalents indigènes (FR *outil de recherche* < EN *search tool*, FR, *courrier électronique* < EN *electronic mail*). Le calque est un procédé de traduction assez rapide et facile à utiliser. D'après Candel (2000 : 358), 70 % des composés syntagmatiques contiennent la

¹¹² 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%).

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 depuis la période 1880-1914 (Candel, 2000 : 345). Elles ont tendance à se développer et la terminologie d'Internet semble aller dans ce sens. Humbley (2000 : 80) atteste de la progression des N + N dans le lexique français et cite les travaux de Riegel (1989), Noailly (1990) et Picone (1996) qui en font état. En informatique (Béciri, 1999 : 138), les juxtaposés N + N sont dominants parmi les composés binaires (58%), 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 22,13% l'apposition d'un adjectif à un substantif, 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, selon Tetet (2000 : 552), les constructions N + A sont couramment utilisées pour dénommer les disciplines sportives : *danse acrobatique*, *danse gymnique*, *danse rythmique*, *danse sportive*. Béciri (1999 : 138) comptabilise 28% de N + A dans son corpus spécialisé mais ne parle pas de la langue générale. Kocourek (1991 : 131) ne donne pas d'indication chiffrée. Comme pour l'anglais, les PS + N ont une influence sur les N + A. Les termes construits avec e- (e-courrier, e-tourisme, e-métier) « remplacent » des N + A (*courrier électronique*, *tourisme électronique*, *métier électronique*). Cela peut expliquer leur présence plus faible 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. Au-delà de la langue anglaise, c'est une spécificité de la terminologie d'Internet. En français, les PS + N sont le quatrième patron le plus fréquent avec 14,75% des composés. Les pseudo-confixés représentent au total 20,38% des composés endogènes. Cette construction est 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 en français (*e-mail*, *webmaster*, *e-business*, *freeware*, etc.). De même que pour l'anglais, à notre connaissance aucune étude (excepté Mopoho 1997) ne mentionne la fréquence « type » élevée des pseudo-confixes.

1.1.2. Les surcomposés à trois éléments

Tableau 28 : Structures compositionnelles endogènes et exogènes des surcomposés à trois éléments du corpus français

Les noms composés anglais français et espagnols du domaine d'Internet :

	Structures	Fréquence « type »	%
1	<u>Composés endogènes :</u>	93	94,89%
1.1	<i>purs</i>	59	60,20%
1.1.1	N + de + N + A	14	14,29%
1.1.2	N + de + N + de + N	11	11,22%
1.1.3	N + en + N	7	7,14%
1.1.4	N + à + N + A	5	5,10%
	N + de + PS + N	5	5,10%
1.1.5	N + non + A	3	3,06%
1.1.6	N + de + N + en + N	2	2,04%
	N + A + N	2	2,04%
	N + N + N	2	2,04%
	N + A + A	2	2,04%
1.1.7	N + à + N + N	1	1,02%
	N + A + de + N	1	1,02%
	N + N + A	1	1,02%
	V + N + N	1	1,02%
	N + sans + N	1	1,02%
	N + hors + N	1	1,02%
1.2	<i>hybrides</i>	34	34,69%
1.2.1	N + de + N + N	7	7,14%
1.2.2	N + N + A	5	5,10%
1.2.3	N + PS + A	3	3,06%
1.2.4	N + N + N	2	2,04%
	N + N + par + N	2	2,04%
1.2.5	N + de + N + A	1	1,02%
	PS + N + de + N	1	1,02%
	N + de + N + PS	1	1,02%
	N + à + N + N	1	1,02%
	N + de + N + à + N	1	1,02%
	N + de + N + de + N	1	1,02%
	N + A + N	1	1,02%
	PS + N + N	1	1,02%
	N + PS + N	1	1,02%
	N + A + PS	1	1,02%
	PS + sans + N	1	1,02%
	N + sur + N	1	1,02%
	N + en + N	1	1,02%
	V + N + N	1	1,02%
	N + non + A	1	1,02%
2	<u>Composés exogènes</u>	5	5,10%
	TOTAL	98	100,00%

Les surcomposés de la terminologie française d'Internet sont presque exclusivement

protégé en vertu de la loi du droit d'auteur.

endogènes (94,89%). Les termes exogènes sont moins présents chez les surcomposés que chez les binaires (5,10% contre 14,63%). La fréquence des composés hybrides à trois éléments est de 10 points inférieure à celle des binaires (34,69% contre 44,26%). Les emprunts entrant dans la formation de surcomposés sont : *Internet* (FR *fournisseur d'accès Internet*), *proxy* (FR *pare-feu proxy*), *cracker* (Fr *cracker de mots de passe*), les sigles (FR *boîte aux lettres POP3, adresse IP publique*) et les pseudo-confixes (FR *cyberpunkitude, page Web musicale*). Aucun surcomposé ne contient un emprunt monosyllabique.

Les raisons du faible taux d'emprunts ont été exposées (Part. II, Ch. 3, Sec. 1.1.1.). Seules les différences entre binaires et surcomposés à trois éléments seront maintenant étudiées. Pourquoi la longueur des termes influence-t-elle le nombre de termes hybrides et d'emprunts ?

Les spécialistes d'Internet privilégient l'utilisation des sigles, surtout dans le cas des anglicismes (*WAP, FTP, HTML, POP3*, etc.). Ils empruntent les sigles sans les modifier¹¹³ (leurs formes développées sont citées dans les articles à titre d'information). Mais d'une manière générale, les surcomposés à trois éléments sont moins fréquents en français qu'en anglais (15,11% contre 27,78% du total du corpus anglais).

Pour le grand public, un composé endogène de plusieurs éléments est plus facile à déchiffrer qu'un emprunt long (FR *mise à jour automatique* < EN *live update*, FR *serveur de courrier entrant* < EN *incoming mail server*). Les calques sont donc préférables aux emprunts.

Tableau 29 : Structures compositionnelles des surcomposés endogènes à trois éléments du corpus français

	Structures	Fréquence « type »	%
	2 structures majeures :	27	29,03%
1	N + de + N + A	15	16,13%
2	N + de + N + de + N	12	12,90%
	7 structures moyennes :	39	41,94%
3	N + en + N	8	8,60%
4	N + de + N + N	7	7,53%
5	N + N + A	6	6,45%
7	N + à + N + A	5	5,38%
8	N + de + PS + N	5	5,38%
9	N + non + A	4	4,30%
10	N + N + N	4	4,30%
11	¹¹⁴ 8 structures mineures ¹¹⁵	27	29,03%
	Total	93	100,00%

¹¹³ Dans la terminologie française de l'informatique, 74% des sigles sont anglais (corpus utilisé : *Dictionnaire de l'informatique*, Pierre Morvan dir., 1996, Paris, Larousse, 351 p.) (Walter, 1997 : 46-47).

¹¹⁴

Les quatre-vingt-treize surcomposés endogènes ont été créés par vingt-sept patrons, dont deux sont plus fréquents : N + de + N + A (type *page d'accueil personnelle*, *fournisseur d'applications hébergées*) et N + de + N + de + N (type *site de comparaison de prix*, *système de détection d'intrusion*). À elles deux, ces structures représentent presque un tiers de la totalité des surcomposés. Elles se démarquent des vingt-cinq autres puisque leur pourcentage est bien supérieur à la moyenne générale de 3,70%. Trois autres patrons dépassent la moyenne : N + en + N (type *achat en ligne*, *dialogue en direct*), N + de + N + N (type *logiciel d'accès Internet*, *vitesse de connexion Internet*), N + N + A (type *Internet bande large*, *adresse IP dynamique*).

Les formations comprenant un pseudo-confixe sont majoritaires. Avec quatorze termes elles totalisent 15,05% de l'ensemble de ces surcomposés endogènes. Cela confirme la tendance observée chez les composés et surcomposés anglais et chez les composés français.

La fréquence du premier patron N + de + N + A peut aisément être expliquée. Les composés N + de + N deviennent les bases de nouvelles créations et/ou sont des calques de composés anglais (*gestionnaire de courrier électronique*, *page d'accueil personnelle*, etc.). Cela s'applique également aux composés N + de + N + N (*fournisseur d'accès Internet*, *logiciel de connexion Internet*, etc.).

Les travaux de Humbley (2000), Kocourek (1991), Mitterrand (2000) ne parlent pas des composés longs. En revanche, l'étude du vocabulaire informatique (Béciri, 1999 : 137) fait ressortir une trentaine de surcomposés à trois éléments ou plus sur 222 termes (soit 13,51%, pourcentage similaire au nôtre).

1.1.3. Les surcomposés à quatre et cinq éléments

Tableau 30 : Structures compositionnelles endogènes et exogènes des surcomposés à quatre et cinq éléments du corpus français

¹¹⁵ N+A+N (3,23%), N+PS+A (3,23%), N+A+A (2,15%), V+N+N (2,15%), N+à+N+N (2,15%), N+N+par+N (2,15%), N+de+N+en+N (2,15%), N+hors+N (1,08%), N+sans+N (1,08%), PS+N+de+N (1,08%), N+de+N+PS (1,08%), N+A+de+N (1,08%), N+de+N+à+N (1,08%), PS+N+N (1,08%), N+PS+N (1,08%), N+A+PS (1,08%), PS+sans+N (1,08%), N+sur+N (1,08%).

	Structures	Fréquence « type »	%
1	<u>Composés endogènes :</u>	6	54,54%
1.1	<i>purs</i>	4	36,36%
1.1.1	N + de + N + A + A	1	9,09%
	N + à + N + A + A	1	9,09%
	N + de + N + sans +N	1	9,09%
	N + par + N + A	1	9,09%
1.2	<i>hybrides</i>	2	18,18%
1.2.1	N + PS + A + N	1	9,09%
	N + N + sans + N	1	9,09%
2	<u>Composés exogènes</u>	5	45,45%
	TOTAL	11	100,00%

Les termes endogènes du tableau 30 représentent 54,54% du corpus contre 94,89% chez les surcomposés à trois éléments. Comment expliquer la différence des pourcentages d'endogénéité et d'exogénéité entre les surcomposés à trois éléments et les termes plus longs ?

Pour répondre à cette question, nous nous pencherons sur la nature des surcomposés à quatre éléments et plus et sur les concepts qu'ils désignent. Ces onze termes longs sont tous relatifs à des aspects techniques et non à l'actualité d'Internet comme la majorité des termes analysés. Ils désignent des protocoles (*Transmission Control Protocol/Internet Protocol*), des langages (*HyperText Markup Language*) ou des modes de connexion (*ligne d'abonnement numérique asymétrique*, *connexion à bande large permanente*). Dans les textes, les formes réduites de ces termes techniques sont privilégiées. Cependant, pour informer et former le grand public sur un protocole ou un langage, les journalistes développent les sigles et utilisent donc des emprunts. La proportion de composés exogènes est plus importante que chez les binaires puisque les journalistes ne précisent pas les équivalents anglais des termes français courts dont les concepts ne sont pas désignés par des sigles.

Pour conclure, les termes très longs sont peu représentés dans notre corpus car : a) le côté technique d'Internet n'est pas le thème majeur des revues (qui est l'actualité) et b) les termes longs sont rapidement raccourcis (par troncation et siglaison). En outre, ces surcomposés rapprochent le lexique d'Internet des lexiques plus soutenus. Il est plus délicat de parler d'absence de frontière entre langue populaire et langue scientifique puisque par leur longueur les termes techniques d'Internet se rapprochent du prototype du terme.

Tableau 31 : Structures compositionnelles des surcomposés endogènes à quatre et cinq éléments du corpus français

	Structures	Fréquence « type »	%
1	N + de + N + A + A	1	16,67%
	N + à + N + A + A	1	16,67%
	N + de + N + sans +N	1	16,67%
	N + PS + A + N	1	16,67%
	N + N + sans + N	1	16,67%
	N + par + N + A	1	16,67%
	Total	6	100,00%

Peu de composés forment cette catégorie : les six composés endogènes comptent quatre éléments. Les données dont nous disposons sont insuffisantes pour aboutir à des résultats satisfaisants. La présence des formations avec pseudo-confixe (16,67%) est similaire à celle de surcomposés à trois éléments (15,05%).

À l'inverse des surcomposés à trois éléments, nous avons des proportions de termes similaires en français (1,69%) et en anglais (1,08%). Nous en avons déjà évoqué la raison (Ch. 2, Sec. 1.1.3.). Les surcomposés à trois, quatre, cinq (et six) éléments sont moins nombreux en français qu'en anglais, contrairement aux composés binaires :

Tableau 32 : Répartition des noms composés anglais et français selon le nombre de formants

Nb de formants	2	3	4	5	6	Total
Corpus anglais	397	129	21	5	1	553
Corpus français	540	98	9	2	0	649

Ceci est vraisemblablement dû au fait que le français adapte les raccourcissements et les réductions des surcomposés anglais. Le nombre inférieur de surcomposés en français est compensé par un nombre supérieur de composés. D'autre part, les composés sont plus nombreux en français qu'en anglais parce qu'il existe une concurrence terminologique : plusieurs équivalents français remplacent un terme anglais. Par exemple, à EN *log file* correspondent trois équivalents français : *fichier journal*, *fichier de log*, *log de connexion*.

1.1.4. Bilan

Tableau 33 : Structures compositionnelles des composés et surcomposés français

	Structures	Fréquence « type »	%
1	Composés endogènes :	560	86,29%
	5 structures majeures :	418	64,41%
1.1	N + de + N	131	20,18%
1.2	N + N	117	18,03%
1.3	N + A	102	15,72%
1.4	PS + N	68	10,48%
	3 structures moyennes :	48	7,40%
1.5	N + PS	21	3,24%
1.6	N + de + N + A	15	2,31%
1.7	N + de + N + de + N	12	1,85%
	38 structures mineures :	94	14,48%
1.8	Flud	10	1,54%
1.9	N + en + N	8	1,23%
1.10	N + de + N + N	7	1,08%
1.11	N + N + A	6	0,92%
1.12	N + à + N + A	5	0,77%
	N + de + PS + N	5	0,77%
1.13	N + non + A	4	0,62%
	N + N + N	4	0,62%
15	N + à + N	3	0,46%
	PS + A	3	0,46%
	N + A + N	3	0,46%
	N + PS + A	3	0,46%
16			

	N + de + PS	2	0,31%
	A + N	2	0,31%
	N + à + N + N	2	0,31%
	N + N + par + N	2	0,31%
	N + de + N + en + N	2	0,31%
	V + N + N	2	0,31%
	N + A + A	2	0,31%
	V + N	1	0,31%
	non + N	1	0,15%
	PS + N + de + N	1	0,15%
	N + de + N + PS	1	0,15%
	N + A + de + N	1	0,15%
	N + de + N + à + N	1	0,15%
	PS + N + N	1	0,15%
	N + PS + N	1	0,15%
	N + A + PS	1	0,15%
	PS + sans + N	1	0,15%
	N + sur + N	1	0,15%
	N + sans + N	1	0,15%
	N + hors + N	1	0,15%
	N + de + N + A + A	1	0,15%
	N + à + N + A + A	1	0,15%
	N + de + N + sans + N	1	0,15%
	N + PS + A + N	1	0,15%
	N + N + sans + N	1	0,15%
	N + par + N + A	1	0,15%
2	Composés exogènes :	89	13,71%
	Emp2 ¹¹⁶	79	12,17%
	Emp3	5	0,77%
	Emp4	3	0,46%
	Emp5	2	0,31%
	Total	649	100,00%

Les 649 composés et surcomposés français ont été créés au moyen de quarante-cinq patrons syntaxiques endogènes qui représentent 86,29% du corpus français. Les 13,71% restant sont les termes exogènes (classés selon le nombre de formants). Les patrons syntaxiques utilisés en français sont moins nombreux qu'en anglais (58). Le pourcentage moyen est de 2% ou 12,98 termes par patron. Du nombre total de composés et surcomposés du corpus français, il résulte que :

- Les structures les plus fréquentes sont au nombre de cinq. Quatre sont endogènes : N + de + N (20,18%), N + N (18,03%), N + A (15,72%), PS + N (10,47%), et une seule est exogène : Emp2 (12,17%). Si ce nombre est proche de l'anglais (six), leur

¹¹⁶ Le chiffre indique le nombre d'éléments formant l'emprunt.

fréquence diffère : 82,10% contre 76,57% pour le français (497 composés sur 649). Avec 23,43%, les patrons mineurs sont plus fréquents en français qu'en anglais.

- Comme en anglais, les composés longs sont les moins nombreux. Les 540 termes composés représentent 83,20% du total, les 98 surcomposés à trois éléments comptent pour 15,10% et les 11 surcomposés à quatre et cinq éléments totalisent 1,69%. Les composés les plus longs sont utilisés sous leur forme siglée, notamment dans le cas des emprunts. Les désignations techniques sont longues mais plus le concept est connu, plus le nombre d'utilisateurs augmente et plus les termes sont raccourcis (Candel, 2000 : 345). Cette affirmation s'applique à Internet qui touche un public très large.
- Tout comme pour les composés anglais, les composés avec pseudo-confixes sont un nouveau procédé de néologie formelle. Au total, ils sont 109 et représentent 16,90% du corpus français (27,67% en anglais). Les pseudo-confixes sont présents dans tous les patrons comme en anglais.
- Les emprunts représentent 13,71% du total des composés du corpus français. Ce pourcentage est comparable à celui de la terminologie de l'informatique dans laquelle les anglicismes (sigles + bases anglaises) représentent 16,5% du corpus de Walter (1997 : 51) (272 termes sur 1 649 entrées).

1.2. Les composés pseudo-confixés

1.2.1. cyber-¹¹⁷

Cybernétique, la première lexie contenant le préfixe *cyber-*, a été créée en français (Part. II, Ch. 2, Sec. 1.2.1.). Par son introduction aux États-Unis, elle a acquis un sens nouveau. Le pseudo-confixe *cyber-* a ensuite été emprunté à l'anglais. Selon le *Grand Dictionnaire Terminologique* (GDT), *cyber-* vient du grec mais son sens actuel tire son origine du nom EN *cyberspace*.

Cyber- est un mot-vedette dans certains ouvrages lexicographiques de langue anglaise, en particulier le MWOD et le CALDO, et dans quelques dictionnaires de langue française. Le *Grand Robert de la langue française* 2001 (GR 2001) présente *cyber-* comme un « élément tiré de *cybernétique* et employé dans les composés dans le contexte des réseaux de communications numériques [...] » (2001 : 919). Cet ouvrage répertorie des termes fréquents : *cybercafé*, *cybercriminalité*, *cyberculture*, *cyberespace* et *cybernaute*.

Tableau 34 : Composés français avec le pseudo-confixe *cyber-*

Pseudo-confixe	Base	Champ	Base	Champ
----------------	------	-------	------	-------

¹¹⁷ Dans les tableaux de cette analyse, les termes en majuscules et caractères gras ont un nombre d'occurrences supérieur à 1 000 000, ceux en majuscules ont entre 1 000 000 et 999 999 occurrences, les composés en caractères gras ont entre 10 000 et 99 999 occurrences, ceux en italique entre 1 000 et 999. La fréquence des composés des tableaux provient de www.google.fr (option sites francophones uniquement).

Les noms composés anglais français et espagnols du domaine d'Internet :

		sémantique		sémantique
cyber	amour	Sé	flic	S
	civisme	Sé	flirt	Sé
	univers	Sé	frontière	Sé
	acheteur	C	guerre	S
	activité	Sé	langue	Sé
	album	Sé	marché	C
	annuaire	Sé	marque	C
	attaque	S	<i>monde</i>	Sé
	<i>boutique</i>	C	piraterie	S
	café	Sé	plébiscite	Sé
	caméra	Sé	protection	S
	commerçant	C	punk	S
	<i>conférence</i>	Sé	punkitude	S
	conflit	S	recrutement	Sé
	<i>consommateur</i>	C	sexé	Sé
	<i>crime</i>	S	soldat	S
	criminalité	S	squatter	S
	criminel	S	<i>squatting</i>	S
	débat	Sé	cyberstress	Sé
	délinquance	S	supermarché	C
	délinquant	S	<i>surveillance</i>	S
	droit	S	<i>terrorisme</i>	S
	éditeur	Sé	terroriste	S
	épicier	C	visiteur	Sé
	espace	Sé		

Légende : C = commerce, S = sécurité, Sé = société

Le corpus français contient quarante-neuf termes avec *cyber-*. Les composés français sont semblables, morphologiquement et sémantiquement, aux composés anglais.

Cyber- :

- a une fonction préfixale et l'ordre des éléments est déterminant/déterminé. Ces composés ne respectent donc pas les règles de formation des composés français. L'ordre déterminant/déterminé est dû à l'influence de l'anglais. Le succès de la lexie EN *cyberespace* a donné naissance à un modèle dont la fréquence « type » est élevée, *cyber- + N*. Ce patron syntaxique est identique au modèle de formation classique utilisé dans la création de termes comme *cybernétique*.
- est ajouté à un mot existant pour transposer la réalité du concept dans le cyberespace. Les termes désignent des activités du monde réel reproduites sur Internet.
- indique un lieu, une chose, une personne qui existe dans le cyberespace, dans Internet ou qui y est associé. Les termes se répartissent dans trois champs sémantiques : le commerce, la société et la sécurité.

1.2.2. hyper-

Dans les ouvrages lexicographiques français consultés, aucune mention d'une nouvelle utilisation de *hyper-* n'est faite. Il a pourtant acquis un nouveau sens avec la création de EN *hypertext*. *Hyper-* est un élément grec exprimant (PR 1996) :

l'exagération, l'excès, le plus haut degré, soit dans des composés empruntés au grec ou au latin, soit dans des formations françaises, savantes ou familières. # super- . REM. Hyper- sert à former des noms et des adjectifs, des mots familiers : c'est *hyper-sympa*, *hyper-chouette*.

Le GR 2001 donne la même définition de *hyper-* et inclut le nom *hypertexte* et son dérivé adjetival, *hypertextuel*, qui sont présentés comme des calques de l'anglais.

Tableau 35 : Composés français avec le pseudo-confixe *hyper-*

Pseudo-confixe	Base	Champ sémantique	Base	Pseudo-confixe	Champ sémantique
hyper	lien	P		Surcomposés avec :	
			lien	hypertexte	P
			<i>HyperText Markup Language</i>		P

Légende : P = procédé

Le corpus français contient la même proportion de termes créés avec *hyper-* que le corpus anglais. Ce nouveau formant semble vraiment restreint à un champ notionnel précis : tout ce qui est lié à l'hypertexte, c'est-à-dire à un moyen de navigation dans un ensemble d'informations grâce à la présence de liens activables dans les documents.

Hyper- a une fonction préfixale. Il est le déterminant de composés qui sont tous endocentriques. Le GDT répertorie d'autres termes : *hypermédia*, *hypertextuel*, *hypertoile*.¹¹⁸ Dans ces termes appartenant au même champ notionnel, le sens du pseudo-confixe est différent de celui donné par le PR et le GR. Dans ces trois termes, *hyper-* est la troncation de EN *hypertext*. Grâce à la popularité de l'hypertexte, en français aussi, la structure *hyper- + déterminé* est devenue fréquente. Dans la lexie *hypertoile*, l'élément de nomination *hyper-* reprend le trait conceptuel *hypertexte* (l'hypertoile est basée sur l'hypertexte) et l'élément de nomination *toile* reprend le concept Internet. L'hypertoile utilise l'hypertexte pour naviguer sur Internet.

1.2.3. télé-

Le patron syntaxique *télé- + déterminé* a une fréquence « type » plus élevée en français qu'en anglais où il n'a produit qu'un seul terme (*teleworking*). Le PR 1996 contient trois entrées pour cet élément :

¹¹⁸ « Système basé sur l'utilisation de l'hypertexte, qui permet la recherche d'information dans Internet, l'accès à cette information et sa visualisation. Synonyme de WWW » (GDT).

télé-

- Élément, du grec *télos* « loin », signifiant « au loin, à distance » : *télévision, téléphone, télégramme*.
- Élément, de *télévision*, signifiant « de télévision, par télévision » : *télédiffusion, téléfilm, téléviseur*.
- Élément, de *telephonique* : *télécabine, télésiège*.

D'après la deuxième acceptation, le préfixe grec avait acquis un statut de pseudo-confixe avant son utilisation dans le lexique d'Internet. Mais dans la terminologie d'Internet, le pseudo-confixetélé- est la troncation d'un autre terme : *télématique*.¹¹⁹ Par exemple, *téléchargement*, dont la date d'apparition est 1985 selon le PR 1996, signifie « introduction d'un fichier dans la mémoire d'un ordinateur » (= *chargement*) « au moyen des techniques de la télématique » (= *télé-*). *Télé-* pourrait éventuellement être la troncation de *télécommunication* (GR 2001), mais la télématique étant un procédé plus récent et complet, il est plus probable que *télé-* soit la réduction de ce dernier. Le terme *télétravail*, créé en 1981, est défini comme une « activité professionnelle exercée hors de l'entreprise (notamment à domicile) grâce à la télématique » (PR 1996). Ce composé est sans doute formé de la troncation antérieure de *télématique* et de la *lexietravail*.

Avec l'ouverture d'Internet au grand public, le terme *téléchargement* a vu sa fréquence s'accroître considérablement. D'ailleurs, Newton définit le téléchargement principalement par rapport à Internet. Le succès du téléchargement a renforcé la fréquence « type » du modèle *télé- + N* qui existait en français avant Internet. Le PR 1996 recense de nombreuses lexies dont certaines ont un sens voisin de celui qui nous intéresse comme *télépaiement* (1986), *télésurveillance* (1968), *téléconférence* (1982).

Tableau 36 : Composés français avec le pseudo-confixetélé-

Pseudo-confixe	Base	Champ sémantique	Base	Pseudo-confixe	Champ sémantique
télé	CHARGEMENT	T		Surcomposés avec:	
	<i>formation</i>	Mt	site de	téléchargement	T
	<i>maintenance</i>	S	assistant de		T
	travail	Mt	temps de		T
	<i>travailleur</i>	Mt	<i>vitesse de</i>		T
			<i>gestionnaire de</i>		T

Légende : T = téléchargement, Mt = monde du travail, S= sécurité

Dans les composés du tableau 36, le pseudo-confixetélé- a une position préfixale. Ces composés sont endocentriques et *télé-* est l'élément déterminant. Un *téléchargement* et un *télétravailleur* sont des sous-types de *chargement* et de *travailleur*. Dans les surcomposés, tous créés à partir de *téléchargement*, cet élément est aussi le déterminant.

119

« La télématique désigne le mariage des techniques de télécommunications et de l'informatique [...] » (Newton, 2000 : 640). Ce dernier terme est un amalgame créé en 1977 à partir de *télé-* et *informatique* (PR 1996).

Un *site de téléchargement* et un *assistant de téléchargement* sont des *sites* et des *assistants* d'un type particulier.

Ce formant, troncation de *télématique*, acquiert une valeur positive due à l'engouement pour le téléchargement qui est rattaché à Internet. De multiples activités sont réalisables sans déplacement. *Télé-* modifie l'acception de lexies antérieures à Internet pour les adapter aux nouvelles technologies. Grâce à Internet, on peut suivre une formation (*téléformation*), faire de la maintenance (*télémaintenance*), travailler (*télétravail*), récupérer des fichiers (*téléchargement*) sans se déplacer.

Un certain flou conceptuel entoure le terme français *téléchargement*. Il est équivalent de l'anglais *download* mais selon les définitions, les traits conceptuels ne sont pas toujours les mêmes. Le GDT définit le téléchargement comme une « opération qui consiste soit à transférer des données ou des programmes stockés dans un ordinateur distant vers un ordinateur local à travers un réseau, soit à transférer des données ou des programmes stockés dans un ordinateur central vers un micro-ordinateur ». Pour la DGLF, un téléchargement est une « copie sur un ordinateur personnel de fichiers (logiciels, images, etc.) se trouvant sur le réseau ». En anglais d'après Netlingo, entre autres, on parle de *download* lorsqu'on récupère des données et de *upload* lorsqu'on en transmet.

1.2.4. métा-

Le PR 1996 présente *méta-* comme un élément grec :

exprimant la succession, le changement, la participation, et en philosophie et dans les sciences humaines « ce qui dépasse, englobe » (un objet, une science) : métalangue, métamathématique.

Quatre acceptations sont mentionnées, comme dans le MWOD pour l'anglais, et trois sont communes aux deux langues : la succession, le changement et le dépassement. Le PR 1996 ne mentionne pas l'acception dans le domaine de la physique.

Tableau 37 : Composés français avec le pseudo-confixe *meta-*

Base	Pseudo-confixe	Champ sémantique	Pseudo-confixe	Base	Champ sémantique
balise	méta	P	meta	tag	P
				balise	P
				moteur	A
				portail	A

Légende : A = application, P = procédé

Dans quatre des cinq composés du corpus français, *méta-* a une position préfixale et il est le déterminant. Seul le composé *balise méta*, calque de *meta tag*, répond à l'ordre déterminé/déterminant du français. En ce qui concerne les termes FR *méta balise*, *méta-moteur* et *méta portail*, ils sont issus d'un modèle, *méta- + N*, né du succès de l'emprunt *meta tag*.

Ce pseudo-confixe, employé en français dans le même champ notionnel qu'en

anglais, est probablement la troncation de EN *metaelement* (FR *méta-élément*) (Part. II, Ch. 2, Sec. 1.2.3.). Dans les nouvelles créations, le français utilise la structure des composés anglais (*meta-* + N) et traduit le déterminé : *tag* > *balise*, *engine* > *moteur*, *portal* > *portail*. D'autres termes, non inclus dans notre corpus, sont formés sur ce modèle : FR *méta-données* (EN *metadata*), *méta-base de données*, *méta recherche* (EN *metasearch*), *méta référencement* (EN *metaregistration/referencing*), *méta-outil* (EN *metatool*).

1.2.5. web-, -web

Web est répertorié dans le GR 2001 comme une abréviation de l'anglais. Aucune mention n'est faite de son utilisation pour former des lexies :

Abrév. De l'angl. World Wide Web, proprt ‘toile (d’araignée) mondiale’. Anglic. Réseau informatique mondial, réunissant les millions de ressources, de pages consultables (cf. Internet, toile). (GR 2001 : 1985)

Le succès et la fréquence « type » élevée du pseudo-confixe anglais sont les motifs de l'emprunt par le français pour former des néologismes.

Tableau 38 : Composés français avec le pseudo-confixe -web-

Base	Pseudo-confixe	Champ sémantique	Pseudo-confixe	Base	Champ sémantique
accès	Web	C	Web	agency	Mt
administrateur		Mt		bar	Sé
adresse		N		bug	S
ANNUAIRE		N		CAM	Co
aspirateur		A		center	Mt
attaque		S		collaborating	Mt
communauté du		Sé		designer	Mt
concepteur		Mt		<i>invisible</i>	N
design		Mt		LOG	Co
designer		Mt		logger	Co
développeur		Mt		<i>logging</i>	Co
ergonome		Mt		MAIL	Co
espace		N		marketeur	Mt
fournisseur d'hébergement		N		MASTER	Mt
gestionnaire		A		MESTRE	Mt
interface		A		mobile	N
interface graphique		A		RADIO	Co
logiciel		A		remixeur	Co
page		N		sans fil	C
serveur		C		<i>visible</i>	N
site		N		(MAGA)ZINE	Co
site webzine		N			
	Surcomposés:				
accès Web haut-débit		C			
page web statique		N			
site Web musical		N			

Légende : A = application, C = connexion, Mt = monde du travail, N = navigation, S = sécurité, Sé = société

Le tableau 38 met en valeur trois structures morphosyntaxiques majeures : N + -web (22 composés), web- + N (17 composés) et web- + A (4 composés). Le modèle le plus fréquent, N + -web, forme des composés répondant à la structure déterminé/déterminant du français (*accès Web*, *ergonome Web*, etc.). Les composés formés sur la structure déterminant/déterminé web- + N sont des emprunts anglais (*web agency*, *weblog*).

Sur le plan sémantique, les composés formés d'après les modèles web- + N et N + -web sont endocentriques : un *accès web*, une *page web*, un *webmaster* et un *web bar* sont des types d'accès, de *page*, de *master* et de *bar*. Quant aux composés contenant un adjectif, ils ne sont pas tous identiques. *Web invisible* et *Web visible* désignent des parties

du *Web*. Ils sont endocentriques. En revanche, *Web mobile* et *Web sans fil* (ou *Web nomade*) sont exocentriques. Ils désignent une technique d'accès à Internet grâce à un terminal sans fil. On pourrait dire un accès *web mobile*, etc.

Les composés français sont généralement des calques des composés anglais *web access* > accès *web*, *web site* > site *web*, *web directory* > annuaire *web*, etc. Aussi, les champs notionnels relevés dans le corpus français sont-ils les mêmes qu'en anglais.

1.2.6. net-, -net

Dans le GR 2001, *net* est un nom venant de l'anglais *net* signifiant 'réseau' dans *Internet*. Il est défini comme un anglicisme désignant le réseau Internet. Le terme dérivé *nétiquette* est répertorié et présenté comme un emprunt à l'anglais *netiquette*, un mot valise formé de *net(work)* signifiant 'réseau' et *étiquette* au sens de convenances, bonnes manières pris au français (2001 : 1884).

Tableau 39 : Composés français avec le pseudo-confixe -net

Base	Pseudo-confixe	Catégorie	Pseudo-confixe	Base	Catégorie
citoyen du	Net	N	Net	ÉCONOMIE	Sé
				(ET)IQUETTE	Sé
				marketeur	Mt
				MARKETING	Mt
				surfer	N
				surfeur	N

Légende : Mt = monde du travail, N = navigation, Sé = société

Les sept composés du tableau 39 sont issus de deux modèles : N + de + -net et net- + N. Le modèle le plus fréquent, *net-* + N, est issu du modèle anglais. Sur les six composés de ce type, deux sont des emprunts (*netsurfer* et *net marketing*) et quatre sont des calques (*net-économie* > *net-economy*, *nétiquette* > *netiquette*, *net marketeur* > *net marketer*, *net surfeur* > *net surfer*).

Le pseudo-confixe occupe la position de déterminant excepté dans le composé *citoyen du Net*, seul terme répondant au modèle nom + de + -net. *Net-économie* est un hyponyme d'*économie* et *net surfeur* est un hyponyme de *surfeur*. Les thèmes français sont identiques aux thèmes anglais.

1.2.7. e-

Le pseudo-confixe *e-* a été emprunté à l'anglais. Le GR 2001 le définit comme un élément tiré de l'anglais *electronic* situé « en haut dans la composition de mots en rapport avec une activité sur le réseau mondial de communication avec Internet. [...] » (2001 : 1753). Quelques exemples sont cités : *e-mail*, *e-commerce*, *e-publicité* ou *e-pub*, *e-book* et *e-médecine*.

Tableau 40 : Composés français avec le pseudo-confixe *e-*

Pseudo-composé	Base	Champ sémantique	Base	Pseudo-suffixe	Champ sémantique
				<i>Surcomposés avec:</i>	
e	<i>achat</i>	C	ADRESSE	email	
	<i>biz</i>	C	alerte		S
	<i>book</i>	C	chiffrement		S
	BUSINESS	C	<i>client</i>		P ?
	<i>card</i>	C	compte		M
	COMMERCE	C	filtrage		S
	<i>courrier</i>	M	fournisseur		M
	<i>économie</i>	C	<i>newsletter</i>		M
	<i>employé</i>	Mt	virus		S
	LEARNING	Mt	intégrateur	e-business	C
	MAIL	M	programme de	e-commerce	C
	métier	Mt	<i>site d'</i>		C
	picerie	C	site	e-zine	Sé
	place de marché	C			
	<i>tourisme</i>	C			
	(e)xagone	Sé			
Dérivés de :					
e-mail	chiffré	S			
	crypté	S			
	non sollicité	S			
Légende : A = application, C = commerce, M = courrier, Mt = monde du travail, S = sécurité, Sé = société					

E- est un des pseudo-suffixes les plus fréquents de notre corpus français. Il a formé 32 termes : 16 composés et 16 surcomposés. Il a toujours une fonction préfixale. L'ordre des éléments est déterminant/déterminé, ordre inverse des constructions françaises. La structure *electronic* + N a été raccourcie en *e-* + N. Le français a conservé cette dernière structure. En revanche, les surcomposés présentent l'ordre déterminé/déterminant (13 termes sur 16) et sont endocentriques. Ce sont des calques des surcomposés anglais : *adresse e-mail* > *email address*, *intégrateur e-business* > *e-business integrator*, etc.

Sur le plan sémantique, *e-* signifie ‘consultable, réalisable sur le réseau Internet’. Les termes contenant *e-* désignent des activités que l'on peut faire de chez soi. Les surcomposés français, qui sont des calques, présentent la même particularité que les surcomposés anglais : ils dénomment des concepts plus spécifiques que les composés, c'est-à-dire des personnes ou des logiciels qui accomplissent des tâches pour l'internaute : *intégrateur e-business*, *filtrage e-mail*, etc.

E- peut être remplacé par *électronique*, l'équivalent de sa forme développée. Les

composés respectent alors l'ordre déterminé/déterminant puisque la structure N + *électronique* remplace le modèle e- + N : *achat électronique*, *métier électronique*, etc. La substitution par la forme développée est plus complexe pour les surcomposés. Si l'on utilise *électronique* et e-, *adresse e-mail* devient *adresse de courrier électronique*, *programme de e-commerce* devient un *programme de commerce électronique*. La structure N (+ de) + PS +N devient N + de + N + A.

1.2.8. -naute

Dans le domaine d'Internet, *-naute* entre dans la composition de termes en tant que troncation d'*astronaute*. Le premier terme formé avec ce pseudo-confixe est *internaute*¹²⁰, le calque de l'anglais *internaut* créé à partir d'*Internet* et d'*astronaut* (Newton, 2000 : 426). *-Naute* prend alors le sens d'"utilisateur". Dans la définition du PR 1996, c'est un élément grec signifiant « relatif à la navigation ». Le GR 2001 donne les mêmes renseignements et répertorie *internaute*. Le pseudo-confixe *-naute* signifie 'utilisateur' mais il ne quitte pas le domaine de la navigation puisque la métaphore de la navigation maritime est fréquente dans le lexique d'Internet.

Tableau 41 : Composés français avec le pseudo-confixe *-naute*

Base	PS	Champ sémantique
entrepre	naute	U
inter		U
wappa		U

Légende : U = utilisateur

-Naute est un des rares pseudo-confixes à avoir une fonction suffixale. Dans la plupart des composés, il est le déterminé : un *internaute* et un *wappanaute* sont des utilisateurs. L'*internaute* est un utilisateur d'Internet. Il « surfe sur le Web et voyage dans le cyberespace » (Colombain, 1998 : 109). Le *wappanaute* utilise le WAP pour surfer sur le Web et voyager dans le cyberespace grâce à une application de poche, un téléphone mobile, par exemple. En revanche, dans *entreprenaute* (*entrepre neur + astro naute*) *-naute* semble être le déterminant. Un *entreprenaute* est un entrepreneur d'Internet créant et gérant une société réalisée par et pour le Réseau.

Ces exemples font ressortir trois particularités. Premièrement, *-naute* signifie 'utilisateur' et, par métaphore, conserve le sens de sa forme développée 'qui navigue'. Cependant, il ne peut pas être remplacé par *astronaute* qui signifie 'qui navigue dans les astres'. Deuxièmement, les patrons syntaxiques sont plus innovants que pour les autres pseudo-confixes : sigle + *-naute*, N tronqué + *-naute*. Troisièmement, une voyelle ou syllabe de liaison est parfois utilisée (*wapernaute* et *waponaute*). On a créé *wappanaute* alors que **wapnaute* ne pose aucun problème orthographique ou de prononciation. D'autres lexies ont été créées sur le modèle nom tronqué + *-naute* : *infonaute* (synonyme d'*internaute*), *ficonaute* (entreprenaute obsédé par l'argent), *intranaute* (utilisateur d'un

¹²⁰ Le GR (2001 : 295) donne 1995 comme date d'apparition d'*internaute*.

intranet), *usenaut* (utilisateur de l'*Usenet*), *cybernaute* (utilisateur du cyberespace), etc.

1.2.9. -ware

-Ware est la troncation du composé EN *software* et en anglais, dans le domaine d'Internet, son sens est différent de celui du formant signifiant 'produit, instrument'.

Tableau 42 : Composés français avec le pseudo-confixe -ware

Base	Pseudo-confixe	Champ sémantique
ad	ware	S
free		A
group		A
open		A
share		A
spy		S
Surcomposé avec :		
logiciel	spyware	A/S

Légende : A = application, A/S = sécurité

Tous les composés du tableau 42 sont des emprunts anglais. Seul le surcomposé est une création hybride. Le pseudo-confixe -ware a un équivalent français dont la fréquence « type » est bien inférieure : -*iciel*, la troncation postérieure de *logiciel*. Un seul terme de notre corpus est construit à partir de -*iciel* : *espiongiciel*. Les termes français correspondant aux termes anglais N + -ware ou A + -ware contiennent le nom *logiciel* et non le pseudo-confixe -*iciel* alors que le modèle N + -*iciel* existe. Il est un calque de la structure anglaise : EN *adware* > FR *logiciel anti popup* et *logiciel espion*, EN *freeware* > FR *logiciel gratuit*, EN *shareware* > FR *logiciel partagé*, EN *openware* > FR *logiciel libre*, EN *spyware* > FR *logiciel spyware*. Cependant, *shareware* a comme équivalents français *partagiciel*, *collecticiel* ou *distributiel*, *spyware* > *espiongiciel*, *groupware* > *synergiciel*, *fatware* ou *bloatware* > *obésiciel*, *charityware* > *caritaticiel*, *freeware* > *graticiel*, *gratuiciel*, etc. Plusieurs hypothèses peuvent expliquer la faible implantation de -*iciel* dans le corpus français :

- La quasi-absence de termes en -*iciel* est une particularité de notre corpus. Peut-être les revues sélectionnées ont-elles tendance à employer une paraphrase ou un terme transparent pour expliquer un terme anglais, d'où la présence dans notre corpus de termes composés comme *logiciel partagé*, *logiciel gratuit* et surcomposés comme *firewall logiciel*, *robot logiciel*, *logiciel d'accès Internet*, *logiciel de chat*, *logiciel de chiffrement*, *logiciel de connexion Internet*, *logiciel de courrier*, *logiciel de cryptage de courrier*, *logiciel de filtrage*, *logiciel de navigation*, *logiciel de news*, *logiciel de serveur*, *logiciel d'hébergement*, *logiciel web*. Cependant, les termes *partagiciel*, *gratuiciel* et *obésiciel* ne semblent pas plus opaques que les termes anglais.
- La forte présence de -*iciel* est une particularité du corpus de Mopoho (1998). D'après

l'auteur, la popularité de *-iciel* est due à l'intensification de l'activité de programmation et de la multiplication de ses allomorphes :*-iel*, *-giciel*.

- Le patron N tronqué + *-iciel* produit des termes dont la fréquence « token » est faible. Le succès des composés anglais est tel que les internautes français les adoptent : *adware*, *freeware*, *groupware*, *openware*, *shareware*, *spyware* sont très fréquemment employés. Ils n'éprouvent pas le besoin de créer des lexies françaises. *-Ware* est peut-être transparent pour la majorité des internautes français.
- Les pseudo-confixes utilisés dans la formation de termes français sont tous d'origine anglaise (*web-*, *net-*, *e-*, *-ware*). Excepté *-ware*, aucun de ces formants n'a été remplacé par un équivalent français. Les instances normatives se sont heurtées à la réticence des internautes en créant des termes avec *-iciel* pour franciser le vocabulaire d'Internet.

1.2.10. Conclusion

Le corpus français compte neuf pseudo-confixes fréquents (*cyber-*, *e-*, *hyper-*, *méta-*, *net-*, *télé-*, *web-*, *-naute*, *-ware*) et le corpus anglais en contient sept (*cyber-*, *e-*, *hyper-*, *meta-*, *net-*, *web-*, *-ware*). Le français recense un pseudo-confixe supplémentaire si on considère *-iciel* et l'anglais en compte deux supplémentaires avec *tele-* et *-naut* qui sont très peu fréquents dans nos corpus. Les pseudo-confixes utilisés en français sont incontestablement des emprunts à l'anglais. Certains sont homographes de confixes classiques (*cyber-*, *e-*, *hyper-*, *méta-*, *télé-*, *-naute*). Le français est passif. Il réemploie les formants anglais et ne crée pas de nouveaux pseudo-confixes. Le seul pseudo-confixe français recensé, *-iciel*, a des problèmes d'implantation. Il est supplanté par *-ware* pour lequel il est une alternative. L'anglais étant la langue source du domaine, les internautes préfèrent les termes à consonance anglaise.

En français, comme en anglais, la fréquence « type » change considérablement d'un pseudo-confixe à un autre :

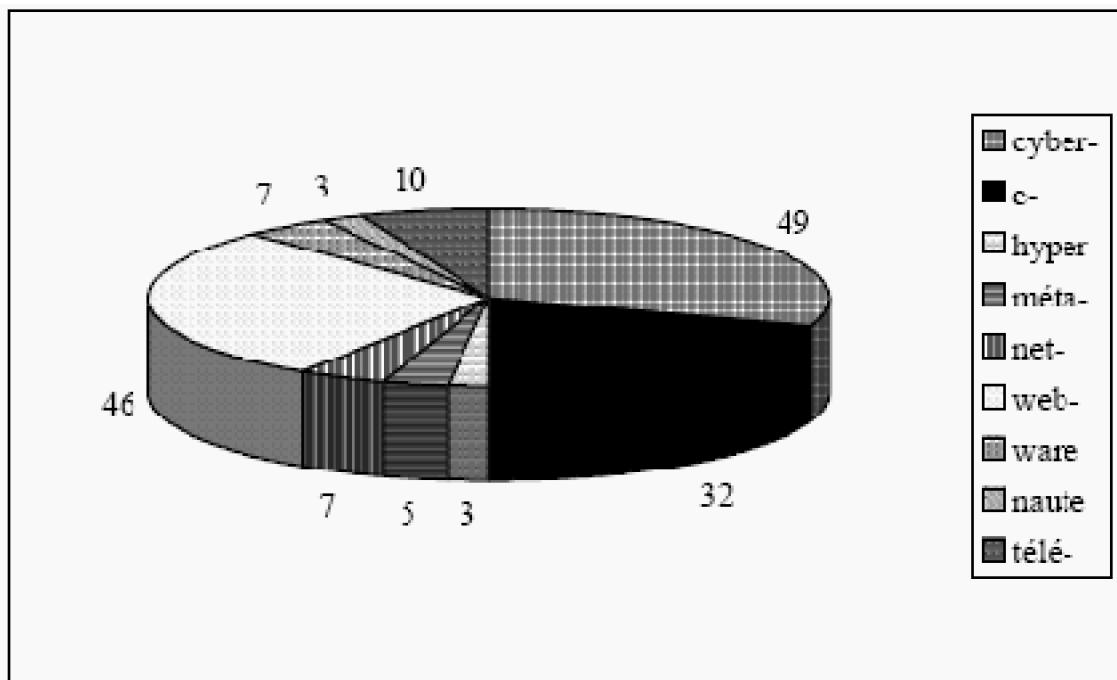


Figure 3 : Fréquence des pseudo-confîxes français

Les pseudo-confîxes les plus fréquents en anglais sont aussi les plus fréquents en français. Avec 49 composés, *cyber-* est le pseudo-confîxe le plus représenté dans le corpus français. Il est suivi de *web-*, présent dans 46 composés. L'écart se creuse avec *e-*, le troisième formant le plus fréquent avec 32 termes. Les pseudo-confîxes *télé-*, *net-* et *-ware*, *métâ-*, *hyper-* et *-naute* arrivent ensuite avec respectivement 10, 7, 5 et 3 composés. L'écart de fréquence entre les pseudo-confîxes résulte de leur champ d'application. Comme pour l'anglais, les pseudo-confîxes sont employés dans des cas précis et certains champs notionnels sont beaucoup plus vastes (*cyber-*, *e-*, *net-*) que d'autres (*hyper-*, *métâ-*), ce qui explique qu'un plus grand nombre de termes soit recensé.

Sur la plan morphosyntaxique, les remarques sont les mêmes que pour l'anglais. Les pseudo-confîxes antérieurs étant les plus nombreux et les plus fréquents, la très grande majorité des composés pseudo-confîxiés sont contraires à la structure française déterminé/déterminant. L'influence de l'anglais sur l'ordre des éléments se ressent surtout dans le cas des pseudo-confîxes clairement identifiables comme des troncations de lexies anglaises (*e-* ou *web-*, par exemple). Les pseudo-confîxes sémantiques sont des homographes de formants classiques. L'ordre déterminant/déterminé n'est pas dû à l'anglais, mais c'est sous son influence que ces pseudo-confîxes sont utilisés en français.

Enfin, nous avons constaté que l'inclusion des pseudo-confîxes dans les ouvrages lexicographiques français est loin d'être systématique. Hormis le GR 2001 et le GDT, les dictionnaires et encyclopédies de langue française récents consultés (en ligne ou version papier) répertorient très rarement les pseudo-confîxes et très peu de termes formés à partir de pseudo-confîxes. Pourtant, certains sont fermement implantés.¹²¹ Plusieurs

¹²¹ Pour FR *cyber-*, *hyper-*, *-naute* et *web-* et leurs dérivés, l'encyclopédie Larousse (www.encyclopedie-larousse.fr/) a comme entrées : *cyber-*, *cybercafé*, *cybermonde*, *cyberespace*, *cybernaute*, *hyperlien*, *hypermédia*, *internaute* et *web*. *E-* est absent.

hypothèses sont envisageables pour expliquer le retard de la lexicographie française sur la lexicographie anglaise dans le domaine d'Internet :

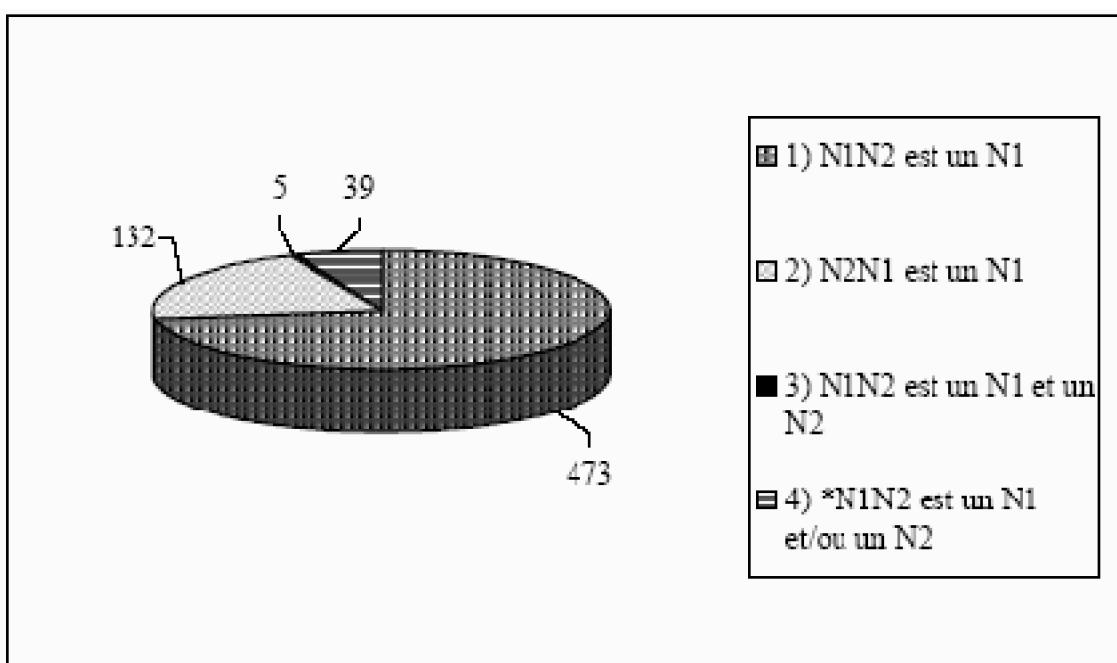
- Les termes sont trop techniques et figurent uniquement dans des dictionnaires spécialisés. Peuvent être concernés des termes créés à partir de *méta-*, *hyper-* mais pas à partir de *cyber-* ou de *web-/web* qui sont massivement utilisés. *Cybersex*, *cyberpunk*, *cyberterrorisme*, *adresse web*, *site web* désignent des concepts souvent utilisés sur Internet.
- Les termes sont trop récents et apparaîtront dans des éditions ultérieures. Les termes sont tirés de revues récentes (2001) et le GR consulté date de 2001. Le GDT contient plus de néologismes car il s'agit d'une ressource électronique et que les mises à jours sont plus aisées et plus fréquentes.
- Les termes sont trop « anglais », ce qui constitue un frein à leur introduction. C'est évidemment le cas des emprunts morphosémantiques mais aussi de termes comme *net surfeur*, *adresse e-mail*, *e-tourisme* (*e-* est prononcé /i/ et non /e/ ou /D/).

2. Le profil sémantique des noms composés collectés

Cette section doit permettre de découvrir l'aspect sémantique des composés du lexique d'Internet du corpus français. Elle permettra peut être de trouver une raison aux exceptions du système d'équivalences types.

2.1. La centricité

La démarche adoptée pour l'étude du corpus français est celle utilisée dans le cadre de l'anglais (Part. II, Ch. 2, Sec. 2.1.).



*Figure 4 : Centricité des composés français du corpus*¹²²

D'après la figure 4, en français comme en anglais, les composés endocentriques sont majoritaires (94% contre 94,57% en anglais). La proportion de termes exocentriques est également très proche (6% en français et 5,53% pour l'anglais).

Comme nous l'avons signalé (Part. II, Ch. 2, Sec. 2.1.), le fait que les composés endocentriques sont motivés contribue à la prédominance de l'endocentricité. La présence de nombreuses métaphores est aussi un facteur non négligeable. Par exemple, hors contexte, les termes *attaque* et *virus* ne sont pas motivés mais dans le lexique d'Internet, la composition *attaque de virus* l'est. Une fois l'arbitraire des deux unités lexicales accepté, une raison de second degré justifie la création de cet assemblage.

Les composés endocentriques les plus nombreux sont de type 'N1N2 est un N1' (72,89%). De fait, le nombre de composés endocentriques 'N2N1 est un N1' paraît élevé (20,34%), surtout par rapport à d'autres terminologies. La présence de nombreux emprunts en est l'explication. Ces composés présentent l'ordre déterminant/déterminé normalement utilisé dans les constructions anglaises. Les composés pseudo-confixés amplifient le phénomène car la plupart des pseudo-confixes ont une position préfixale (*e-courrier*, *net-économie*, *cyber-conférence*, etc.).

Les composés endocentriques 'N1N2 est un N1 et un N2' dont les éléments ont un poids identique (ou composés additifs) sont aussi peu représentés en français qu'en anglais (moins de 1%).

Les composés exocentriques sont, comme en anglais, des termes empruntés à d'autres domaines (*cheval de Troie*, *cookie dans cookie permanent* par exemple, etc.). Comme cela a été envisagé, les composés exocentriques seront un obstacle au système d'équivalences types anglais-français. Pour traduire les termes anglais exocentriques, le français ajoute, sous influence de l'anglais, un sens à une lexie existante (EN *Trojan Horse* > FR *cheval de Troie*) ou adopte un emprunt externe partiel ou total (EN *temporary cookie* > FR *cookie temporaire*, *bookmark* > FR *bookmark*).

2.2. Les relations sémantiques

La démarche de l'étude des relations sémantiques unissant le déterminé (N1) et le déterminant (N2) des noms composés anglais sera reprise. Pour garantir l'homogénéité de l'analyse trilingue, nous réutiliserons les paraphrases élaborées pour les relations sémantiques des composés anglais, en supprimant et ajoutant les relations nécessaires. Si des composés métaphoriques présentent des ambiguïtés catégorielles, nous déterminerons les traits conceptuels les plus importants dans le domaine d'Internet¹²³ pour leur attribuer une catégorie.

Tableau 43 : Synthèse des relations sémantiques des composés français du corpus

Code	Description	Exemples	Fréquence
------	-------------	----------	-----------

¹²² La typologie complète se trouve en Annexe (Annexe 9).

¹²³ Voir Part. II Ch. 2, Sec. 2.2.

			« type »
1	N1 est désigné par analogie à une propriété de N2	<i>courrier escargot, serveur mandataire</i>	7
2	N1 s'adresse à N2 (e-commerce, échange)	<i>peer-to-peer, business-to-business</i>	5
3a	N1 (logiciel, matériel, technique) a pour activité, fonction N2	<i>annuaire de sites, générateur de virus</i>	110
3b	N1 (humain, entreprise) a pour object d'activité, fonction N2	<i>concepteur web, fournisseur de courrier</i>	39
3c	N1 (virus, attaque, pirate) a pour fonction N2	<i>pirate informatique, attaque DoS</i>	14
4	N1 a un fonctionnement dont le procédé, le mode est N2	<i>connexion dédiée, accès haut-débit</i>	118
5	*N1N2 est un N1 et/ou un N2	<i>porte de service, jeune pousse</i>	31
6	N1 se distingue des autres N1 car il est N2 (nature)	<i>lien consulté, e-mail encrypté</i>	99
7	N1 a N2 comme localisation, environnement	<i>cyberattaque, cyberépicier</i>	77
8	N1 est combiné à N2	<i>entreprenaute, apprenti pirate</i>	6
9	N1 est engendré, provoqué par N2	<i>habitude de navigation, infection virale</i>	13
10	N1 est constitué, formé par N2	<i>communauté hacker, réseau des réseaux</i>	7
11	N1 évalue, mesure N2 (durée, vitesse, nombre)	<i>session de chat, temps de téléchargement</i>	8
12	N1 est transféré en direction N2	<i>download, upload</i>	2
13a	N1 est un instrument, mode, moyen utilisé par N2	<i>compte utilisateur, interface client</i>	3
13b	N1 est un instrument, mode, moyen utilisé pour N2	<i>clé de chiffrement, barre de navigation</i>	109
14	omission	<i>non-connecté</i>	1
	Total		649

Les 649 noms composés français se répartissent en 14 catégories. Le fait le plus marquant est la parfaite similitude des relations déterminé/déterminant des composés français et anglais. Une seule relation concernant un composé (*non-connecté*) est nouvelle : 14 'N1 est omis, N2 est autonome'. Deux raisons peuvent expliquer cette similitude :

- Parmi les 649 noms composés français, 79 sont des composés exogènes (14,63% du corpus). Leur relation identique se retrouve donc dans les analyses anglaises et françaises. Pour tous les termes exocentriques français présents dans le corpus anglais, les relations sont identiques. Par exemple, pour EN *backdoor* et FR

, *porte de derrière* la paraphrase est ‘N1N2 est un accès secret’. Ces composés entrent dans la catégorie 5.

- Bon nombre de noms composés français sont des calques des composés anglais. Le procédé du calque, consistant à emprunter le signifié anglais et à traduire littéralement le signe en français, garde la relation sémantique. Par exemple, la relation sémantique entre le déterminé et le déterminant de EN *high-speed access* et FR *accès haut-débit* est 4 ‘N1 a un fonctionnement dont le procédé, le mode est N2’. Celle entre les formants de EN *discussion group* et FR *groupe de discussion* est 13b ‘N1 est un instrument, mode, moyen utilisé pour N2’. Dans ces deux exemples, le sens d'une lexie du xénolexique est emprunté et attribué à une forme française traduite littéralement et la relation N1-N2 est conservée.

Comme en anglais, il est évident que des noms composés formés par un même patron syntaxique et des composés ayant un même déterminé ou un même déterminant peuvent présenter des relations sémantiques différentes.

La catégorie 5 ‘*N1N2 est un N1 et/ou un N2’ contient essentiellement des composés exocentriques, comme dans le corpus anglais. La relation N1-N2 ne peut pas être découpée et paraphrasée autrement que ‘*N1N2 n'est un N1 et/ou un N2’. Par métaphore, un *goulot d'étranglement* est un bouchon virtuel, qui se produit lorsque plusieurs routines exécutées en même temps tentent d'accéder aux mêmes ressources simultanément, ce qui provoque une insuffisance des ressources disponibles et une diminution des performances de l'ordinateur. Une *jeune poussée* est une entreprise innovante et dynamique lancée depuis peu et qui est promise à une croissance rapide.¹²⁴

Les relations dont la fréquence « type » est la plus élevée sont :

- La fonction avec N1 actif (n°3a, 3b et 3c) avec 25,12% des composés.
- Le principe de fonctionnement (n°4) est présent dans 18,18% des termes du corpus français.
- La nature (n°6) est la relation unissant le déterminé et le déterminant de 15,25% des composés.
- La localisation, l'environnement (n°7) est présente dans 11,86% des termes.

Ces relations occupent des rangs similaires en anglais et en français et leurs proportions diffèrent très peu :

Tableau 44 : Résumés des relations sémantiques fréquentes en français et anglais

¹²⁴ Source des deux définitions : www.granddictionnaire.com

Relation sémantique	Français	Anglais
3a 'N1 (logiciel, matériel, technique) a pour activité, fonction N2', 3b 'N1 (humain entreprise) a pour objet d'activité, fonction N2' et 3c 'N1 (virus, attaque, pirate) a pour fonction N2'	25,12%	23,33%
4 'N1 a un fonctionnement dont le procédé, le mode est N2'	18,18%	18,81%
6 'N1 se distingue des autres N1 car il est N2'	15,25%	14,10%
7 'N1 a N2 comme localisation, environnement'	11,86%	13,38%

Pour conclure, les équivalents français des composés anglais étant majoritairement des calques et des emprunts, les composés français et anglais présentent les mêmes relations sémantiques. Nous examinerons si les difficultés rencontrées par le système d'équivalences correspondent à une ou plusieurs relations sémantiques et si les composés de la catégorie 5 sont les plus imprévisibles (Part. III).

3. La Combinatoire intermatricielle

Rappelons brièvement le contexte considéré et la méthode de calcul de la combinatoire.¹²⁵ Premièrement, nous ne prenons en considération que les matrices entrant dans la formation du terme dans le domaine d'Internet. Lorsqu'une matrice a été utilisée avant l'introduction d'une lexie dans la terminologie d'Internet, elle n'est pas comptabilisée, sauf si le domaine est l'informatique (frontière interdisciplinaire floue et changement de dimension). Deuxièmement, chaque phase de création n'est pas détaillée. Dans les néologismes constitués d'un signe existant et d'un autre élément, l'élément déjà existant n'est pas recensé car il n'est pas pensé au moment de la nouvelle création ; il est réutilisé.

3.1. Analyses

Tableau 45 : Combinaisons des matrices lexicogéniques du corpus français

¹²⁵ Part. II, Ch. 2, Sec. 3.2.1. et 3.2.4. pour les détails.

Rang	Combinasions	Fréquence « type »	%	Exemples
1	C M	121	18,6%	adresse électronique
2	C M M	112	17,3%	adresse d'hébergement
3	C Ps	63	9,7%	cyberdélinquant
4	C M C	59	9,1%	gestionnaire de courrier électronique
5	C Ps M	56	8,6%	e-courrier
6	C C	38	5,9%	entreprise point com
7	C M Si	35	5,4%	accès FTP
8	M	26	4,0%	boîte aux lettres
9	C M A	24	3,7%	aspirateur Internet
10	C	19	2,9%	groupe de contact
11	C A	16	2,5%	logiciel partagé
12	C Si	9	1,4%	pack ADSL
	C A C	9	1,4%	logiciel spyware
13	A	6	0,9%	hacktiviste
	C M Me	6	0,9%	fichier Napster
14	C C C	5	0,8%	accès web haut débit
	C Me	5	0,8%	opérateur booléen
15	C Ps C	4	0,6%	e-place de marché
16	C Me Ap	3	0,5%	point com
	C Ps S	3	0,5%	webmarketeur
	C Ps Ah	3	0,5%	webzine
17	C A Si Me	2	0,3%	modem ADSL
	C P	2	0,3%	protection antispam
	C Me Si	2	0,3%	voix sur IP
	C S M	2	0,3%	bloqueur de popup
	C Ps Ap	2	0,3%	webcam
18	A M Ap	1	0,2%	shopbot
	C A P	1	0,2%	logiciel anti-pup
	C C Me	1	0,2%	mise à jour antivirus
	C A Ps	1	0,2%	logiciel web
	C M Ap	1	0,2%	site perso
	A M Ps	1	0,2%	espiongiciel
	A Ps Ah	1	0,2%	nétiquette
	C A Me	1	0,2%	modem satellite
	C A Ps M	1	0,2%	logiciel anti popup
	C M A C	1	0,2%	connexion Internet sans fil
	C M M S	1	0,2%	site hacké
	C M Si Me	1	0,2%	site MP3
	C Me Me	1	0,2%	click and mortar
	C Ps M Ah	1	0,2%	cyberpunkitude
	C Si Ps	1	0,2%	wappanaute
	C Si S	1	0,2%	waplockage

	M S	1	0,2%	<i>hotliner</i>
	Total	649	100,00%	

Les noms composés du corpus français ont été créés par quarante-trois schémas combinatoires différents. Pourquoi ce nombre est-il plus élevé que pour le corpus anglais ? Plusieurs équivalents français peuvent correspondre à un terme anglais. Il arrive que ces équivalents ne soient pas tous formés sur le même schéma, ce qui augmente le nombre de combinaisons par rapport à l'anglais. Par exemple, *bulk e-mail* (C C) a comme équivalents français *courrier non-sollicité* (C M C), *publicité non-sollicitée* (C C), *message non-sollicité* (C M C), *e-mail non-sollicité* (C C C).

Le nombre de matrices utilisées pour créer un terme varie, comme pour l'anglais, de une (*hacktiviste*) à quatre (*cyberpunkitude*). La moyenne est de 15 termes par schéma combinatoire. Neuf des dix schémas supérieurs à cette moyenne se démarquent très nettement : C M, C M M, C Ps, C M C, C Ps M, C C, C M Si, M, C M A.

La composition est présente dans quarante-et-une combinaisons sur quarante-trois et a créé 95,8% des termes du corpus français (et non 100% puisque quelques composés ont été créés avant leur utilisation dans le domaine d'Internet). La composition est récurrente dans onze schémas combinatoires, ce qui représente 25,5% des termes.

La métaphore est la matrice avec laquelle la composition est la plus fréquemment combinée. Au total, 69,7% des termes du corpus (18 schémas sur 43) contiennent au moins un élément métaphorique. La composition et la métaphore ont ensemble permis la création de 65,5% des noms composés du corpus. Cette combinaison est présente dans seize schémas sur quarante-trois. Le pourcentage des composés renfermant une métaphore est de 48 et celui des composés avec deux métaphores est de 17,5. En français, comme en anglais, la composition et la métaphore sont les matrices qui ont la plus grande affinité.

En français aussi, la deuxième combinaison la plus fréquente est composition-pseudo-confixation : 9,7% C Ps, 8,6% C Ps M, 0,6% C Ps C, 0,5% C Ps S, 0,5% C Ps Ah, 0,3% C Ps Ap, 0,2% A M Ps, 0,2% A Ps Ah, 0,2% C A Ps, 0,2% C A Ps M, 0,2% C Ps M Ah. Au total, le nombre de termes pseudo-confixés est de 21,4%.

Avec 3,6% des composés, la métonymie est peu présente dans le corpus français. Dans un cas seulement, elle est récurrente : *click and mortar* est totalement métonymique.

3.2. Bilan

Tableau 46 : Fréquence des combinaisons de matrices dans le corpus français

Combinaisons	%	Fréquence « type »	% anglais
Composition + métaphore	65,5	16	64,5
Composition + pseudo-confixation	21,4	12	21,5
Composition + siglaison	8	7	7,5
Composition + métonymie	3,6	9	3,4
Composition + apocope	1,2	4	0,6
Composition + suffixation	1,4	5	0,2
Composition + aphérèse	0,9	3	0,2
Composition + préfixation	0,5	2	0

Les pourcentages du tableau 46 sont équivalents à ceux observés pour le corpus anglais. Ces résultats appellent deux questions : la prédominance de la combinaison composition-métaphore dans le lexique français d'Internet est-elle due à l'influence de l'anglais ? Les métaphores des composés anglais sont-elles reprises en français ? Nous reprendrons l'analyse de la traduction des composés métaphoriques anglais en français de Brolles (2001).

Meyer et al. (1997) ont enregistré trois stratégies de traduction des termes métaphoriques : le maintien, le changement ou la suppression des métaphores. Une métaphore est conservée si elle est transparente – identifiable et compréhensible – pour les français. Par exemple, *blacklist* est une métaphore empruntée à l'anglais et *mot de passe* est le calque de *password*. Ces deux métaphores sont compréhensibles pour les locuteurs français.¹²⁶ Parmi les trente-huit termes anglais métaphoriques du corpus de Brolles (cinquante termes au total), vingt-huit équivalents français contiennent une métaphore transparente. Lors du passage vers le français les métaphores sont calquées (*agent intelligent*) ou empruntées (*surf, cookie*), aucune n'est abandonnée. Les dix métaphores opaques se répartissent en neuf emprunts (*bookmark, newsgroup*) et un calque. Le maintien des métaphores suppose que les pays source et cible partagent la même culture ou ont des environnements culturels proches, ou bien que la communauté emprunteuse souhaite appréhender une autre culture.

Certaines des métaphores conservées sont opaques pour le locuteur français (*bookmark, weblog*, etc.). Plusieurs facteurs expliquent l'emprunt de termes opaques pour l'internaute français :

- Les emprunts sont monosémiques. En France, nous ne connaissons pour *popup* que le sens utilisé dans le cadre d'Internet.¹²⁷
- La langue française a une bonne capacité d'assimilation. L'attribution d'un genre à un substantif est systématique (*créer une startup, recevoir un e-mail*, etc.). La prononciation de l'emprunt est en général facilement mémorisable et/ou adaptable au système phonologique français (*e-mail* est prononcé /imDI/ et *chat* /tGat/).
- Certains emprunts s'avèrent productifs en français (EN to *chat* > FR *chater*, EN

¹²⁶ Par locuteur français, nous entendons locuteur francophone moyen.

¹²⁷ <http://www.granddictionnaire.com> : « Petite fenêtre publicitaire s'ouvrant automatiquement sur la page principale d'un internaute, une fois que celui-ci accède à certains sites commerciaux ».

startup > start-upeur).

Brolles conclut son analyse de la combinatoire par le fait que la métaphore est la matrice lexicogénique la plus utilisée dans le cadre d'Internet et qu'elle semble être le moteur principal de la néologie. La néologie formelle intervient en second¹²⁸ pour adapter le terme aux particularités de la langue française. Étant donné les principales caractéristiques de la composition, nous ne pensons pas que la métaphore soit le moteur principal. La composition forme des dénominations complètes et claires – décrivant bien le concept – et souvent monosémiques, comme *gestionnaire de téléchargement* ou *comparateur de prix* – contrairement à des termes métaphoriques simples comme *forum*, *portail* ou *fournisseur*.

À l'instar de l'auteur, mais pour des motifs différents, nous pensons que la métaphore n'est pas la principale raison justifiant le recours massif au calque et à l'emprunt. Internet est un domaine dont la terminologie n'est pas régie par une autorité centrale et dans lequel les innovations sont rapides et multiples. Ces réalités font du calque et de l'emprunt les deux procédés les plus appropriés pour créer des termes dénommant des concepts nouveaux car ils sont rapides à appliquer et permettent l'identification à la communauté internaute anglophone.

Chapitre 4. Les noms composés espagnols

1. Le profil morphosyntaxique des noms composés collectés

Situer la terminologie espagnole d'Internet par rapport à d'autres terminologies sera moins aisés. Le manque de sources rend impossible les comparaisons interdisciplinaires comme dans les chapitres 2 et 3. Cependant, cette analyse ajoutée à d'autres, à venir ou hors de notre portée, permettra de pallier cette insuffisance.

1.1. Les structures compositionnelles et leur fréquence « type »¹²⁹

Cette partie nous permettra de savoir si les patrons syntaxiques espagnols, qui ont formé les 576 noms composés du corpus, sont proches des patrons anglais et français. Grâce à deux études – Bustos Gisbert (1986) et Alvar (1999) –, nous tenterons de situer la terminologie d'Internet par rapport à la langue espagnole générale.

1.1.1. Les composés binaires

¹²⁸ La combinaison métaphore-composition représente 72% du corpus de l'auteur (Brolles, 2001 : 32). « La composition conserve un rôle important ; la juxtaposition reste le procédé le plus employé [...] » (*ibid.* : 37).

¹²⁹ Pour les définitions des notions de fréquence et de productivité et la liste des abréviations utilisées se reporter au Part. II, Ch. 1, Sec. 1.1.

Tableau 47 : Structures compositionnelles endogènes et exogènes des composés binaires du corpus espagnol

	Structures	Fréquence « type »	%
1	<u>Composés endogènes :</u>	435	88,23%
1.1	<i>purs</i>	208	42,19%
1.1.1	N + de + N	111	22,52%
1.1.2	N + A	79	16,02%
1.1.3	N + N	7	1,42%
1.1.4	N + a + N	3	0,61%
1.1.5	PS + N	2	0,41%
	N + PS	2	0,41%
1.1.6	no + N	1	0,20%
	V + N	1	0,20%
	N + en + N	1	0,20%
	A + N	1	0,20%
1.2	<i>hybrides</i>	227	46,04%
1.2.1	N + N	78	15,82%
1.2.2	PS + N	54	10,95%
1.2.3	N + de + N	41	8,32%
1.2.4	N + PS	23	4,67%
1.2.5	N + A	22	4,46%
1.2.6	PS + A	4	0,81%
1.2.7	Fdiv	3	0,61%
1.2.8	N + a + N	2	0,41%
2	<u>Composés exogènes</u>	58	11,76%
	TOTAL	493	100,00%

Comme dans l'étude morphosyntaxique française, on retrouve la subdivision composés endogènes purs/composés endogènes hybrides (Ch. 3, Sec. 1.1.1.). Avec 88,23%, les composés endogènes sont majoritaires. Ce pourcentage est légèrement supérieur à celui du français (85,37%). L'influence de la RAE, qui n'apprécie guère les emprunts, n'est certainement pas étrangère à cette différence (Ahronian, 2001 : 19-20). Dans les deux langues cibles, les emprunts occupent une faible place malgré le fait a) qu'Internet soit un des domaines les plus marqués par l'anglais, b) que son développement et sa diffusion soient rapides et c) que sa médiatisation soit importante.¹³⁰ Les raisons de la place minime des emprunts sont :

- La nature des unités lexicales étudiées. Nous analysons des termes longs alors que les emprunts sont généralement des lexies courtes.
- La nature de notre corpus. Il est écrit et constitué de revues grand public.

¹³⁰ Ces raisons ne sont que brièvement rappelées ici (Part. II, Ch. 3, Sec. 1.1.1.).

- Notre manière de comptabiliser les emprunts. Nous comptons les emprunts morphosémantiques uniquement.

Avec 46,04%, les composés hybrides sont les plus nombreux parmi les binaires. Ce pourcentage est de deux points supérieurs au français (44,26%). Le taux élevé de termes hybrides s'explique par la diffusion rapide du vocabulaire par les médias. En effet, les emprunts sont implantés avant que les spécialistes n'aient le temps de créer des équivalents espagnols. Les éléments anglais des termes hybrides sont des sigles (*cliente FTP, fichero MP3, mensajería SMS*), des amalgames (*Internet móvil, módem analógico*) et des pseudo-confixes (*email comercial, e-compra, gestor web*). Ces emprunts sont monosyllabiques (*chat, mail, news*) ou dissyllabiques (*cookie, software, hacker*), sauf certains amalgames. Ces formations hybrides distinguent le lexique d'Internet des lexiques spécialisés soutenus. L'absence manifeste de frontières entre la langue populaire et la langue scientifique dans le domaine d'Internet se retrouvant dans les deux langues cibles, on peut sans doute parler d'anglicisation des langues.

Tableau 48 : Structures compositionnelles endogènes des composés binaires du corpus espagnol

	Structures	Fréquence « type »	%
1	N + de + N	152	34,94%
2	N + A	101	23,22%
3	N + N	85	19,54%
4	PS + N	56	12,87%
5	N + PS	25	5,75%
6	7 structures mineures ¹³¹	16	3,68%
	TOTAL	435	100,00%

Douze patrons syntaxiques endogènes sont recensés. Les composés les plus fréquents sont ceux formés de deux substantifs reliés par *de* (type *página de inicio, contador de visitas, grupo de discusión*), ceux constitués d'un substantif et d'un adjectif (type *página inicial, sitio visitado, ADSL multipuesto*) ou de deux substantifs (type *código fuente*), les pseudo-confixes précédant un nom (type *ciberterrorista, e-compra, teletrabajo*) et les noms suivis d'un pseudo-confixe (type *dirección web, jubilnauta, navegador web*). Ces cinq structures représentent 96,32% des composés. Les sept autres se partagent les 3,68% restants.

N + de + N est, comme en français, le type le plus fréquent, mais le pourcentage relevé pour l'espagnol est plus important (34,94% contre 28,42% pour le français). Cet écart est compensé par la proportion de N + N plus élevée en français (25,38%) qu'en espagnol (19,54%). L'espagnol usant moins de la juxtaposition, au moins dans la terminologie d'Internet, il crée beaucoup de composés syntagmatiques. À cet égard, les avis des auteurs divergent. Pour Belot (1987 : 47), « on assiste, aussi bien en France qu'en Espagne, à un mode de création par gémellation ou juxtaposition » sous l'influence

¹³¹ Il s'agit de : N+a+N (1,15%), PS+A (0,92%), Fdiv (0,69%), A + N (0,23%), no+N (0,23%), V+N (0,23%), N+en+N (0,23%).

de l'anglo-américain alors que pour Bustos (1986 : 182) les N + N sont peu fréquents en espagnol – et dans les langues romanes – par rapport aux langues germaniques pour des raisons historiques (pérennité du génitif) et formelles (ordre déterminant/déterminé).

Dans les composés syntagmatiques, la relation syntaxique est le plus souvent rendue par la préposition *de*. Les éléments peuvent également être reliés par *a*, – probablement sous influence étrangère (Alvar Ezquerro, 1999 : 22). Cette influence est peu marquée ici puisque seulement 1,15% des composés ont la préposition *a*. Toujours selon Alvar, la présence d'autres prépositions, comme *en*, est rare. Notre corpus corrobore cette affirmation : 0,23% des termes sont formés avec *en* et aucune autre préposition n'a été relevée.

Les composés de type N + A sont légèrement plus fréquents (23,22%) qu'en français (22,13%) et qu'en anglais (18,09%). La différence de pourcentage entre l'espagnol et le français n'est pas significative au regard du nombre de termes.¹³²

En espagnol, les pseudo-confixes sont, comme dans la terminologie française, le quatrième patron le plus fréquent, mais les pourcentages diffèrent (respectivement 14,75% et 12,87%). La totalité des composés ayant un pseudo-confixe représente 19,54% des binaires. Ce mode de formation est passé de l'anglais (26,14% des binaires) au français (20,38%) et à l'espagnol. La fréquence est plus faible en espagnol qu'en anglais puisqu'une partie des composés pseudo-confixés sont des créations exogènes (*adware*, *e-business*, *webmaster*, etc.).

1.1.2. Les surcomposés à trois éléments

Tableau 49 : Structures compositionnelles endogènes et exogènes des surcomposés à trois éléments du corpus espagnol

¹³² Le corpus espagnol compte 100 composés N + A et le corpus français 101. La différence de pourcentage est due au fait que le corpus espagnol compte moins de binaires (493) que le corpus français (536).

Les noms composés anglais français et espagnols du domaine d'Internet :

	Structures	Fréquence « type »	%
1	<u>Surcomposés endogènes :</u>	59	85,51%
1.1	<i>purs</i>	33	47,83%
1.1.	N + de + N + de + N	10	14,49%
	N + de + N + A	10	14,49%
1.1.2	N + A + A	3	4,35%
1.1.3	N + de + A + N	2	2,90%
1.1.4	N + N + A	1	1,45%
	N + N + PS	1	1,45%
	N + en + N	1	1,45%
	N + más + N	1	1,45%
	N + N + N	1	1,45%
	N + a + A + N	1	1,45%
	N + N + a + N	1	1,45%
	N + fuera + de + N	1	1,45%
1.2	<i>hybrides</i>	26	37,68%
1.2.1	N + N + A	6	8,70%
1.2.2	N + de + N + de + N	3	4,35%
	N + de + N + a + N	3	4,35%
	N + de + N + PS	3	4,35%
1.2.4	N + PS + A	2	2,90%
1.2.5	N + A + A	1	1,45%
	N + A + PS	1	1,45%
	V + N + N	1	1,45%
	N + sobre + N	1	1,45%
	N + de + A + N	1	1,45%
	N + de + PS + N	1	1,45%
	N + A + a + N	1	1,45%
	N + N + a + N	1	1,45%
	N + de + N + N	1	1,45%
2	<u>Surcomposés exogènes :</u>	10	14,49%
	TOTAL	69	100,00%

Les surcomposés du tableau 49 sont massivement endogènes (85,51%), comme en français, mais avec un pourcentage inférieur (94,89%), alors que dans les deux langues cibles les proportions sont identiques chez les composés binaires (FR 85,37%, ES 88,23%).

Les composés exogènes (14,49%) sont plus fréquents que chez les surcomposés français à trois éléments (5,10%) et que chez les binaires espagnols (11,06%), donnée étonnante puisque plus un emprunt comporte d'éléments de nomination, plus le concept peut être difficile à appréhender pour les internautes ayant une faible maîtrise de la langue

anglaise. Sur les dix composés exogènes, la moitié est très certainement peu employée. Ce sont des développements de sigles (*Domain Name System, File transfer Protocol, Frequently Asked Questions, World Wide Web, Wireless Application Protocol*) et deux sont relatifs au piratage¹³³ et sont plus courts que leurs équivalents espagnols (*password cracker, shadow password*).

À l'inverse des composés binaires, les composés hybrides (37,68%) sont moins fréquents que les composés endogènes purs (47,83%). La compréhension des termes longs du corpus contenant un emprunt n'est pas plus complexe que celle des composés binaires car les emprunts inclus dans les surcomposés hybrides sont les mêmes que chez les binaires (*newsletter electrónica, Internet de alta velocidad, página web estática*) ou sont des sigles (*conexión ADSL multipuesto, dirección IP dinámica*). Deux faits nous éclairent sur le taux plus faible de termes hybrides :

- Les composés contenant des sigles sont moins nombreux chez les surcomposés que chez les binaires.
- Les composés pseudo-confixés sont moins fréquents que chez les binaires (13,54% contre 19,54% chez les binaires).

Tableau 50 : Structures compositionnelles des surcomposés endogènes à trois éléments du corpus espagnol

	Structures	Fréquence « type »	%
	3 structures majeures :	30	50,85%
1	N + de + N + de+ N	13	22,03%
2	N + de + N + A	10	16,95%
	N + N + A	7	11,86%
	4 structures moyennes :	13	22,03%
3	N + A + A	4	6,78%
4	N + de + A + N	3	5,08%
5	N + de + N + PS	3	5,08%
	N + de + N + a + N	3	5,08%
	14 structures mineures :	16	27,12%
134			
	Total	59	100,00%

Vingt-et-un patrons de surcomposés à trois éléments sont recensés dans le corpus espagnol. Les trois plus fréquents représentent la moitié des surcomposés (50,85%) : N + de + N + de + N (type *ataque de denegación de servicio, proveedor de servicios de*

¹³³ Ce thème est certainement un de ceux qui comportent le plus d'emprunts.

¹³⁴ N+PS+A (3,39%), N+N+a+N (3,39%), N+A+PS (1,69%), N+N+PS (1,69%), V+N+N (1,69%), N+sobre+S (1,69%), N+en+N (1,69%), N+más+A (1,69%), N+N+N (1,69%), N+a+A+N (1,69%), N+de+PS+N (1,69%), N+A+a+N (1,69%), N+de+N+N (1,69%), N+fueras de+N (1,69%).

Internet), N + de + N + A (type *red de área local*, *virus de correo electrónico*) et N + N + A (type *dirección IP pública*, *conexión ADSL multipuesto*). L'autre moitié (49,15%) est issue de dix-huit patrons. Les cinq structures avec pseudo-confixe totalisent 13,54% des composés endogènes.

La formation de la terminologie d'Internet est homogène dans la mesure où N + de + N + de + N est le patron de surcomposés le plus fréquent, tout comme N + de + N chez les binaires. Ces surcomposés sont des N + de + N auxquels des précisions ont été ajoutées : *programa de intercambio de ficheros*, *fichero de firmas de virus*, etc. Les N + de + N + de + N sont plus représentés en espagnol qu'en français (22,03% contre 12,90%). Cela semble cohérent puisque, comme nous l'avons vu, les N + de + N sont plus nombreux en espagnol qu'en français.

Le troisième patron le plus fréquent, N + de + N + A (16,95%), a la même raison d'être que N + de + N + de + N. Ce sont des composés N + de + N auxquels une extension a été ajoutée : *servidor de correo entrante*, *servidor de correo saliente*, etc. La fréquence des N + de + N + A est similaire à celle du français (16,13%).

Les procédés les plus fréquents en anglais étant N + N + N et A + N + N, il est naturel que les structures citées précédemment créent des termes dans les langues cibles. Ces surcomposés sont des calques des surcomposés anglais : *Protocolo de Transferencia de Ficheros* et *File Transfer Protocol*, *página de inicio personal* et *personal home page*, *sistema de detección de intrusiones* et *intrusion detection system*, etc.

La présence des pseudo-confixés est plus faible (13,54%) qu'en anglais (26,98%) et en français (15,05%). La différence de pourcentage entre l'anglais et les langues cibles vient de la nature de nos catégories. Un grand nombre des ces surcomposés sont des créations exogènes. Les créations espagnoles sont cependant moins fréquentes que les créations françaises. Là où le français reprend un pseudo-confixe :

- L'espagnol ne le fait pas : EN *e-mail client* / FR *client e-mail* / ES *cliente de correo*, *cliente de mail* ; EN *e-mail provider* / FR *fournisseur d'e-mail* / ES *proveedor de correo*, etc. ;
- Les équivalents sont absents de notre corpus : EN *e-mail contact* / FR *contact e-mail* / ES ?, EN *e-mail encryptor* / FR *encodeur d'e-mail*, *chiffrement d'e-mail* / ES ?, EN *cybersupermarket* / FR *cybersupermarché* / ES ?, EN *e-commerce package* / FR *programme d'e-commerce*, EN *e-mail alert* / FR *alerte e-mail* / ES ?, etc.

1.1.3. Les surcomposés à quatre et cinq éléments

Tableau 51 : Structures compositionnelles endogènes et exogènes des surcomposés à quatre et cinq éléments du corpus espagnol

	Structures	Fréquence « type »	%
1	Surcomposés endogènes :	12	85,71%
1.1	purs	8	57,14%
1.1.1	N + de + N + en + N	2	14,29%
1.1.2	N + A + no + N	1	7,14%
	N + por + N + A	1	7,14%
	N + A + de + N + A	1	7,14%
	N + de + N + A + A	1	7,14%
	N + de + N + de + N + A	1	7,14%
	N + A + de + todo + N	1	7,14%
1.2	hybrides	4	28,57%
1.2.1	N + a + N + a + A + N	1	7,14%
	N + de + N + de + PS + N	1	7,14%
	N + a + N + por + N	1	7,14%
	N + por + N + a + N	1	7,14%
2	Surcomposés exogènes :	2	14,29%
	TOTAL	14	100,00%

Avec 85,71%, les termes endogènes sont autant représentés que chez les surcomposés à trois éléments. Cependant, à l'intérieur des formations endogènes, les termes hybrides sont moins fréquents (28,57% contre 37,68% chez les trois éléments). On ne trouve qu'un seul composé pseudo-confixé.

En revanche, les composés exogènes sont bien moins présents en espagnol (14,29%) qu'en français (45,45%). Peut-on trouver une explication à cela ? Comme nous l'avons montré, en français, les surcomposés de plus de trois éléments de notre corpus sont tous relatifs à des aspects techniques d'Internet (protocoles, langages, modes de connexion). En espagnol, la plupart des termes longs sont liés à ces mêmes thèmes : modes de connexion (*conexión por línea conmutada, acceso a Internet a alta velocidad, conexión permanente de banda ancha*, etc.) et langages (*Lenguaje de Marcado de Hipertexto*). Les autres termes désignent des concepts du courrier électronique (*correo electrónico no deseado*) et des logiciels (*aplicación libre de todo coste*). Contrairement aux journalistes français, les journalistes espagnols emploient des composés endogènes sans mentionner les équivalents anglais. Il en découle un faible taux d'emprunts.

Tableau 52 : Structures compositionnelles des surcomposés endogènes à quatre et cinq éléments du corpus espagnol

	Structures	Fréquence « type »	%
1	N + de + N + en + N	2	16,67%
2	N + A + no + N	1	8,33%
	N + por + N + A	1	8,33%
	N + a + N + a + A + N	1	8,33%
	N + A + de + N + A	1	8,33%
	N + de + N + de + PS + N	1	8,33%
	N + de + N + A + A	1	8,33%
	N + de + N + de + N + A	1	8,33%
	N + a + N + por + N	1	8,33%
	N + por + N + a + N	1	8,33%
	N + A + de + todo + N	1	8,33%
	TOTAL	12	100,00%

Il existe onze patrons de surcomposés à quatre éléments. Aucun n'a une fréquence plus élevée que les autres, comme pour l'anglais et le français. Le tableau 52 met en évidence la faible fréquence des structures longues. Comme nous l'avons vu pour l'anglais et le français, les termes longs sont les moins nombreux (absence de termes très techniques dans notre corpus et tendance aux réductions) (Part, II Ch. 2 et 3, Sec. 1.1.3.) : 2,08% pour l'espagnol, 1,08 % pour l'anglais, 1,45% pour le français.

Contrairement à l'anglais et au français, une seule structure comprend un pseudo-confixe N + de + N + de + PS + N et elle n'a donné qu'un terme (*Lenguaje de Marcado de Hipertexto*).

1.1.4. Bilan

Tableau 53 : Les structures compositionnelles des composés et surcomposés espagnols

	Structures	Fréquence « type »	%
1	Composés endogènes	506	87,85%
	6 structures majeures :	419	72,74%
1.1	N + de + N	152	26,39%
1.2	N + A	101	17,53%
1.3	N + N	85	14,76%
1.4	PS + N	56	9,72%
1.5	N + PS	25	4,34%
	1 structure moyenne:	13	2,26%
1.6	N + de + N + de+ N	13	2,26%
	41 structures mineures :	74	12,85%
1.7	N + de + N + A	10	1,74%
1.8	N + N + A	7	1,22%
1.9	N + a + N	5	0,87%
1.10	PS + A	4	0,69%
	N + A + A	4	0,69%
1.11	Fdiv	3	0,52%
	N + de + A + N	3	0,52%
	N + de + N + PS	3	0,52%
	N + de + N + a + N	3	0,52%
1.12	N +PS + A	2	0,35%
	N + N + a + N	2	0,35%
	N + de + N + en + N	2	0,35%
1.13			

	A + N	1	0,17%
	no + N	1	0,17%
	V + N	1	0,17%
	N + en + N	1	0,17%
	N + A + PS	1	0,17%
	N + N + PS	1	0,17%
	V + N + N	1	0,17%
	N + sobre + N	1	0,17%
	N + en + N	1	0,17%
	N + más + A	1	0,17%
	N + N + N	1	0,17%
	N + a + A + N	1	0,17%
	N + de + PS + N	1	0,17%
	N + A + a + N	1	0,17%
	N + de + N + N	1	0,17%
	N + fuera + de + N	1	0,17%
	N + A + no + N	1	0,17%
	N + por + N + A	1	0,17%
	N + a + N + a + A + N	1	0,17%
	N + A + de + N + A	1	0,17%
	N + de + N + de + PS + N	1	0,17%
	N + de + N + A + A	1	0,17%
	N + de + N + de + N + A	1	0,17%
	N + a + N + por + N	1	0,17%
	N + por + N + a + N	1	0,17%
	N + A + de + todo + N	1	0,17%
2	Composés exogènes :	70	12,15%
2.1	Emp2	58	10,07%
2.2	Emp3	10	1,74%
2.3	Emp4	1	0,17%
2.4	Emp5	1	0,17%
	TOTAL	576	100,00%

Les 576 termes du corpus espagnol ont été créés par quarante-quatre patrons endogènes et quatre patrons exogènes. Le pourcentage moyen est 2,08% – soit 12 termes par patron. Cinq structures sont bien supérieures à ce chiffre et totalisent 72,74% : N + de + N (26,39%), N + A (17,53%), N + N (14,76%), PS + N (9,72%) et N + PS (4,34%). Les 43 autres procédés se partagent les 27,26% restants. En résumé, cinq structures ont donné naissance à plus des quatre cinquièmes des termes du corpus d'espagnol.

Après l'anglais et le français, l'espagnol confirme le fait que plus les composés sont longs, moins leur fréquence est élevée. On passe de 493 composés à 69 surcomposés de trois éléments, à 14 surcomposés de quatre et cinq éléments. Ces chiffres représentent respectivement 85,60%, 11,97% et 2,43% du corpus.

Le français et l'espagnol présentent un nombre similaire de structures (50 et 48) alors que l'anglais en possède une dizaine de plus (58). Le pourcentage moyen de termes créés par structure est cependant quasiment identique dans les trois langues : 1,72 pour l'anglais, 2 pour le français et 2,08 pour l'espagnol.

Le nombre total de composés comprenant un pseudo-confixe s'élève à 94 – soit 16,32 % du corpus. Ce pourcentage atteint 16,90% pour le français et 27,67% pour l'anglais. Globalement, les deux langues cibles utilisent les mêmes structures dans des proportions assez semblables. Elles calquent donc toutes deux les structures de la langue source.

Tableau 54 : Répartition des noms composés anglais et espagnols selon le nombre de formants¹³⁵

Nb de formants	2	3	4	5	6	Total
Corpus anglais	397	129	21	5	1	553
Corpus espagnol	493	69	13	1	0	576

1.2. Les composés pseudo-confixés

L'analyse des pseudo-confixes du corpus prend fin avec l'étude des pseudo-confixes du lexique espagnol d'Internet. L'historique des pseudo-confixes étant retracé (Part. II, Ch. 2, Sec. 1.2.), seuls leur inclusion dans deux dictionnaires espagnols, les DRAE 1996 et 2001 et le DGLE (*Diccionario General de la Lengua Española* en ligne de Vox), et l'ajout d'un nouveau sens seront vérifiés. Les acceptations mentionnées seront comparées à celles du MWOD et du CALDO pour l'anglais et du PR 1996 et du GR 2001 pour le français. Nous aurons ainsi un début d'aperçu diachronique et saurons s'il s'agit des mêmes éléments dans les trois langues.

1.2.1. ciber-

Le formant *ciber-* est un mot-vedette du DRAE 2001. Il a un statut d'élément de composition signifiant ‘cybernétique’, comme dans *ciberespacio* et *cibernauta*. *Ciberespacio* est une vedette (« *ámbito artificial creado por medios informáticos* »). *Cibernética* n'apparaît pas dans le DGLE. Seuls le substantif *cibernética* et l'adjectif *cibernético/a* sont répertoriés. En revanche, des termes relatifs au domaine d'Internet sont l'objet d'articles : *ciberartista*, *ciberespacio* et *cibernauta*. Pour chacun d'entre eux, l'ouvrage fournit une définition sans aucune information sur leur origine et leur formation. L'entrée de *ciberartista*¹³⁶ est inattendue puisque, contrairement à *ciberespacio* et *cibernauta*, ce terme ne désigne pas un concept fondamental du domaine.

Tableau 55 : Noms composés espagnols avec le pseudo-confixe *ciber-*¹³⁷

Pseudo-confixe	Base	Champ	Base	Champ

¹³⁵ Part. II, Ch. 3, Sec. 1.1.3. pour les commentaires.

¹³⁶ Un ciberartista es un artista que empleo los medios cibernéticos (DGLE).

Les noms composés anglais français et espagnols du domaine d'Internet :

		sémantique		sémantique
ciber	actividad	Sé	frontera	Sé
	adicto	Sé	guerra	S
	álbum	Sé	lenguaje	Sé
	amor	Sé	ley	S
	ataque	S	mayor	Sé
	<i>café</i>	Sé	mercado	S
	<i>centro</i>	Sé	mundo	Sé
	comerciante	C	<i>nauta</i>	Sé
	comprador	C	ocupa	Sé
	conferencia	Sé	ocupación	Sé
	conflicto	S	piratería	Sé
	contratación	Sé	plebiscito	Sé
	crimen	S	protección	S
	criminal	S	<i>punk</i>	S
	criminalidad	S	sexo	Sé
	debate	Sé	soldado	S
	delincuencia	S	supermercado	C
	delincuente	S	tendero	C
	delito	S	terrorismo	S
	derecho	S	terrorista	S
	directorio	N/A	tienda	C
	editor	Sé	universo	Sé
	espacio	Sé	visitante	Sé

Légende : C = commerce, N = navigation, S = sécurité relative au cyberespace, Sé = société

L'espagnol adapte la graphie du pseudo-confixecyber- au système espagnol en remplaçant le -y- de l'anglais par un -i-. L'espagnol adapte couramment les emprunts anglais à son système phonologique et graphique (*e-líder*, *clic*, etc) (Ahronian, 2001).

Ciber- entre dans la formation de quarante-six composés. Il a une fonction préfixale et l'ordre des éléments est invariablement déterminant/déterminé. Ces composés sont non-conformes à la structure déterminé/déterminant des composés espagnols. Pour justifier cet ordre, la raison évoquée pour le français peut être avancée : *ciber-*, la troncation de EN *cyberspace*, répond aux règles de formation des composés savants qui ont la tête à droite (*cibernético*). Il est homographe du confixe *ciber-*, mais c'est le succès de EN *cyberspace* qui a contribué à augmenter la fréquence « type » du modèle *ciber- + N*. Les termes espagnols – comme les termes français – sont des calques des composés anglais : EN *cyber activity* > ES *ciber actividad*, EN *cyber delinquency* > ES *ciber*

¹³⁷ La fréquence des composés des tableaux provient de www.lycos.es (option langue espagnole uniquement). Les termes en majuscules plus caractères gras ont un nombre d'occurrences supérieur à 1 000 000, ceux en majuscules ont entre 1 000 000 et 999 999 occurrences. Les lexies en caractères gras ont une fréquence comprise entre 10 000 et 99 999, celles en italique ont une fréquence entre 1 000 et 999. les termes en caractères normaux comptent moins de 999 occurrences.

delicuencia.

Aucun changement morphosyntaxique ou sémantique n'est à noter par rapport à l'anglais et au français. La structure *ciber-* + N a une fréquence « type » similaire dans les trois langues : 46 composés en espagnol, 44 en anglais et 49 en français. *Ciber-* modifie le sens traditionnel des lexies auxquelles il est accolé pour les adapter au monde du cyberespace. Le champ d'application des termes espagnols est identique à celui du français et de l'anglais. Les thèmes dans lesquels on trouve *ciber-* sont identiques ou proches des thèmes anglais et français : la vie dans le cyberespace, le commerce, la navigation et la sécurité des internautes.

1.2.2. *hiper-*

Pour le DRAE 1996 et le DGLE, *hiper-* est un élément grec entrant dans la formation de composés et signifiant « ‘superioridad’ o ‘exceso’: HIPERTensión, HIPERmercado, HIPERchlorhidria ». D'après les ouvrages lexicographiques consultés, *hiper-* est identique sémantiquement à ses équivalents anglais et français. Le DRAE 2001 inclut la lexie *hipertexto*.

Tableau 56 : Composés espagnols avec le pseudo-confixe *hiper-*

Pseudo-confixe	Base	Champ sémantique	Base	Pseudo-confixe	Champ sémantique
hiper	vínculo	P		Surcomposés avec:	
	texto	P	enlace de	hipertexto	P
			lenguaje de Marcado de hiperexto		P
			video	hipervinculado	P

Légende : P = procédé

Le corpus espagnol contient la même proportion de composés en *hiper-* que les corpus anglais et français. *Hiper-* désigne uniquement les concepts en rapport avec un procédé de navigation. Le champ notionnel auquel *hiper-* fait référence est donc très restreint.

Hiper- a une fonction préfixale et a perdu sa valeur intensive. Comme nous l'avons dit (Part. 2, Ch. 2, Sec. 1.2.2.), le succès d'*hiper-* a commencé avec l'*hypertext* de Ted Nelson. L'emploi de cet élément s'est ensuite étendu. D'autres composés non inclus dans notre corpus existent : *hiperenlace* (synonyme de *hipervínculo*), une *hiperficción* est une fiction qui utilise l'hypertexte, un *hiperdiccionario* est un dictionnaire basé sur l'hypertexte, un *hipermedio* est l'ensemble des moyens d'expression qui intervient dans un travail hypertextuel et une *hipertulia* est une « tertulia » axée sur les thèmes relatifs à l'hypertexte.¹³⁸ Les composés et surcomposés espagnols sont des calques des termes anglais : *hyperlink* > *hipervínculo*, *hypertext* > *hipertexto*, etc.

¹³⁸ Termes répertoriés par José Antonio Millan sur son site web : http://www.jamillan.com/v_index.htm

1.2.3. tele-

D'après le DRAE 1996, *tele-* est un élément de composition qui signifie ‘à distance’, comme dans *teléfono* et *televisión*. Ce dictionnaire contient également une entrée *tele*, troncation de *televisión*. La définition du DGLE est plus complète. *Tele-* est un élément préfixal ayant quatre acceptances :

A distancia, lejos de : teléfono, televisión, teleférico ; televisión : teleclub, teledebate ; teleférico : telesquí ; telefotografía : teleobjetivo

L'article *tele-* du DRAE 1996 est moins précis que ceux du PR 1996 et du DGLE. *Tele-* n'apparaît pas comme troncation de *teleférico* alors que le mot tête de l'article *telecabina* est *teleférico* (« teleférico de cable único para la tracción y la suspensión, dotado de cabina »). Comme en français, *tele-* existait en tant que pseudo-confixe avant son utilisation dans le lexique d'Internet. Dans *telefilme* et *telenovela*, *tele-* est la troncation de *televisión* et dans *telecabina* et *telesquí*, *tele-* est la troncation de *teleférico* (créé d'après *téléphérique* en français selon le DRAE 1996). Dans le lexique d'Internet, *tele-* est la réduction de *telemática* (ou de *telecomunicación*, Part. II, Ch. 3, Sec. 1.2.3.).

Tableau 57 : Composés espagnols avec le pseudo-confixe *tele-*

Pseudo-confixe	Base	Champ sémantique
tele	<i>centro</i>	Mt
	<i>trabajador</i>	Mt
	<i>trabajo</i>	Mt

Légende : Mt = monde du travail

Avec trois termes, ce pseudo-confixe est moins fréquent en espagnol qu'en français (le corpus français compte 5 composés et 5 termes dérivés de *téléchargement*). Aucun de ces trois termes ne possède une entrée dans le DRAE. *Teletrabajo* est inclus dans le DGLE (« trabajo que realiza habitualmente una persona para una empresa desde un lugar alejado de la sede de ésta a través de un sistema de telecomunicación »).

En français, la fréquence d'emploi de *téléchargement* a renforcé la fréquence du modèle *télé- + N*. *Tele-* n'étant pas fréquent en anglais, à quoi est due sa présence en espagnol ? Deux hypothèses peuvent être formulées :

- Puisqu'un seul composé est contenu dans notre corpus anglais (*teleworking*), le français est probablement à l'origine de l'implantation du modèle *tele- + N* en espagnol.
- L'implantation des termes *telecentro*, *teletrabajo* et *teletrabajador* est peut-être due au succès des concepts très utilisés que ces termes désignent (comme c'est le cas pour le téléchargement en France). Le concept de *download* est rendu en espagnol par le terme *descarga* ou *bajada* et non par un terme en *tele-*.

Ce pseudo-confixe occupe une fonction préfixale et forme des composés endocentriques dont il est le déterminant. Dans *teletrabajador* l'élément tête est *trabajador*. À l'instar de

ciber-, tele- modifie l'acception traditionnelle de lexies qui existaient avant Internet pour les adapter aux nouvelles technologies. Le champ notionnel concerné, 'activités rendues possibles sans se déplacer grâce à la télématique/Internet', est le même qu'en français. Une valeur positive se dégage de ces termes : en entendant parler de *teletrabajo*, on pense 'travail facilité par la télématique et Internet'.

1.2.4. meta-

Selon le DRAE 1996, *meta-* est un élément grec de composition qui a pour sens 'près de' (« junto a »), 'après' (« después de »), 'entre' (« entre ») ou 'avec' (« con »). Les exemples cités sont *metacentro* et *metatórax*. À la différence du MWOD et du PR 1996, aucune mention n'est faite de l'acception de *meta-* indiquant un changement. En revanche, le DGLE présente *meta-* comme un élément préfixal à cinq acceptions :

Meta- (gr. *meta*) 1 *Elemento prefijal que entre en la formación de palabras con el significado de cambio, mutación, más allá de, que engloba ; después, posterior ; por el medio. 2 QUÍM. Se utiliza en química para formar los nombres de compuestos menos hidratados y para designar la sustitución de los cuerpos de la serie bencénica.*

Tableau 58 : Composés espagnols avec le pseudo-confixemeta-

Base	Pseudo-confiexemeta	Champ sémantique	Pseudo-confiexemeta	Champ sémantique
etiqueta	meta	P	meta	tag
tag		P		<i>buscador</i>
				<i>portal</i>

Légende : A = application, P = procédé

Le corpus espagnol contient cinq composés pseudo-confixés avec *meta-*. Dans deux des cinq termes, *meta-* a une position suffixale et les termes répondent à l'ordre déterminé/déterminant des composés espagnols. ES *etiqueta meta* est la traduction littérale de EN *meta tag* et ES *tag meta* est l'adaptation au système espagnol de l'emprunt EN *meta tag*. Aucune des deux structures, *meta- + N* ou *N + meta-*, ne semble être privilégiée. Trois composés contiennent *meta-* en position préfixale. *Meta tag* est un emprunt et conserve l'ordre anglais. *Metabuscador* et *metaportal* sont nés du succès de l'emprunt EN *meta tag* qui a provoqué une hausse de la fréquence « type » de la structure *meta- + N* déjà existante en espagnol où elle a donné des lexies comme *metabolismo*, *metafísica*, *metalinguaje*, etc. La structure de ces exemples est identique à celle des composés d'Internet, mais *meta-* est un confixe classique.

En espagnol, l'élément *meta-*, employé dans les mêmes conditions qu'en anglais et en français, est probablement également issu de la troncation du terme EN *metaelement* (ES *metaelemento*). Dans ces nouvelles créations, l'espagnol conserve la structure des composés anglais (*meta- + N*) et traduit l'élément déterminé : *engine* > *buscador*, *portal* > *portal*. D'autres termes, non inclus dans notre corpus, sont formés sur ce modèle : ES *metadatos* (EN *metadata*), *metabúsqueda* (EN *metasearch*), *metaindexación* (EN *metaregistration / referencing*), *metaherramienta* (EN *metatool*).

1.2.5. web-

Web est absent du DGLE mais présent dans le DRAE 2001 où il est un mot venant de l'anglais signifiant ‘réseau informatique’. Le composé *página web* est aussi une vedette du DRAE 2001. Les espagnols étant plutôt réticents aux emprunts morphosémantiques, la fréquence « type » élevée du pseudo-confixe en anglais et la fréquence « token » élevée des termes avec *web-* en anglais sont vraisemblablement les raisons de l'utilisation de *web-* en espagnol.

Tableau 59 : Composés espagnols avec le pseudo-confixe *web-*

Base	Pseudo-confixe	Champ sémantique	Pseudo-confixe	Base	Champ sémantique
acceso	web	C	web	bug	S
administrador		Mt		invisible	N
aplicación		A		musical	N
ataque		S		personal	N
cámara		Co		radio	Co
correo		Co		visible	N
desarrollador		Mt		cam	Co
dirección		N		log	Co
directorio		N		logger	Co
diseñador		Mt		MASTER	Mt
diseño		Mt		SITE	N
espacio		N	Dérivé de:		
explorador		A	página web	estática	N
gestor		A			
interfaz		A			
interfaz gráfica		A			
web					
marketing		Mt			
navegador		A			
PÁGINA		N			
radio		Co			
servidor		C			
SITIO		N			
	Surcomposés avec :				
alojamiento de	sitio web	N			
empresa de	alojamiento web	N			
proveedor de		N			

Légende : A = application, C = connexion, Mt = monde du travail, N = navigation, S = sécurité, Sé = société

Web- est présent dans 37 termes. Les principales structures sont : N + *web* (21 composés), *web-* + N (6 composés) et *web* + A (4 composés). Les six autres termes sont des surcomposés. Le modèle le plus fréquent, *web-* + N, crée des termes répondant à l'ordre déterminé/déterminant. Ces composés sont hybrides (*administrador web*, *servidor*, *web*, *página web*, etc). Les composés présentant l'ordre déterminant/déterminé sont des emprunts (*web bug*, *webmaster*, etc).

Sémantiquement, les modèles *web-* + N et N + -*web* créent des composés endocentriques. Un *diseñador web*, un *explorador web* et un *webmaster* sont des types de *diseñador*, *explorador* et *master*. Concernant la structure *web-* + A, les remarques sont identiques à celles du corpus français. Certains composés désignent des parties du Web et sont endocentriques (*web invisible* et *web visible*), les autres désignent des types de sites web (*web musical*, *web personal*). *Web* peut alors être la troncation de *website* ou de *World Wide Web*.

Les champs notionnels relevés dans le corpus espagnol sont les mêmes que dans les corpus anglais et français : application, connexion, monde du travail, navigation, sécurité et société.

1.2.6. *net-*

En anglais, *net* faisait référence à un instrument de pêche et de chasse et a acquis d'autres sens au XX^e siècle. L'espagnol a connu une évolution similaire : l'utilisation métaphorique de *red* (EN *net*) est antérieure à Internet. *Red* s'emploie pour parler de choses connectées qui peuvent être immatérielles comme *red de espías*, *red de distribución*, etc.¹³⁹ *Net* n'est répertorié ni par le DRAE 2001 (*Internet* est également absent), ni par le DGLE.

Tableau 60 : Composés espagnols avec le pseudo-confixenet-

Pseudo-confixe	Base	Champ sémantique
net	cam	Sé
	economía	C
	etiqueta	Sé

Légende : C = commerce, Sé = société

Ce pseudo-confixe est peu fréquent en espagnol. Le corpus espagnol contient seulement trois termes alors que le corpus anglais en contient neuf et le corpus français sept. Contrairement au français, l'espagnol n'adapte pas les composés en *net-* à son système. Seul le modèle *net-* + N est fréquent alors que le français utilise l'ordre déterminé/déterminant avec la structure N + -*net* après avoir emprunté la formation *net-* + N. En espagnol, il n'y a pas de différence formelle entre un emprunt (*netcam*) et un terme espagnol (*neteconomía*). Dans les deux cas, l'ordre est déterminant/déterminé.

Les thèmes abordés sont semblables à ceux de l'anglais et du français. Faire de la

¹³⁹ Ces informations se trouvent sur le site web de Millán : http://www.jamillan.com/v_index.htm

neteconomía, respecter la *netiqueta* et utiliser une *netcam* ne demandent aucun savoir informatique.

Comment expliquer la faible fréquence « type » de *net-* dans notre corpus espagnol ? Deux hypothèses sont envisageables :

- Les journalistes utilisent plutôt *web*, qui forme, cependant, des termes plus marqués géographiquement puisqu'ils renvoient à une partie du réseau seulement. Il y a souvent confusion entre Internet/ le Net et le Web.
- L'absence de composés avec *net-* peut être une spécificité de notre corpus. Mais comme la macrostructure du glossaire de Millán ne recense aucun composé formé avec *net-* nous ne pensons pas que ce soit le cas. Le substantif *net* n'y possède même pas sa propre entrée. Dans le glossaire de Rafael Fernández Calvo,¹⁴⁰ les termes anglais contenant *net* sont rendus en espagnol par d'autres structures : EN *net surfing* > ES *navegación por la red*, EN *netiquette* > ES *etiqueta de la red*, *cibernetiqueta*, *ciberurbanidad*, EN *netizen* > ES *ciudadano de la red*, *ciuredano*. Il semblerait que l'espagnol préfère la structure nom + *Internet / de la red*.

1.2.7. e-

Le formant e- a sa propre entrée dans le DRAE 1996 et le DGLE en tant que préfixe latin ayant trois acceptations : ‘hors de’ comme dans *eliminar*, ‘origine, provenance’ comme dans *emanar*, *emigrar*, ou ‘extension, dilatation’ comme dans *efusión* et *emoción*. E- est de la troncation de l'adjectif anglais *electronic* (Part. II, Ch. 2, Sec. 1.2.4). Dans les corpus espagnol et français, il ne s'agit pas du préfixe latin répertorié dans le PR 1996, le DGLE et le DRAE 1996.

Tableau 61 : Composés espagnols avec le pseudo-confixee-

¹⁴⁰ Source : <http://www.ari.es/novatica/glointv2.html>

Pseudo-confixe	Base	Champ sémantique	Base	Pseudo-confixe	Champ sémantique
e	biz	C		Surcomposés de:	
	<i>book</i>	P	chequeador de	e-mail	A
	<i>business</i>	C	cuenta		C
	<i>card</i>	C	dirección de		M
	<i>commerce</i>	C	junk		M
	<i>compra</i>	C	integrador	e-business	A
	<i>empleado</i>	Mt			
	<i>empleo</i>	Mt	programa de	e-commerce	A
	<i>learning</i>	Mt	sitio de		C
	mail	M			
	<i>market place</i>	C			
	<i>marketing</i>	M			
	<i>turismo</i>	C			
Surcomposés avec :					
email	comercial	M			

Légende : A = application, C = commerce, M = courrier, Mt = monde du travail, P = périphérique

Dans le corpus espagnol, e- entre dans la formation de 21 termes : 13 composés et 8 surcomposés (dérivés de *email*, *e-business* et *e-commerce*). Ce pseudo-confixe est donc bien moins fréquent qu'en anglais et en français, où il a produit respectivement 40 et 32 lexies. D'après Millán, l'anglais est beaucoup plus souple que l'espagnol en matière d'abréviation. En espagnol, l'adjectif suit le nom. On devrait donc dire et écrire **libro-e* et **comercio-e* et non *e-libro* et *e-comercio*. La langue préfère les composés N + *electrónico* comme *libro electrónico* et *comercio electrónico*. Ces composés N + A respectent l'ordre déterminé/déterminant. La structure N + *electrónico* est l'équivalent du modèle e + N. Notre corpus contient six termes répondant à ce modèle (*negocio electrónico*, *correo electrónico*, *bulo electrónico*). C'est une autre illustration de la résistance de l'espagnol aux emprunts et au changement. La structure e- + N est fréquente en français alors que, comme en espagnol, l'adjectif n'y est normalement pas antéposé.

D'après le tableau 61, e- est toujours le déterminant. L'ordre des éléments étant déterminant/déterminé, ces composés ne respectent pas l'ordre de l'espagnol. Quant aux surcomposés, ils présentent l'ordre déterminé/déterminant (sept termes sur huit) et sont endocentriques. Ce sont des constructions hybrides calquées sur les surcomposés anglais : *dirección de e-mail* > *email address*, *integrador e-business* > *e-business integrator*, etc. Une *dirección de e-mail* est une *adresse* et un *integrador e-business* est un *integrador*.

Les termes espagnols étant des emprunts et des constructions hybrides, e- désigne

toujours des activités du monde réel reproduites sur le Net. Les thèmes relevés sont eux aussi identiques à ceux des corpus anglais et français. Les termes désignent des acteurs, des occupations et des pratiques ne nécessitant pas de connaissances en informatique. Les surcomposés espagnols sont des calques et présentent la même particularité que les surcomposés anglais et français. Ils dénomment des personnes ou des logiciels qui accomplissent des tâches à la place de l'internaute : *integrador e-business*, *chequeador de e-mail*, etc.

1.2.8. -nauta

Le DRAE 2001 et le DGLE comprennent chacun une entrée *nauta*. Dans le DRAE, il s'agit d'un substantif signifiant 'homme de la mer' (« hombre del mar »). Le DGLE donne deux acceptions : *marinero*, *marino*, *navegante / calamento*.

Millán précise qu'à travers l'anglais de nouvelles formations comme *internauta* (« el que navega por la Internet ») ont fait leur apparition. Un *cibernauta* désigne « el que navega por el ciberespacio ». Mais l'auteur ne précise pas que dans le domaine d'Internet *-nauta* est la troncation d'*astronauta*. Le préfixe classique est relatif à la navigation (dans le sens de déplacement en mer) alors que le nouveau formant *-nauta* entre dans les dénominations d'utilisateurs (mais la métaphore de la navigation est fortement présente).

Tableau 62 : Composés espagnols avec le pseudo-confixe *-nauta*

Base	Pseudo-confixe	Champ sémantique
inter	nauta	U
jubil		U
Surcomposé avec :		
comunidad	internauta	U

Légende : U = utilisateur

-Nauta a une fonction suffixale. Les composés du tableau 62 sont endocentriques. Un *internauta* est un utilisateur d'Internet. Un *jubilnauta* (*jubilado* + *astro nauta*) est un utilisateur d'Internet à la retraite. Ce pseudo-confixe est toujours utilisé dans le sens d' 'utilisateur de'.

-Nauta conserve le sens de sa forme développée, il signifie 'utilisateur' et, par métaphore 'navigateur' mais il ne peut pas être remplacé par sa forme développée *astronauta*. Le patron syntaxique utilisant *-nauta* (N tronqué + *-nauta*) est plus innovant que ceux formés contenant les autres pseudo-confixes. Aucune voyelle ou syllabe de liaison n'est utilisée en espagnol : *cibernauta*, *infonauta*, *intranauta*.

1.2.9. -ware

Cet élément n'est inclus ni dans le DRAE, ni dans le DGLE, mais ce dernier possède une entrée pour *software*. La traduction espagnole de *software* est *programa* ou, sous l'influence du français, N + *lógico* comme *soporte lógico*. *Programa* semble être utilisé dans la sphère privée et *software* dans le monde de l'entreprise.¹⁴¹

Tableau 63 : Composés espagnols avec le pseudo-confixe *-ware*

Elément	Pseudo-confixe	Champ sémantique	Pseudo-confixe	Base	Champ sémantique
ad	ware	Sé	Surcomposés avec :		
free		A	software	compartido	A
group		A		de acceso a Internet	A
open		A		de filtrado	Sé/A
share		A		de navegación	A
spy		Sé		de servidor	A
	Surcomposés avec :			espía	Sé
firewall	hardware	A			
firewall	software	A			
programa	spyware	Se/A			
Légende : A = application, Sé = sécurité					

Dans le tableau 63, *software* entre dans la composition de nombreux termes hybrides. D'après Millán, cette lexie a fait son apparition dans la langue espagnole en 1967 et, depuis, ce terme anglais dont la prononciation est adaptée à la phonologie espagnole (/sóftguar/) est fréquemment employé. Aucun patron espagnol n'a été créé comme alternative à déterminant + *-ware* (comme déterminant + *-iciel* en français, par exemple).

Dans le glossaire de Millán, qui explique les termes anglais, *freeware* a un synonyme : le composé hybride *software público*. Pourquoi l'espagnol ne cherche-t-il pas d'équivalents à ces emprunts anglais ? Nous formulons deux hypothèses :

- Pour respecter l'ordre déterminé/déterminant, les internautes forment des composés hybrides (*software compartido*, *programa spyware*, *software espía*).
- L'espagnol a emprunté *software* et l'utilise pour traduire les termes avec *-ware* (qui en est la forme réduite). Les internautes espagnols estiment que cet emprunt est suffisant (*shareware* > *software lógico*, *spyware* > *software espía* et *freeware* > *software libre*).

1.2.10. Conclusion

Neuf pseudo-confixes fréquents sont recensés dans le corpus espagnol (*ciber-*, *e-*, *hiper-*, *meta-*, *-nauta*, *net-*, *tele-*, *-ware*, *-web-*). Ce sont les emprunts de troncations de lexies anglaises. Certains sont homographes de confixes classiques (*ciber-*, *e-*, *hiper-*, *meta-*, *tele-*, *-nauta*). L'espagnol, comme le français, réemploie les formants anglais dont la

¹⁴¹ Source : http://www.jamillan.com/v_index.htm

fréquence « type » varie beaucoup :

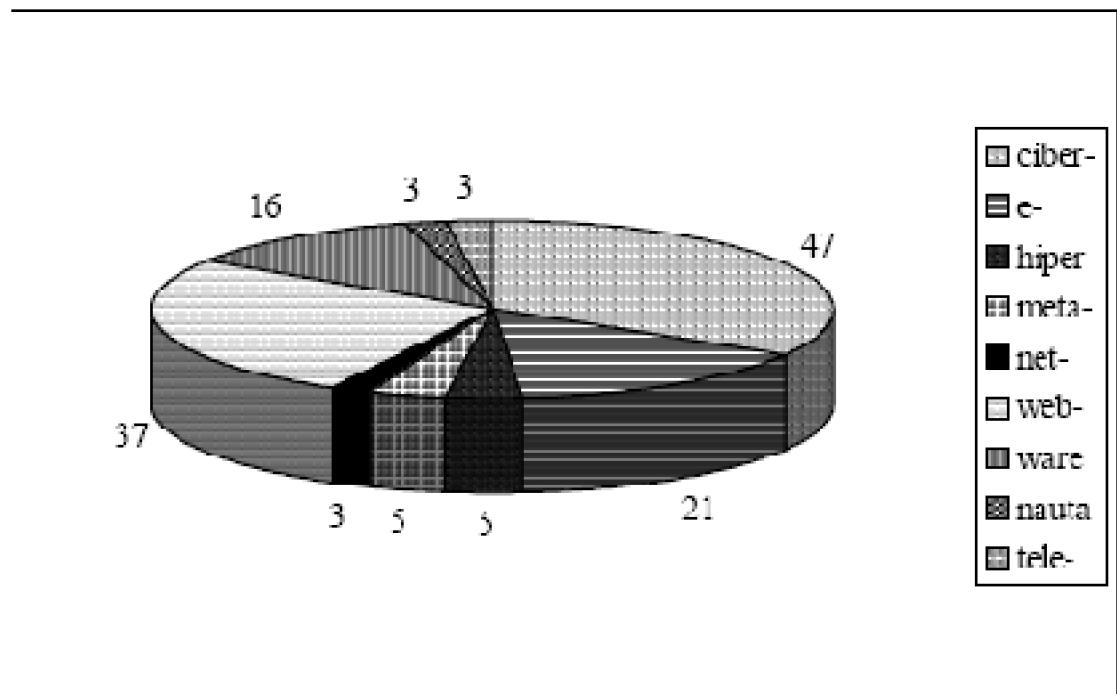


Figure 5 : Fréquence des pseudo-confixes espagnols

Les pseudo-confixes qui forment le plus de termes sont les mêmes qu'en anglais et français. *Ciber-* et *web* sont les formants les plus fréquents avec respectivement 47 et 37 composés. *E-* arrive en troisième position mais sa fréquence « type » est bien inférieure (21 termes). Ce formant est moins fréquent qu'en anglais (40 termes) et français (32 termes) car l'espagnol est plus réfractaire aux emprunts et préfère la structure N + *electrónico*. Les autres pseudo-confixes sont moins fréquents : *-ware* (16 termes), *hiper-* (5 termes), *meta-* (5 termes), *-nauta* (3 termes), *net-* (3 termes), *tele-* (3 termes). La raison des écarts de fréquence entre les pseudo-confixes a été développée pour l'anglais et le français (Part. II, Ch. 2, Sec. 1.2.8. / Ch. 3, Sec. 1.2.10) : les pseudo-confixes sont employés dans des champs sémantiques précis et certains champs notionnels, comme celui de *ciber-*, sont beaucoup plus vastes que d'autres, comme celui de *meta-*.

Les formants antérieurs, qui occupent en général une position de déterminant, sont plus nombreux et plus fréquents que les formants postérieurs. Par conséquent, la majeure partie des composés pseudo-confixés est contraire à l'ordre déterminé/déterminant des composés espagnols. Du point de vue morphosyntaxique, l'influence de l'anglais n'est sentie que dans le cas des emprunts des troncations morphosémantiques. Les pseudo-confixes sémantiques ont la même structure que les confixes classiques dont ils sont homographes.

L'entrée dans les dictionnaires de ces nouveaux formants à fréquence « type » élevée est encore moins systématique pour l'espagnol que pour le français, y compris pour les termes dont la fréquence « token » est élevée comme *internauta*, *Internet*, *cibercrimen*. Seul le DGLE recense quelques termes. Les mêmes hypothèses que pour le français se dessinent (Part. II, Ch. 3, Sec. 1.2.10.) mais la troisième hypothèse, exclusion

des emprunts, a plus de poids dans le cas de l'espagnol. Les ouvrages lexicographiques rejettent les termes qui sont trop « anglais » (surtout le DRAE). Le DGLE, ressource électronique, répertorie moins de pseudo-confixes et de termes pseudo-confixés que le GR 2001, version papier.

2. Le profil sémantique des noms composés collectés

2.1. La centricité

Les critères pour qu'un composé soit considéré comme endocentrique ont été exposés précédemment (Part. II, Ch. 2, Sec. 2.1.).

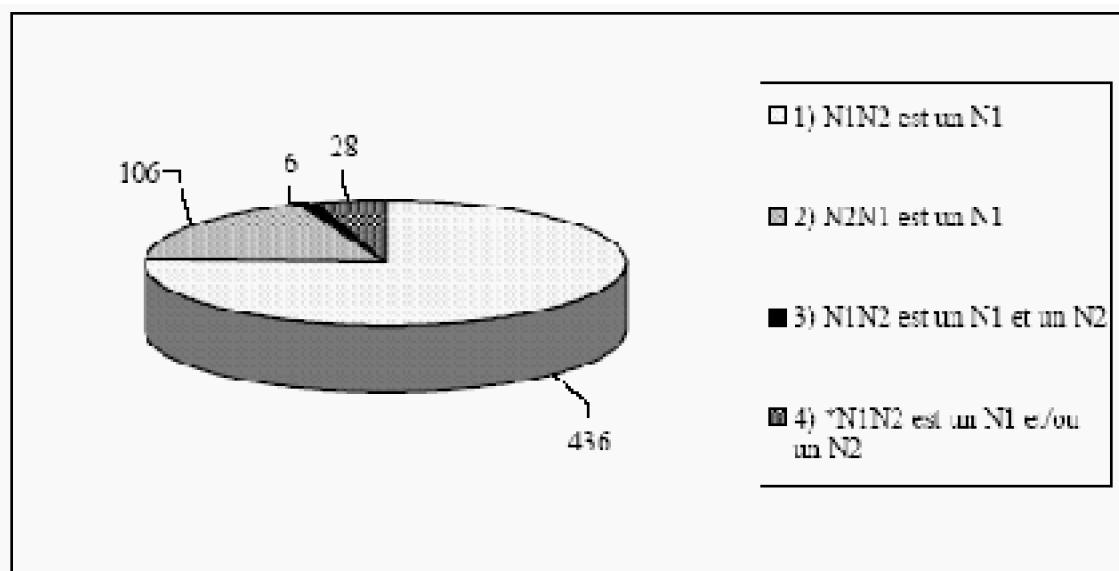


Figure 6: Centricité des composés espagnols du corpus

Tout comme dans les corpus anglais et français, les composés endocentriques abondent dans le corpus espagnol. Avec 95,13% du corpus, ils sont autant représentés qu'en français (94%) et en anglais (94,57%). Avec 4,87%, les composés exocentriques sont présents dans des proportions similaires (6% pour le français et 5,43% pour l'anglais).

La motivation des composés endocentriques ainsi que le flot de métaphores concourent à la prédominance de ce type de composés.¹⁴² En espagnol, comme en français, le type endocentrique le plus fréquent est 'N1N2 est un N1' puisque l'ordre classique de formation des composés est déterminé/déterminant. Les composés endocentriques répondant au modèle 'N2N1 est un N1' sont plutôt nombreux (18,40%) pour une langue romane en raison de la présence des emprunts anglais (*bookmark*, *message board*, etc.) et des composés pseudo-confixés (*cibereditor*, *net-economía*, *metabuscador*, etc.).¹⁴³

Quant aux composés exocentriques, modèle 'N1N2 est un N1 et/ou un N2', le corpus

¹⁴² Part. II, Ch. 2, Sec. 2.1. pour l'analyse détaillée.

espagnol en contient 4,87%. Ils sont, comme dans les deux autres langues étudiées, partiellement ou totalement métaphoriques ou métonymiques (*agente inteligente*, *puerta trasera*, *puntocom*, etc.).

Dans le système d'équivalences types, les composés exocentriques constitueront un obstacle puisque les composés anglais sont rendus en espagnol par un emprunt total ou partiel (EN *temporary cookie* = ES *cookie temporal*, EN *firewall* > ES *firewall*, etc.) ou par un emprunt interne comme dans la langue source (EN *backdoor* > ES *puerta trasera*, EN *firewall* > *cortafuegos*, etc.).

2.2. Les relations sémantiques

La démarche utilisée pour étudier les relations déterminé/déterminant des noms composés est détaillée dans le chapitre consacré à l'anglais (Part. II, Ch. 2, Sec. 2.2.). Comme pour l'analyse des composés français, nous réemploierons les paraphrases mises en place dans le chapitre 2 lorsque ce sera possible. Le poids des traits conceptuels dans le domaine d'Internet sera, en espagnol aussi, évalué pour déterminer la relation N1-N2 des noms composés métaphoriques.

Tableau 64 : Synthèse des relations sémantiques des composés espagnols du corpus

¹⁴³

Les pseudo-confixes occupent pour la plupart une position préfixale, et précèdent donc le déterminé. (Part. II, Ch. 4, Sec. 1.2.)

Code	Description	Exemples	Fréquence « type »
1	N1 est désigné par analogie à une propriété de N2	<i>servidor proxy, sitio mirror</i>	6
2	N1 s'adresse à N2 (e-commerce, échange)	<i>empresa a empresa, peer-to-peer</i>	5
3a	N1 (logiciel, matériel, technique) a pour activité, fonction N2	<i>bloqueador de popup, gestor de descarga</i>	86
3b	N1 (humain, entreprise) a pour object d'activité, fonction N2	<i>administrador de red, proveedor de correo</i>	28
3c	N1 (virus, attaque, pirate) a pour fonction N2	<i>ataque DoS, ataque de denegación de servicio</i>	11
4	N1 a un fonctionnement dont le procédé, le mode est N2	<i>correo electrónico, conexión de alta velocidad</i>	110
5	*N1N2 est un N1 et/ou un N2	<i>agente inteligente, puerta trasera</i>	21
6	N1 se distingue des autres N1 car il est N2 (nature)	<i>ADSL individual, hacker blanco</i>	99
7	N1 a N2 comme localisation, environnement	<i>ciberamor, ciberconferencia</i>	68
8	N1 est combiné à N2	<i>jubilnauta, hactivismo</i>	6
9	N1 est engendré, provoqué par N2	<i>ataque vírico, infección de virus</i>	11
10	N1 est constitué, formé par N2	<i>comunidad hacker, red de redes</i>	7
11	N1 évalue, mesure N2 (durée, vitesse, nombre)	<i>sesión de charla, velocidad de conexión</i>	8
12	N1 est transféré en direction N2 / N1 a N2 comme destination	<i>download</i>	1
13a	N1 est un instrument, mode, moyen utilisé par N2	<i>cuenta de usuario, interfaz de usuario</i>	3
13b	N1 est un instrument, mode, moyen utilisé pour N2	<i>barra de direcciones, protocolo POP</i>	104
14	omission	<i>no conectado, recién conectado</i>	2
	Total		576

Quatorze relations sémantiques unissent les éléments des 576 noms composés espagnols. De manière générale, l'espagnol ne se démarque pas des deux autres langues étudiées : les relations N1-N2 sont les mêmes. Deux termes espagnols (*no conectado, recién conectado*) et un terme français (*non connecté*) présentent une relation non présente en anglais : 14 'N1 est omis, N2 est autonome'. Notre étude étant basée sur

un corpus clos, cela ne signifie pas que cette relation est inexisteante en anglais.

Les raisons avancées pour expliquer la proximité des résultats entre l'anglais et le français sont valables pour l'espagnol :

- Pour les composés qui sont des créations exogènes, les relations sont bien évidemment identiques à celles observées dans les termes anglais.
- Les calques sont nombreux. Les relations N1-N2 des composés anglais sont conservées dans les calques espagnols. Dans EN *email alert* et ES *alerta de correo* la relation est 9 'N1 est engendrée, provoquée par N2'. La relation 2 'N1 s'adresse à N2 (e-commerce, échange)' est au cœur de EN *business-to-business* et ES *empresa a empresa*.

Comme en anglais et en français, les composés dont la relation N1-N2 ne peut pas être décrite autrement que 'N1N2 est un ...' sont regroupés dans la catégorie 5 '*N1N2 est un N1 et/ou un N2'. En espagnol aussi, ces termes sont majoritairement exocentriques (21 composés sur 30), comme *agente inteligente*, *banda ancha* ou *cortafuegos*. Ces équivalents espagnols sont soit des emprunts (11 cas), soit des calques des termes anglais (10 cas).

Le tableau 65 résume les relations les plus fréquentes en espagnol et présente les chiffres obtenus pour l'anglais et le français, à titre comparatif :

Tableau 65 : Relations sémantiques fréquentes

Relations sémantiques	Espagnol	Anglais	Français
3a 'N1 (logiciel, matériel, technique) a pour activité, fonction N2', 3b 'N1 (humain, entreprise) a pour object d'activité, fonction N2', 3c 'N1 (virus, attaque, pirate) a pour fonction N2'	21,70%	23,33%	25,12%
4 'N1 a un fonctionnement dont le procédé, le mode est N2'	19,10%	18,81%	18,18%
6 'N1 se distingue des autres N1 car il est N2 (nature)'	17,19%	14,10%	15,25%
7 'N1 a N2 comme localisation, environnement'	11,81%	13,38%	11,86%

En conclusion, la prépondérance des calques et des emprunts en espagnol, comme en français, explique que les chiffres ci-dessus soient proches. Les quatre relations N1-N2 les plus fréquentes sont identiques dans les trois langues et leur poids est constant (moins de 4% d'écart entre le pourcentage le plus faible et le plus fort). Il nous restera à déterminer si une ou plusieurs relations sémantiques peuvent être à l'origine des cas d'échecs du système d'équivalences types anglais-espagnol.

3. la combinatoire intermatricielle

Dans le lexique espagnol d'Internet, les termes sont fréquemment formés par plusieurs matrices lexicogéniques. Selon nos critères (Part. II, Ch. 2, Sec. 3.2.1. et 3.2.4.), parmi les 576 composés du corpus espagnol, seulement 52 (9,03%) sont formés par un seul processus.

3.1. Analyse

Tableau 66 : Combinaisons des matrices du corpus espagnol

Les noms composés anglais français et espagnols du domaine d'Internet :

Rang	Combinaisons	Fréquence « type »	%	Exemples
1	C M	127	22,2%	<i>agente inteligente</i>
2	C M M	100	17,4%	<i>administrador de red</i>
3	C Ps	59	10,2%	<i>cibercrimen</i>
4	C Ps M	46	8,0%	<i>ciberalbum</i>
5	C M C	43	7,5%	<i>alojamiento de sitio web</i>
6	C M Si	34	5,9%	<i>acceso FTP</i>
7	C C	29	5,2%	<i>mensaje de correo electronico</i>
8	M	25	4,3%	<i>lista negra</i>
9	C	21	3,6%	<i>negocio a negocio</i>
10	C A	14	2,4%	<i>café Internet</i>
	C M A			<i>navegador de Internet</i>
11	C Si	7	1,2%	<i>lenguaje HTML</i>
12	C A C	6	1,0%	<i>Internet de alta velocidad</i>
	A			<i>modem</i>
	C M Me			<i>proveedor de antivirus</i>
13	C C C	5	0,9%	<i>conexion permanente de banda ancha</i>
	C Me			<i>click and mortar</i>
14	C M S M	4	0,7%	<i>bloqueador de pop up</i>
15	C M S	3	0,5%	<i>pagina hackeada</i>
	C Ps Ap			<i>webcam</i>
16	C Me Ap	2	0,3%	<i>dot org</i>
	C S Ps			<i>desarollador Web</i>
	C Si Me			<i>voz sobre IP</i>
	C A Si Me			<i>modem ADSL</i>
	C Ap			<i>empresa.com</i>
17	C M Si Me	1	0,2%	<i>sitio MP3</i>
	C Ps Ps			<i>cibernauta</i>
	A M Ah			<i>shopbot</i>
	A Ps			<i>jubilnauta</i>
	C C Ps			<i>interfaz grafica web</i>
	C P			<i>protección antispam</i>
	C M P M			<i>programa anti popup</i>
	Total	576	100,0%	

Les schémas combinatoires à l'origine de la création des noms composés espagnols sont au nombre de trente-deux. En espagnol, comme en anglais et en français, entre une et quatre matrices (*módem / módem ADSL*) interviennent dans la formation de chaque composé du corpus. En moyenne, le corpus espagnol comprend 18 composés par schéma combinatoire. Neuf combinaisons sont supérieures à cette moyenne (C M, C M M, C Ps, C Ps M, C M C, C M Si, C C, M, C) avec 84,3% du total des noms composés.

La composition, présente dans 95,7% des termes du corpus espagnol, est récurrente

dans huit combinaisons (C M C, C C, C A, C M A, C A C, C C C, C A Si Me, C C Ps), c'est-à-dire pour 19,8% des termes. La composition est le plus souvent combinée à la métaphore. Dans 66,2% des composés, au moins un élément métaphorique est recensé. La métaphore est récurrente dans trois schémas (C M M, C M S M et C M P M), soit 18,3% des composés espagnols. Avec 3% des composés du corpus espagnol, la métonymie est beaucoup moins représentée que la métaphore.

Après la métaphore, la pseudo-confixation est la matrice qui se combine le plus fréquemment avec la composition. Six combinaisons contiennent au moins un pseudo-confixe (19,1% des composés). Cette matrice a une récursivité très faible. Un seul terme constitué de deux pseudo-confixes a été relevé (*cibernauta*).

3.2. Bilan

Tableau 67 : Fréquence des combinaisons de matrices dans le corpus espagnol

Combinatoire	%	Fréquence « type »	% anglais	% français
Composition + métaphore	66,2	12	64,5	65,5
Composition + pseudo-confixation	19,1	8	21,5	21,4
Composition + siglaison	7,9	5	7,5	8
Composition + métonymie	3	6	3,4	3,6
Composition + apocope	1,1	3	0,6	1,2
Composition + aphérèse	0,2	1	0,2	0,9
Composition + suffixation	1,5	2	0,2	1,4
Composition + préfixation	0,2	1		0,5

Les pourcentages du tableau 67 sont semblables à ceux observés pour l'anglais et le français. Seule la combinaison composition-suffixation offre une fréquence plus élevée en espagnol et en français. Certains emprunts anglais deviennent productifs : *sitio hackeado*, *chequeador de mail*, *diseñador web*, etc.

La même question que pour le français se pose : la présence des métaphores observées en espagnol est-elle due à une influence de l'anglais ? Certaines des métaphores conservées sont opaques pour le locuteur espagnol (*hacker blanco*, *firewall*, etc.). Comme pour le français, plusieurs éléments peuvent expliquer l'emprunt de lexies opaques :

- Les emprunts sont monosémiques. En Espagne, il n'y a pas d'autre sens pour *postmaster* que celui utilisé dans le cadre d'Internet (« persona responsable de solucionar problemas en el correo electrónico, responder a preguntas sobre usuarios y otros asuntos de una determinada instalación »).
- La prononciation de l'emprunt est adaptable au système phonologique espagnol. EN *software* est prononcé /sCftgwar/ et *spam* /espam/. Certains verbes se conjuguent comme les verbes espagnols : EN *to hack* > ES *hackear*, EN *to chat* > *chatear*.

Les noms composés anglais français et espagnols du domaine d'Internet :

Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol

Chapitre 1. Pré-enquête

Dans cette troisième partie, l'étude de la traduction des noms composés anglais du domaine d'Internet en français et en espagnol a pour but de :

- fournir au traducteur et rédacteur spécialisés les informations nécessaires à une bonne traduction des noms composés,
- mettre en place un système d'équivalences types multilingue d'aide à la traduction des noms composés.

Dans le premier chapitre, une ‘enquête de terrain’ à travers l'étude de la traduction du lexique d'Internet, du sens de circulation des termes du domaine et un test de faisabilité des équivalences permettra de déterminer la nécessité de la mise en place d'un outil d'aide à la traduction des noms composés. Le deuxième chapitre présentera la manière dont les équivalences anglais-français et anglais-espagnol seront créées, détaillera les

cas d'échec de ce système et examinera l'efficacité des outils de traduction automatique en matière de noms composés. Le troisième chapitre retracera l'étape ultime, l'automatisation du système d'équivalences types.

1. La traduction du lexique d'Internet

Dans le domaine d'Internet, les concepts, les nouvelles technologies et les produits (logiciels de téléchargement, de sécurité, d'aide à la navigation, sites Web, langages de programmation, jeux en ligne, etc.) sont développés par des entreprises américaines. Et comme tout concept conçu au sein d'une communauté linguistique se traduit par la création des signes linguistiques nécessaires à sa dénomination, les pays technologiquement plus faibles – les pays francophones et hispanophones ici – ont un volume considérable de termes anglais (et potentiellement de textes) à traduire. Les spécialistes d'Internet français et espagnols consultent de nombreux ouvrages qui sont des traductions de l'anglais. Sur quarante et un livres ayant pour thème la création de sites Web (recherche effectuée sur le SCD de la BU de Lyon 2 le 09 juillet 2004 avec le mot clé « création site web »), quatorze sont des traductions (plus d'un tiers). Par ailleurs, la rapidité d'apparition des concepts entraîne un besoin urgent de traduction, ou au moins d'adaptation, des termes liés à ces concepts pour les pays importateurs. Les traducteurs spécialisés sont peu nombreux par rapport au volume de traduction. Pour pallier ce manque, des outils d'aide à la traduction sont nécessaires.

1.1. Les difficultés de traduction des noms composés

Le nombre considérable de composés¹⁴⁴ et leur fréquence élevée dans les textes engendrent des difficultés de traduction. Les trois étapes nécessaires à la traduction d'un nom composé ou autre sont parsemées d'obstacles : identification, compréhension et mise en rapport du composé anglais avec son (ou ses) équivalent(s) de traduction en langue cible (lorsqu'il en existe). Un chapitre entier étant consacré à l'identification des noms composés (Part. I, Ch. 1), nous commencerons par examiner la compréhension des composés.

1.1.1. La compréhension des noms composés

1.1.1.1. Le décodage des noms composés

Internet est un domaine hautement métaphorique. Or, plus un domaine de spécialité est métaphorique, plus le décodage des relations entre les éléments des composés est difficile. Dans EN *address bar* la relation N1-N2 est 'N1 est le moyen utilisé pour saisir N2'. L'équivalent français *barre d'adresses* rend cette relation. Dans le cas des composés exocentriques, les relations ne sont pas analysables autrement que par une paraphrase très générale (*N1N2 est un N1 et/ou un N2). Peut-être serait-il judicieux de ne pas

¹⁴⁴ Dans les glossaires spécialisés, 80% de la terminologie d'un domaine de spécialité est composé de groupes de noms. Ce chiffre varie peu d'un domaine à un autre (Béjoint, 1988 : 356). Les noms composés en font partie.

prendre en compte les relations sémantiques N1-N2 lors de la traduction des exocentriques.¹⁴⁵ En effet, dans le domaine d'Internet, dire que dans EN *thumbnail* la relation est 'N1 est situé sur N2' est absurde et inutile pour créer un équivalent. Le rôle didactique de la métaphore, par une analogie de taille, est probablement à l'origine de l'utilisation de EN *thumbnail* pour désigner le concept /image d'une page web de la taille d'un timbre-poste sur laquelle l'internaute peut cliquer pour accéder à une version agrandie/.

1.1.1.2. Les reformulations et les anaphores

Les termes complexes sont parfois repris dans les textes par leur hyperonyme. Plus un terme est long, plus tôt dans le texte il sera repris par son hyperonyme. Cela répond à la loi d'économie linguistique. Le coût linguistique est en rapport direct avec la longueur des termes (Tournier, 1991b : 63). Plus le terme utilisé est court, plus l'effort articulatoire et sans doute mémoriel est réduit. On prononce, écrit et mémorise plus facilement EN *ADSL* que *Asymmetric Digital Subscriber Line*. Les choses se compliquent lorsque plusieurs composés et surcomposés ont le même hyperonyme. Imaginons la publication d'un article comparatif sur les différents types de fournisseurs dans le domaine d'Internet. Les termes EN *antivirus provider*, *ADSL provider*, *space provider*, *Application Service Provider*, *Internet Access Provider*, *Internet content provider*, *Internet Service Provider*, *Managed Service Provider*, *email provider*, *Online Service Provider* désignent tous des fournisseurs et ont tous *provider* comme hyperonyme, ce qui peut poser des problèmes d'identification du concept à traduire au traducteur.

Toujours en rapport avec la loi d'économie linguistique, dans les textes, les surcomposés de plus de trois éléments sont remplacés par leur sigle. Par exemple, *HTTP* remplace *HyperText Transfer Protocol*, *SMTP* remplace *Simple Mail Transfer Protocol* et *SOHO* remplace *Small Office Home Office*. Avant la première traduction ou le premier emprunt d'un sigle, le traducteur cherchera le terme complet pour identifier le concept.

1.1.2. La mise en correspondance avec un équivalent de traduction de la langue cible

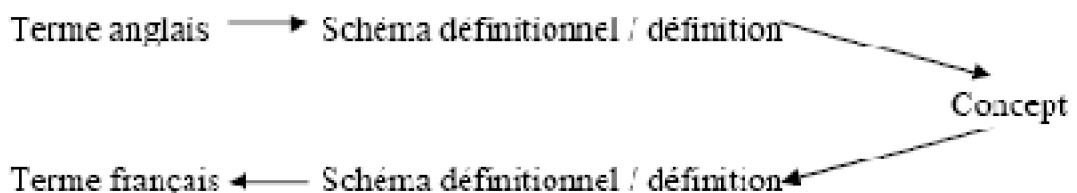
Nous commencerons par répondre à deux interrogations concernant la notion d'équivalence terminologique : Qu'est-ce qu'un équivalent terminologique ? Comment créer de bons équivalents terminologiques français et espagnols dans le domaine d'Internet ?

Des termes équivalents désignent des concepts homologues dans plusieurs langues (Thoiron et al., 1996 : 512). Dans le domaine d'Internet, la traduction des termes d'Internet est une affaire de communauté. Les internautes et les traducteurs procèdent à la « terminologisation traductive » (Brownlie, 2002 : 298) des termes, c'est-à-dire produisent des termes cibles qui seront réutilisés chaque fois que le terme source apparaîtra. Les équivalents français et espagnols d'Internet, contrairement à la plupart des autres terminologies techniques et scientifiques, ne sont pas créés et normés par une

¹⁴⁵ Pour le déterminer, un test sera effectué ultérieurement (Part. III, Ch. 2, Sec. 2.).

commission de spécialistes mais par les entreprises, les internautes et les traducteurs.

La démarche terminologique pour créer un équivalent est constituée de plusieurs étapes. Le terminologue observe le terme anglais et cherche dans des sources unilingues anglaises le concept désigné (sémasiologie). Une fois le concept identifié, il consulte des sources unilingues françaises et espagnoles pour trouver les termes français et espagnols désignant ce même concept (onomasiologie).



Voici une définition du terme EN *chat room* et de FR *salon de chat* et ES *sala de charla* ses équivalents supposés :

EN « *A variation on the interactive message board, it is a website for live, online conversation in which any number of computer users can type messages to each other and communicate [...]* ». (www.netlingo.com) **FR** « *Lieu de rencontre virtuel, accessible à partir d'un site de bavardage, que l'internaute peut choisir, selon le sujet proposé, afin de converser en temps réel et par clavier interposé, avec un nombre relativement restreint de participants* » (www.granddictionnaire.com) **ES** « *Lugar virtual de la red, llamado también canal, donde la gente se reúne para charlar con otras personas que hay en la misma sala* » (www.ati.es)

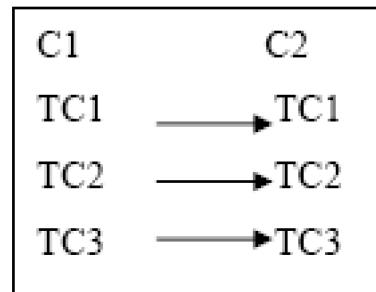
Dans les définitions ci-dessus, une partie des informations données est présente dans le terme (lieu + finalité). Les éléments nouveaux concernent les moyens utilisés pour atteindre la finalité. Les trois définitions sont basées sur un schéma définitionnel homologue :

	Anglais	Français	Espagnol
Nature	site web	lieu virtuel	lieu virtuel
Finalité	discuter à plusieurs en temps réel	discuter à plusieurs en temps réel	discuter à plusieurs
Moyen	saisie d'un message	saisie d'un message	

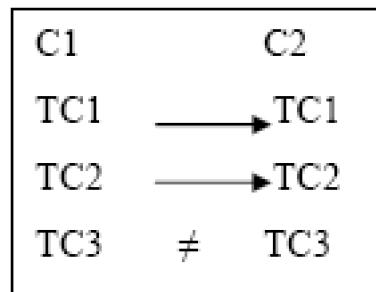
Dans cet exemple, en anglais, en français et en espagnol, des traits conceptuels homologues sont utilisés pour la dénomination et les équivalents français et espagnol comportent les mêmes éléments de nomination que le terme anglais mais ce n'est pas toujours le cas. La traduction des éléments de nomination du terme EN *bottleneck* donnerait un terme existant en français, *goulot de bouteille*, mais qui, dans le domaine d'Internet, ne correspond pas au concept désigné par le terme anglais. Le terme équivalent est FR *goulot d'étranglement*. Recourir au schéma définitionnel lorsqu'on sélectionne un équivalent en langue cible, que les langues source et cible sont proches et que le terme source est métaphorique permet d'éviter d'accorder trop d'importance à la forme du terme et de traduire par un faux ami terminologique.

On l'a vu, deux termes sont équivalents lorsqu'ils sont similaires sur le plan conceptuel. Il existe plusieurs degrés d'équivalence conceptuelle dont Arntz propose une classification (1993 : 13-14) qui sera illustrée par des schémas :

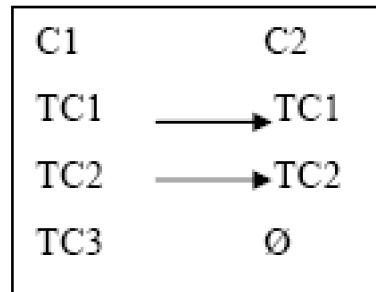
- Équivalence conceptuelle totale. Exemple¹⁴⁶ : C1 EN *bookmark* et C2 FR *signet*



- Chevauchement conceptuel. Exemple : C1 EN *informatics* et C2 FR *informatique*



- Inclusion. Exemple : C1 EN *pet* et C2 FR *animal*



- Absence d'équivalence conceptuelle. C1 EN *location* et C2 FR *location*

¹⁴⁶ C = concept, TC = trait conceptuel.

C1		C2
TC1	≠	TC1
TC2	≠	TC2
TC3	≠	TC3

Lorsque les termes anglais n'ont pas d'équivalents de traduction préexistants en langue cible, Arntz (1993 : 15-16) propose trois procédés de traduction : l'emprunt ou la traduction littérale, la traduction non littérale et la paraphrase. En comparant les termes anglais et les termes français et espagnols équivalents dans le corpus, il est manifeste que les spécialistes d'Internet optent de préférence pour la première solution. Ils privilient les dénominations présentant une similitude de signifiants d'une langue à l'autre (EN *virus definition*, *mailbox* / FR *définition de virus*, *boîte aux lettres*). Ils semblent aspirer à l'internationalisme. Pour créer des équivalents dans le domaine d'Internet ayant un bon potentiel d'implantation, il faudra respecter cette préférence des internautes qui vont réutiliser les termes créés.

Si un équivalent terminologique en langue cible doit être créé, plusieurs faits devront être considérés pour fournir aux internautes des équivalents reconnaissables et pour assurer la stabilisation de la terminologie du domaine :

- Un terme anglais doit être traduit par un terme en français et en espagnol et non par une paraphrase explicative.
- Une même traduction devra être maintenue lors de chaque apparition du terme anglais et autant que possible lors du réemploi d'un formant dans un autre composé ou dans un surcomposé. Par exemple, le néologisme EN *bookmark converter* formé avec *bookmark* qui a déjà un équivalent en langues cible, aura pour équivalent FR *convertisseur de signets* ou *convertisseur de favoris*. La stabilisation des termes traduits garantit la compréhension des internautes, l'efficacité de la communication et la dissémination des concepts dans les cultures cibles.

D'autre part, un emprunt à l'anglais (FR *webmaster*, ES *firewall*), l'emploi d'un néologisme (FR *bande passante*, ES *cibercafé*), un emprunt interne (FR *virus polymorphique*, ES *acuse de recibido*) indiquent au lecteur que le concept est nouveau dans la culture cible et qu'il s'agit d'un terme.

Lors de la création d'un équivalent, un facteur important doit être pris en compte : la qualité de l'équivalent terminologique. Selon Rondeau (1980), la qualité d'un bon terme doit être évaluée sur deux plans :

- Le plan scientifico-technique. La dénomination X doit correspondre à la notion x dans l'esprit des spécialistes. Si plusieurs dénominations existent pour un même concept, sont-elles de vrais synonymes terminologiques ? Par exemple, les termes FR *mél*, *email* et *courriel* désignent tous le concept /courrier électronique/.

- Le plan linguistique. Il faut considérer non seulement la qualité linguistique interne, c'est-à-dire la conformité de la dénomination aux règles morphologiques, syntaxiques, graphiques de la langue, mais aussi la valeur notionnelle comparée. Par exemple, les termes FR *fournisseur d'accès Internet* et ES *proveedor de acceso a Internet* sont les équivalents de EN *Internet access provider* puisqu'ils dénomment un concept homologue /entreprise qui fournit à des utilisateurs des connexions permettant d'accéder aux services disponibles sur Internet/.

Dans le domaine d'Internet, d'autres critères garantissent l'implantation d'un équivalent :

- la brièveté. FR *haut débit* et ES *lista negra* contiennent trois et quatre syllabes, ce qui est assez peu pour des composés.
- la transparence. FR *logiciel de courrier* et ES *virus de arranque* sont transparents pour les internautes francophones et hispanophones puisque les lexies FR *logiciel*, *courrier*, ES *virus* et *arranque* sont implantées et fréquentes. Les déterminés/*logiciel* et *virus* entrent déjà dans la formation de composés : ES *virus de programa*, *virus de correo*, FR *logiciel libre*, *logiciel de chiffrement*.
- L'emprunt interne. Les termes sont souvent des lexies existantes auxquelles a été attribué un nouveau sens (FR *moteur* dans *moteur de recherche*, *moteur d'indexation*, *moteur à assistance humaine*, ES *portal* dans *portal específico*, *portal generalista*, *portal Internet*, *portal temático*) ou un sens plus spécifique (FR *adresse* dans *adresse électronique*, *adresse e-mail*, *adresse IP*, ES *alojamiento* dans *alojamiento de sitio web*, *empresa de alojamiento web*).
- la productivité. Dans le lexique d'Internet, les modèles FR *fournisseur + de + N* et ES *servidor + de + nom*, ont créé de nombreux termes : FR *fournisseur d'hébergement*, *fournisseur d'accès*, *fournisseur de contenu*, ES *servidor de chat*, *servidor de correo*, *servidor de Internet*, *servidor de noticias*.
- l'hybridation. De nombreux composés FR *protection antispam*, *netsurfeur*, *site hacké* et ES *banner publicitario*, *buzón de mail*, *firewall personal* contiennent des emprunts anglais.

1.1.2.1. La transparence translinguistique

Dans le cas du vocabulaire existant, nous l'avons vu, en recourant à la traduction par transparence le traducteur peut commettre une erreur. Les traductions françaises et espagnoles d'un composé anglais peuvent être des composés traduits littéralement (EN *user interface* > FR *interface utilisateur*, EN *search engine* > ES *motor de búsqueda*), des lexies simples (EN *bookmark* > FR *signet*, *favori*, EN *history folder* > FR *historique*, EN *password* > ES *contraseña*), ou des surcomposés (EN *ad blocker* > FR *filtre anti pub*, ES *programa anti pop-up*). Dans le dernier exemple, une traduction par transparence ne correspondrait pas à l'équivalent employé en langue cible.

1.1.2.2. Le niveau de transparence des équivalents de traduction des noms composés anglais

Blanco (1997 : 2) établit une typologie de la traduction des noms composés français-espagnols adaptable aux noms composés anglais-français et anglais-espagnols du lexique d'Internet. Lors de la mise en correspondance d'un nom composé anglais avec son ou ses équivalents français et espagnols, on s'aperçoit que le niveau de transparence de l'équivalent est variable :¹⁴⁷

- Les signes qui constituent le composé français ou espagnol sont considérés comme les équivalents de traduction immédiats des signes du composé anglais : EN *domain name* > FR *nom de domaine*, ES *nombre de dominio*, EN *anonymous FTP*, FR *FTP anonyme*, ES *FTP anónimo*. C'est le niveau de transparence le plus fréquent dans le corpus.
- Les signes du composé français ou espagnol peuvent être des équivalents de traduction des signes anglais, mais une traduction littérale serait inexacte : EN *digital signature* > FR *signature numérique*, et non **signature digitale*, EN *malicious code* > FR *code malveillant*, et non **code malicieux*.
- Au moins un des signes du composé français ou espagnol ne peut pas être vu comme un équivalent de traduction du signe anglais correspondant : EN *fire wall* > FR *coupe-feu*, EN *guest book* > ES *libro de visita*. La traduction par transparence conduirait à l'erreur : EN *thumbnail* > FR **ongle de pouce*, EN *smurf attack* > ES **ataque pitufo*).

1.1.2.3. Variation du nombre d'un formant

Le déterminant d'un nom composé anglais singulier peut être pluriel en français et en espagnol. C'est le cas dans les composés anglais N + N où le premier nom joue le rôle d'un adjectif. Il est alors généralement au singulier et reste invariable : EN *address book* > FR *carnet d'adresses*, ES *libreta de direcciones, agenda de direcciones*.

1.1.2.4. Cumul des difficultés

Les difficultés énumérées ci-dessus peuvent être cumulées. Par exemple, un nom composé anglais N + N dont les deux éléments sont singuliers peut être traduit par un N + à + N avec un élément pluriel en français ou en espagnol : EN *mailbox* > FR *boîte aux lettres, *boîte de courrier*.

Si le travail d'identification, de compréhension et de mise en correspondance des termes anglais avec leurs équivalents français et espagnols ou de création d'un équivalent était effectué en amont et mis à disposition du traducteur spécialisé, cela augmenterait sa productivité.¹⁴⁸

1.2. L'acquisition de connaissances en traduction

¹⁴⁷ Humbley (1998) a mené une étude similaire sur le degré de rapprochement des composés anglais et français dans le domaine de l'informatique.

¹⁴⁸ Pour être optimum, le système d'équivalences pourra ultérieurement être étendu aux autres unités lexicales.

1.2.1. Les types de sources

Les sources d'acquisition de connaissances sont :

- terminologiques, comme les dictionnaires et les glossaires. Elles informent principalement sur le sens des termes, leur utilisation pour dénommer des concepts et des objets et donnent des équivalences translinguistiques des termes.
- non terminologiques. Ce sont les textes sans visée terminologique comme les articles thématiques, les livres ou encore les catalogues. Elles délivrent néanmoins des informations terminologiques (les termes apparaissent en contexte).

1.2.2. Les sources terminologiques

Nous allons orienter notre travail sur les forces et les faiblesses des sources dictionnaires qui sont les sources terminologiques les plus connues et les plus usitées.

1.2.2.1. Les dictionnaires bilingues et multilingues

Leur accès tant externe qu'interne¹⁴⁹ présentant des avantages non négligeables pour le traducteur (coût en temps et en argent réduits et organisation alphabétique), les dictionnaires bilingues et plurilingues peuvent servir de point de départ.¹⁵⁰ Cependant, si le dictionnaire a un but commercial, le critère de fiabilité pâtit car, pour gagner du temps et de l'argent, le dictionnaire est conçu à partir de dictionnaires existants et peut reprendre leurs erreurs et, si le dictionnaire est le fruit du travail du comité terminologique d'une organisation internationale, c'est la rapidité qui en pâtit (Gile, 1995 : 140).

D'autre part, le nombre de termes d'Internet répertorié dans les dictionnaires bilingues est faible étant donné que a) le domaine d'Internet est récent, b) le temps nécessaire à la compilation d'un dictionnaire est considérable et c) la fréquence des rééditions est peu élevée. De plus, il n'y a pas pléthore de dictionnaires bilingues spécialisés. En janvier 2005, le site Amazon.fr recense cinq dictionnaires anglais-français parus entre 1996 et 2004 (recherche avec « dictionnaire Internet »). Le site Casadelibro.com répertorie deux dictionnaires anglais-espagnol.¹⁵¹

1.2.2.2. Les dictionnaires monolingues

¹⁴⁹ Cinq variables permettent d'estimer l'utilité des sources d'acquisition de connaissances : l'existence, l'accès externe, l'accès interne, le taux de couverture et la fiabilité (Gile, 1995 : 134-135). Par *accès externe*, Gile désigne le coût d'accès à la source en temps, argent et effort. *Accès interne* signifie temps et effort requis pour extraire l'information recherchée dans la source.

¹⁵⁰ Pour garantir la fiabilité des informations, l'acquisition de connaissance peut se faire en trois étapes. Certaines sources servent de point de départ, d'autres d'intermédiaire et d'autres encore de point final.

¹⁵¹ Aucun dictionnaire anglais-français n'est exclusivement consacré à Internet (les domaines relevés sont : l'informatique, le multimédia, les télécommunications).

Un double problème se pose au traducteur spécialisé qui cherche des informations sur des concepts du domaine d'Internet désignés par des noms composés puisque ces unités lexicales sont à la fois longues et spécialisées. Or, la longueur et le degré de spécialisation des termes sont primordiaux pour leur inclusion dans un dictionnaire de langue générale (LG) ou de langue spécialisée (LS). Les mots simples sont plus fréquemment répertoriés que les mots composés et surcomposés. Si la présence de termes dans les grands dictionnaires de LG est une tradition – ils représentent une proportion notable de la nomenclature¹⁵² –, constituent-ils une source de départ fiable pour le traducteur spécialisé dans le domaine d'Internet ? Les définitions et les informations obtenues sont-elles appropriées aux besoins du traducteur ? Pour répondre à ces questions, nous nous intéresserons à l'identification des termes par les lexicographes (1.2.1.2.1.), à leur traitement (1.2.1.2.2.) et à l'élaboration des définitions (1.2.1.2.3.).

1.2.2.2.1. Identification des termes par les lexicographes

Les informations recueillies pour mener cette analyse proviennent essentiellement de *Scientific and Technical Words in General Dictionaries* (Béjoint, 1988). Les lexicographes utilisent plusieurs critères définitoires pour déterminer, parmi toutes les lexies de leur corpus, celles qui sont des lexies scientifiques et techniques (Béjoint, 1988 : 359) (nous verrons ensuite si ces critères s'appliquent au domaine d'Internet) :

- les lexies scientifiques et techniques occupent une place prépondérante dans une taxinomie spécialisée,
- leur fréquence d'utilisation est significative uniquement dans les discours très spécialisés,
- elles sont monosémiques,
- leur sens a peu de chances de changer brusquement,
- leur origine peut être retracée,
- elles n'ont pas d'extension métaphorique.

Les lexies réunissant le plus de critères sont les lexies très spécialisées, comme EN *key idea pointer* ou *kilo-logical inferences per second*. Elles sont répertoriées dans les grands dictionnaires de LG, voire uniquement dans les dictionnaires de LS. Ce sont les lexies scientifiques et techniques prototypiques. Ensuite, viennent les lexies occupant une place dans la taxinomie du domaine, utilisées dans les discours très spécialisés dont le sens est stable (*kurchatovium*), puis celles dont le concept n'est pas très spécialisé (*kleptomania*). Enfin, les lexies les plus éloignées du prototype occupent une place dans la taxinomie du domaine, sont plus fréquentes dans les discours spécialisés mais ne répondent à aucun autre critère. Ces lexies sont communes ou générales.

Dans quelle mesure les critères définitoires des lexies spécialisées s'appliquent-ils au

¹⁵² Landau (1984), cité par Béjoint (1988 : 354), estime que les termes scientifiques et techniques représentent plus de 40% des entrées du *Webster's Third New International Dictionary*.

domaine d'Internet ?

Tableau 68 : Application des critères définitoires des lexies scientifiques et techniques au domaine d'Internet

Critères	Application réussie ?	Exemples anglais
Place importante dans la taxinomie du domaine	Oui	<i>World Wide Web, email, FTP</i>
Fréquence d'utilisation uniquement significative dans les discours très spécialisés	Non (nombreuses lexies issues de la langue générale ou d'autres domaines)	<i>mail, server, cookie, chat</i>
Lexies monosémiques	Non (lexies empruntées à d'autres domaines)	<i>cookie, proxy,</i>
Sens peu enclin à changer soudainement	Pas toujours (évolution rapide des concepts et objets du domaine)	<i>forum, worm</i>
Origine retracée des lexies	Non (nombreux emprunts sémantiques)	<i>backdoor</i>
Pas d'extension métaphorique	Non (nombreuses métaphores)	<i>Trojan horse, script kiddy</i>

L'abondance des emprunts internes et la proximité du lexique d'Internet avec la langue populaire rendent l'application des critères difficile. Ces particularités éloignent les lexies d'Internet du prototype de la lexie spécialisée. Seules les lexies de plus de trois éléments touchant les aspects les plus techniques d'Internet s'en rapprochent comme EN *Asymmetric Digital Subscriber Line, always-on broadband connection* ou *high-speed Internet access* qui sont relatives aux modes de connexion. La plupart des critères définitoires n'étant pas réunis (tableau 68), il est peu probable de voir des lexies d'Internet répertoriées dans les dictionnaires de LG.

1.2.2.2. Traitement des lexies scientifiques et techniques dans les dictionnaires

Trois aspects du traitement des lexies spécialisées dans les dictionnaires monolingues seront étudiés : l'attribution des étiquettes, le nombre de lexies répertoriées et la nomenclature.

- L'attribution des étiquettes

Dans les dictionnaires, les lexies sont souvent repérables grâce à des étiquettes. Ces étiquettes sont utilisées par les traducteurs lorsqu'ils extraient des lexies de la nomenclature des dictionnaires de LG et LS pour concevoir et/ou compléter leur source terminologique (travail que les traducteurs effectuent avant d'entreprendre toute traduction). Mais selon Béjoint (1988 : 360) :

- pour un même terme, l'étiquette peut varier d'un dictionnaire à un autre,

- les étiquettes sont parfois manquantes,
- des lexies qui ne sont pas vraiment scientifiques ou techniques peuvent avoir une étiquette.

Comme le montre le tableau 69, ces irrégularités se retrouvent dans les dictionnaires de LS et les glossaires d'Internet. Les étiquettes de Netgloss sont plus générales que celles du GDT : « technical term » vs. « Internet » pour *anonymous surfing* et « technical term » vs. « computer programming language » pour *alias*.

Tableau 69 : Etiquettes de domaine attribuées par le GDT et Netgloss

Composés anglais	GDT	Netgloss
<i>ad banner</i>	electronic commerce	online marketing
<i>alias</i>	computer programming language	technical term
<i>anonymous surfing</i>	Internet	technical term
<i>freeware</i>	Internet	online jargon
<i>hacker</i>	computer crimes / computer security /Internet	hacker jargon

- Le nombre de lexies scientifiques et techniques répertoriées

Il est évident que le traducteur doit consulter de grands dictionnaires de LG et LS pour bénéficier d'un maximum d'informations sur les lexies d'Internet car plus un dictionnaire contient de lexies d'Internet, plus il permet au traducteur en phase de recherche documentaire de réaliser des économies de temps (nombre limité d'ouvrages à consulter) et d'améliorer sa productivité (non dispersion). Plus la nomenclature d'un dictionnaire de langue générale est riche, plus elle contient de termes scientifiques et techniques (Béjoint, 1988 : 361).

Gile (1995 : 137) évoque plusieurs contreparties. Une seule source couvre rarement tous les besoins terminologiques. Les grands dictionnaires sont coûteux, difficilement maniables, normatifs et surtout ils ne peuvent être exhaustifs en raison de la prolifération naturelle des termes. Les dictionnaires sont dépassés avant même leur publication. Cela est d'autant plus vrai pour le domaine d'Internet qui est en constante évolution.

Béjoint (1988 : 362) aborde un autre problème. Les dictionnaires de LG incluent rarement les termes scientifiques et techniques composés dont le sens semble évident. Ces dictionnaires sont plus créés pour le décodage que pour l'encodage. En revanche, ces composés sont répertoriés dans les glossaires et dictionnaires de LS car ils sont indispensables au discours scientifique et technique et à la taxinomie du domaine.

- Place des noms composés dans la nomenclature

Les dictionnaires anglais de LG consultés (MWOD, CALDO et OED 1992) accordent une faible place au lexique d'Internet. À titre d'exemple, sur les 201 composés N + N recherchés dans le MWOD,¹⁵³ 19 seulement sont répertoriés en tant qu'entrées. Le sens

de 70 composés peut être reconstitué puisque les éléments constitutifs sont des entrées et l'acception « Internet » est précisée. Cependant, du fait de l'assemblage des définitions de lexies simples, cette reconstitution du sens n'est que partielle, certains traits conceptuels sont absents. Par exemple, le MWOD définit EN *multipartite* et *virus*. La reconstitution du sens du terme composé *multipartite virus* à partir des deux définitions présente des lacunes. *Multipartite* a deux acceptations : « divided into several or many parts » et « having numerous members or signatories » et *virus* est défini de la manière suivante : « a computer program usually hidden within another seemingly innocuous program that produces copies of itself and inserts them into other programs and that usually performs a malicious action (as destroying data) ».¹⁵⁴

1.2.2.2.3. Définition des lexies spécialisées dans les dictionnaires

Les dictionnaires de LG ne reflètent pas toujours la précision des lexies spécialisées qui idéalement doivent être monosémiques pour que la communication entre spécialistes fonctionne. Ces définitions sont imprécises parce que (Béjoint, 1988 : 363) :

- le lexicographe ne peut délimiter le sens de manière plus précise.
- elles doivent être comprises du public auquel le dictionnaire est destiné, autrement dit elles ne sont pas scientifiques. Le langage utilisé dans la définition est simple. Or, les concepts sont désignés par les lexies spécialisées et ne peuvent être définis que par des mots qui sont eux aussi spécialisés.¹⁵⁵

1.2.2.2.4. Choix des dictionnaires de langue de spécialité

Il faut que le traducteur soit vigilant lors du choix des dictionnaires de LS qu'il consultera car certains sont des traductions et d'autres sont mauvais car ils donnent des équivalents qui ne sont pas ou peu employés en langue cible. Par exemple, le *Norton nouveau dictionnaire des télécommunications, de l'informatique et de l'Internet* (NND) est une traduction de l'anglais. Il définit, explique les termes anglais et présente des propositions d'équivalents français. Ces dernières ne sont pas toujours utilisées par les spécialistes. C'est le cas pour EN *Asymmetric Digital Subscriber Line*. Les propositions de termes sont : FR *raccordement numérique asymétrique* (RNA), *liaison numérique à débit asymétrique* et *ligne numérique à paire asymétrique*. Le traducteur devra consulter une autre source pour obtenir le terme réellement employé.

1.3. Conclusion

¹⁵³ Les difficultés posées par la traduction des noms composés du domaine d'Internet et les Recherche effectuée en mai 2004.

¹⁵⁴ L'assemblage est très éloigné de la définition du fabricant d'antivirus McAfee « multipartite viruses are a combination of techniques including infecting documents, executables and boot sectors to infect computers. Most multipartite viruses first become resident in memory and then infect the boot sector of the hard drive. Once in memory, multipartite viruses may infect the entire system ». Source: http://fr.mcafee.com/VirusInfo/VIL/glossary_app.asp

¹⁵⁵ Comme EN *executable*, *boot sector* dans la définition de *multipartite virus* (note précédente).

faiblesses des sources d'acquisition de connaissances révélées dans les sections 1.1. et 1.2. soulignent l'importance de la mise en place d'outils d'aide à la traduction. Rappelons les faits les plus marquants :

- Les chances de trouver les composés du domaine d'Internet dans des dictionnaires de LG et LS sont limitées. Les définitions des dictionnaires de LG ne sont pas pertinentes. Il faut faire attention à l'aspect normatif des dictionnaires.
- Les dictionnaires de LS sont une source valable à condition d'être suffisamment récents et régulièrement mis à jour.
- Un tri doit être opéré dans les sources terminologiques (élimination des traductions et des mauvais ouvrages).

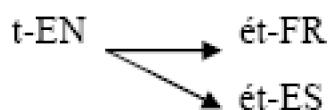
Pour créer un système d'aide à la traduction et rédaction technique, le sens de circulation des termes entre les trois langues étudiées doit être déterminé avec certitude. Dans le système multilingue, les termes sources seront-ils toujours les termes anglais ? Les couples de langues seront-ils anglais-français, anglais-espagnol ou français-espagnol, par exemple ? Autrement dit, les termes anglais passent-ils directement en français et en espagnol, sans langue intermédiaire ?

2. La circulation des termes dans le domaine d'Internet

S'il est indéniable que la langue source du vocabulaire d'Internet est l'anglais, le sens de circulation des termes entre l'anglais, le français et l'espagnol n'est, lui, pas avéré. Trois hypothèses seront examinées dans le but de le déterminer.

- Hypothèse 1 (H1)¹⁵⁶

Hypothèse 1 (H1)¹⁵⁵



Les termes anglais passent directement en français et en espagnol, c'est-à-dire sans langue intermédiaire. Les procédés de traduction sont le plus souvent le calque et l'emprunt. Les lexies françaises et espagnoles sont des équivalents de traduction des lexies anglaises.

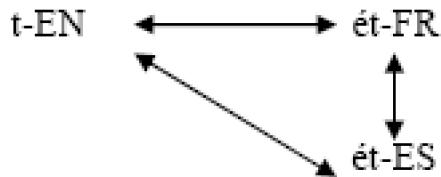
- Hypothèse 2 (H2)



¹⁵⁶ Légende : Les termes ét = équivalent de traduction, EN = anglais, FR = français et ES = espagnol H2 aussi, constituées

d'équivalents de traduction des lexies anglaises. Mais contrairement à H1, seuls les termes français sont créés directement à partir des termes anglais. Les termes espagnols sont des adaptations indirectes des termes anglais, le français est la langue intermédiaire.

- Hypothèse 3 (H3)



Les termes français et espagnols sont majoritairement des équivalents de traduction des termes anglais, mais cette fois il y a un échange entre les trois langues. Certaines lexies formées en français peuvent être adaptées en anglais et en espagnol et des lexies espagnoles peuvent entrer en français et en anglais. Le manque d'exemples rend cette hypothèse difficilement concevable. H3 pourrait expliquer la présence de synonymes et l'instabilité de certains termes dans les trois langues – mais un flou terminologique existe souvent dans les vocabulaires récents même sans échange entre langues.

Dans H1, H2 et H3, la langue source des lexies d'Internet est l'anglais. H1 semble l'hypothèse la plus probable dans la mesure où l'anglais est la langue véhiculaire du domaine, le vecteur de transmission des termes pour des raisons historiques et de diffusion¹⁵⁷ mais aussi la langue de fabrication. Un exemple parmi tant d'autres, Lavasoft, une entreprise suédoise, a lancé un produit avec un nom anglais : *AD-aware personal*.¹⁵⁸ Ces observations desservent H3 qui ne peut être retenue. Il est très difficile de trouver des termes créés en français ou en espagnol et ensuite « transposés » en anglais puisque les produits français ou espagnols sont issus de technologies déjà existantes développées aux États-Unis.¹⁵⁹ H2 est peu envisageable, l'Espagne d'aujourd'hui ne vit plus tournée vers la France. Selon Pratt (1980), les néologismes entrés en espagnol entre 1800 et 1950 sont des adaptations de modèles français et ceux entrés depuis 1950 sont des adaptations de modèles anglais. Il affirme que, depuis deux cents ans, la langue espagnole ne crée pas de néologismes. Elle se limite à copier et à adapter les lexies étrangères :

Hay muchas pruebas (aunque lingüísticamente no son incontrovertibles) que apoyan la opinión de que en los doscientos últimos años el español apenas ha acuñado una voz nueva, ni ex nihilo (es decir, de raíces latinas, griegas, etc.), ni de palabras existentes, por muy polémica que resulte esta afirmación. Lo que sí

¹⁵⁷ Les agences de presse sont en grande majorité anglaises (Humbley 1988).

¹⁵⁸ Ad-aware est un utilitaire qui scanne la mémoire, la base de registre et les disques durs pour détecter les spyware connus et qui permet de les enlever (www.telecharger.01net.com).

¹⁵⁹ C'est le vocabulaire relatif aux FAI qui se développe le plus en France et en Espagne à cause des enjeux commerciaux. Il ne semble pas y avoir de retour vers la langue source.

ha hecho es copiar y adaptar neologismos extranjeros, anglicismos en la época de la posguerra, y galicismos en los dos siglos anteriores. (1980: 45).

3. Des équivalences types anglais-français et anglais-espagnol sont-elles possibles ?

Pour répondre à cette question, quarante noms composés anglais et leur(s) équivalent(s) français et espagnol(s) ont été sélectionnés dans plusieurs revues présentant les mêmes caractéristiques – revues d'Internet généralistes s'adressant à un large public de personnes expérimentées ou débutantes. Certains des composés ne figurent pas dans les listes définitives – Annexe 7 – car ils sont tirés de Ahronian (2001) et de revues ne faisant pas partie de l'ensemble final.¹⁶⁰

3.1. Présentation de l'échantillon

Les composés appartiennent à des thèmes variés (courrier électronique, forum de discussion, navigation, moteur de recherche, piratage, téléchargement, mode de connexion) et font partie du lexique réel du domaine puisqu'ils ont été relevés dans des magazines. Ce sont des termes attestés, répertoriés ou non. En cas de concurrence, si plusieurs composés désignent un même concept, les hapax et les lexies peu utilisés sont exclus de l'échantillon.

Tableau 70 : Échantillon trilingue de termes

¹⁶⁰ Test effectué avant la constitution du corpus présenté dans la deuxième partie (Ch. 1).

Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol

Patrons syntaxiques anglais	Composés	Équivalents français	Patrons syntaxiques français	Équivalents espagnols	Patrons syntaxiques espagnols
N + N	backbone	backbone	emprunt	eje principal	N + A
				backbone	emprunt
	bandwidth	bande passante	N + A	ancho de banda	N+de+N
				anchura de banda	N+de+N
	banner ad	bannière de publicité	N + de + N	banner publicitario	N + A
	bookmark	signet	N	bookmark	emprunt
		favori	N		
		bookmark	emprunt		
	cross-post	envoi multiple	N + A	envío cruzado	N + A
	domain name	nom de domaine	N+de+N	nombre de dominio	N+de+N
	firewall	pare-feu	V + N	cortafuegos	V + N
		firewall	emprunt	firewall	emprunt
	gateway	passerelle	N	pasarela	N
	home page	page d'accueil	N+de+N	página de inicio	N+de+N
	IP address	adresse IP	N + N	dirección IP	N + N
	mail-bombing	mitraillage de courrier	N+de+N	bombardeo postal	N + A
	mailbot	répondeur de courrier	N+de+N	administrador de correo	N+de+N
	mail-filter	filtre courrier	N + N	filtro de correo	N+de+N
	mailing list	liste de diffusion	N+de+N	lista de correo	N+de+N
	Net access	accès Internet	N + N	acceso a Internet	N+a+N
	newsgroup	forum de discussion	N+de+N	grupo de noticias	N+de+N
		groupe de dicussion	N+de+N	grupo de discusión	N+de+N
		forum	N		
	search engine	moteur de recherche	N+de+N	motor de búsqueda	N+de+N
	snail mail	snail mail	emprunt	snail mail	emprunt
		courrier postal	N + A	correo tradicional	N + A
A + N					

Les noms composés anglais français et espagnols du domaine d'Internet :

	anonymous FTP	FTP anonyme	N + A	FTP anónimo	N + A
	clever agent	agent intelligent	N + A	agente inteligente	N + A
	dedicated line	ligne dédiée	N + A	línea dedicada	N + A
	electronic mail	courrier électronique	N + A	correo electrónico	N + A
	hotlist	liste de signets	N+de+N	lista de enlaces	N+de+N
	Internet	Internet	emprunt	Internet	emprunt
	trojan horse	cheval de Troie	N+de+N	caballo de Troya	N+de+N
V + N	hacktivist	hacktiviste	N + N	hacktivista	N + N
	password	mot de passe	N+de+N	contraseña	N
Adv + N	download	téléchar-gement		descarga	N
	upload	téléchar-gement		descarga	N
N + Adv	login	identifiant	N	nombre de usuario	N+de+N
		login	emprunt		
	startup	startup	emprunt		
N + P + N	B to B	BtoB	emprunt	BtoB	emprunt
	Point of Presence	Point d'accès	N+de+N	punto de presencia	N+de+N
N + N + N	Domain name system	Système de Nom de Domaine	N+de+N+de+N	Sistema de nombre de dominio	N+de+N+de+N
Adv + A+ N	Frequently Asked Questions	Foire aux Questions	N+à+N	Frequently Asked Questions	emprunt
		Frequently Asked Questions	emprunt		
A + N+ N	high-speed access	accès haut-débit	N + A+ N	acceso a alta velocidad	N+a+A+N
N + A + N	World Wide Web	World Wide Web	emprunt	World Wide Web	emprunt
PS + N + N	email address	adresse e-mail	N + N	dirección de e-mail	N+de+N
PS + N	webmaster	webmaster	emprunt	webmaster	emprunt
		webmestre	PS + N		
	Web page	page Web	N + PS	página Web	N + PS

3.2. Observations et hypothèses

protégé en vertu de la loi du droit d'auteur.

-
- Les types compositionnels sont très variés. Douze structures ont été utilisées dans la création des quarante composés anglais. Onze types sont à l'origine des quarante-neuf équivalents français et des quarante-quatre équivalents espagnols. Plusieurs patrons français et espagnols correspondent parfois à un seul patron anglais. Par exemple, la structure anglaise N + N est rendue en français par un emprunt ou par les structures N + A, N + de + N, et V + N.
 - Le français et l'espagnol utilisent généralement des patrons compositionnels pré-existants. Un patron que l'on trouve quelquefois dans le lexique du commerce est fréquemment utilisé ici : la formation à partir de pseudo-confixes (dans FR *téléthon*, *Sofitel*, *Novotel* –*thon* et –*tel* sont les troncations de *marathon* et *hotel*). Les composés N + de + N, N + A existent dans la langue générale et dans de nombreuses langues de spécialité françaises et espagnoles (FR *couteau de cuisine*, *corps gras*, ES *parada de autobús*, *ataque cardíaco*).
 - Les lexies composées sont très majoritairement rendues en français et en espagnol par des calques ou des emprunts. Exemples : EN *Web page* > FR *page Web*, ES *página web* ; EN *webmaster* > FR *webmaster*, *webmestre*, ES *webmaster*.
 - Les termes anglais et français ont souvent une similitude morphologique. Les composés créés par les structures anglaises N + N et A + N sont généralement traduits par des composés français N + de + N et N + A. Cependant, certains termes français et espagnols sont assez éloignés des termes anglais aussi bien sur le plan morphologique que sémantique. Exemples : EN *home page* > ES *página de inicio*, EN *newsgroup* > FR *forum de discussion*, ES *grupo de discusión*, EN *mailbot* > FR *répondeur de courrier* et EN *login* > ES *nombre de usuario*.
 - Un terme anglais peut avoir plusieurs traductions dans les deux langues cibles. Exemples : *newsgroup*, *login*.
 - Certains termes anglais ont un signifiant nouveau (*search engine*, *bug hunter*) et d'autres un signifiant qui existait avant la création d'Internet (*thumbnail*, *address book*). Ces termes ne sont pas nécessairement adaptés de la même manière. Par exemple, le nom *composégateway*, antérieur à Internet, est rendu en français par *passerelle*, une lexie simple. Le français a eu recours à l'emprunt sémantique, il n'a pas créé une nouvelle forme. Le composé *newsgroup*, créé dans le cadre d'Internet, a pour équivalent français *forum de discussion*, un nom composé. Les néologismes sémantiques seront-ils problématiques pour les équivalences types ?
 - Si le terme anglais est un amalgame ou est issu d'une construction inhabituelle, il semble ne pas y avoir d'équivalence type anglais-français et anglais-espagnol. Par exemples, *Internet* reste *Internet* dans les deux langues cibles mais *mailbot* devient *répondeur de courrier* en français et *administrador de correo* en espagnol – les deux équivalents ne partagent ni la forme, ni les éléments de nomination de la lexie d'origine.
 - Le français et l'espagnol semblent privilégier certains patrons syntaxiques pour traduire un patron anglais précis (N + N et A + N sont majoritairement traduits par N + de + N et N + A). Il semble possible de dresser un parallèle entre la structure

morphosyntaxique des composés sources et cibles et de se baser sur la fréquence comme nous l'avions envisagé (Part. II, Ch. 1, Sec. 3.1.). Par exemple, au patron anglais A + N correspondent x composés français de structure N + A et y de structure N + de + N, etc.

Pour résumer cette pré-enquête, il est possible d'aider le traducteur en traitant les difficultés de traduction liées aux composés et en effectuant les recherches documentaires en amont. Cette aide peut prendre la forme d'un système d'équivalences types multilingue. Il est possible de créer des équivalences compositionnelles types.

Chapitre 2. Les noms composés anglais et leurs équivalents français et espagnols

Ce chapitre, consacré à la conception d'un système d'équivalences types destiné en particulier au traducteur et au rédacteur spécialisés, commencera par la mise en place d'équivalences de traduction anglais-français et anglais-espagnol. Nous testerons ensuite un échantillon de composés anglais pour identifier les exceptions au système d'équivalences. Enfin, nous utiliserons des traducteurs automatiques pour traduire quelques composés anglais dans le but de connaître les performances de ces outils dans la traduction des composés du lexique d'Internet.

1. Les équivalences types anglais-français et anglais-espagnol

Une manière, pour le terminologue/terminographe d'intervenir dans l'activité de traduction est d'établir dans quelle mesure les caractéristiques morphosyntaxiques et sémantiques des composés sources et cibles peuvent être exploitées pour créer des modèles d'équivalences. Le système d'équivalences reposera sur les patrons syntaxiques des composés anglais ainsi que sur ceux des composés français et espagnols. L'aspect sémantique interviendra ultérieurement (Sec. 1.4.) car il aura une importance dans le traitement des exceptions. À partir de la structure compositionnelle d'un nouveau terme anglais, le système d'équivalences permettra de créer un équivalent de traduction dans les langues cibles – le français et l'espagnol. Pour avoir les meilleures chances d'implantation, les équivalents de traduction devront être basés sur les modèles des termes implantés dans le domaine et, bien sûr, respecter les contraintes du système morphosyntaxique.

1.1. La méthode de création

Cinq étapes président à la mise en place des modèles d'équivalences morphosyntaxiques anglais-français et anglais-espagnol :

- Regrouper les composés anglais selon leur patron compositionnel. Pour cela, nous

reprendrons les typologies construites (Part. III, Ch. 1, Sec. 3, liste Annexe 8).

- Rechercher les termes français et espagnols qui sont les équivalents de traduction de chaque composé anglais en explorant les corpus français et espagnol. Si certains équivalents ne figurent pas dans nos corpus, nous consulterons, entre autres, des glossaires et des dictionnaires sérieux pour nous procurer les termes manquants.¹⁶¹
- Créer une base de données mettant en parallèle les termes anglais – classés selon leur morphologie –, et leurs équivalents français et espagnols et faisant ressortir la structure de ces derniers.
- Calculer la fréquence « type » des modèles morphosyntaxiques français et espagnols. Autrement dit, calculer le nombre de fois qu'une structure morphosyntaxique française ou espagnole est utilisée pour traduire les termes d'une structure anglaise particulière, toutes occurrences confondues.
- Classer les structures françaises et espagnoles selon la fréquence « type » calculée précédemment (toujours par rapport au modèle compositionnel anglais).

Ces étapes permettent de découvrir si un patron compositionnel anglais est traduit en langue d'arrivée par un ou plusieurs patrons compositionnels ou par d'autres types de formation néologique (dérivation, lexie simple, etc.) et de connaître la fréquence de ces structures.

Un exemple bilingue anglais-français permettra de mieux saisir la démarche. À chaque composé anglais N + N (*mail server*, *newsletter*, *address book*) correspondent un ou plusieurs équivalents français. Regrouper dans un tableau les termes anglais, les traductions françaises et leurs structures morphosyntaxiques donne les moyens de calculer la fréquence des modèles français utilisés.

Tableau 71 : Structures françaises équivalentes au patron compositionnel anglais N + N

	Composés anglais	Équivalents français	Structure des équivalents français
1	<i>address book</i>	<i>carnet d'adresses</i>	N + de + N
2	<i>bandwidth</i>	<i>bande passante</i>	N + A
		<i>bande large</i>	N + A
3	<i>log file</i>	<i>fichier journal</i>	N + N
		<i>fichier log</i>	N + N
		<i>log de connexion</i>	N + de + N
		<i>journal de connexion</i>	N + de + N
4	<i>domain name</i>	<i>nom de domaine</i>	N + de + N

Pour savoir comment traduire un futur composé anglais N + N, nous calculons la fréquence « type » des structures françaises utilisées pour rendre ces composés anglais (N + N, N + A et N + de + N). Dans notre exemple, la structure la plus fréquente est N + de + N (4 équivalents français), suivie à égalité de N + N et N + A (2 équivalents chacune).

¹⁶¹ Les ressources françaises consultées sont : www.granddictionnaire.org , www.telecharger.com , Les ressources espagnoles se situent aux adresses suivantes : www.rae.es , www.rae.es/avlex.htm.

serviront de base à la création néologique dans les langues cibles lorsque de nouveaux composés anglais feront leur apparition. Lorsqu'un néologisme composé anglais NC_a sera créé dans le domaine d'Internet, le traducteur saura qu'il serait préférable de créer un équivalent français E_f de structure S_{f X} (ou S_{f Y} si E_f est irréalisable ou n'a pas de bonnes capacités d'implantation).¹⁶² Le traducteur pourra ainsi construire un équivalent en langue cible conforme au vocabulaire existant.

1.2. Les résultats

Les résultats obtenus grâce à la démarche décrite (Sec. 1.1.) seront retranscrits sous forme de tableaux.

1.2.1. Méthode d'exploitation des équivalences types

Pour construire un équivalent en langue cible, le traducteur utilisera la première équivalence de chaque tableau (équivalence la plus fréquente). Si elle ne permet pas la création d'un terme satisfaisant (Part. III, Ch.1, Sec.1.1.2.), alors il pourra se reporter à la deuxième équivalence proposée. Lorsqu'un patron compositionnel anglais est rendu en langue cible par plusieurs structures, seules les quatre ou cinq premières structures seront utilisées pour créer des termes.

La méthode d'exploitation des équivalences est expliquée à l'aide d'exemples. Le patron du néologisme EN *bookmark converter* est N + N + N. L'équivalent français sera de préférence un N + de + N + de + N (21,28%). On peut construire le composé FR *convertisseur de marque de page* qui n'est pas très satisfaisant dans la mesure où *bookmark* – dont le signifiant existait avant Internet – a déjà plusieurs équivalents français (*favori*, *signet* ou *bookmark*) dont *marque de page* ne fait pas partie. FR *Convertisseur de signets* est un bon candidat. Il répond à N + de + N, le troisième modèle le plus fréquent (12,77%) en français pour traduire les composés anglais N + N + N. *Convertisseur de bookmark* – N + de + N + N est la deuxième structure la plus fréquente avec 14,89% – serait un bon équivalent.

Le nom propre anglais *Crack Killer*¹⁶³ (structure N + N) pourrait devenir un terme. Il devrait alors de préférence être rendu par un composé N + de + N en français (fréquence : 39,27%) comme *exterminateur de cracks* ou *tueur de cracks*, qui contiennent l'emprunt *crack* puisque le terme est implanté en français et dans lesquels la métaphore véhiculée par *killer* a été conservée. De plus, ces termes s'apparentent à FR *chasseur de bugs*.

1.2.2. Les équivalences types anglais-français

¹⁶² Voir Part. III, Ch. 1, Sec. 1.1.2. pour les détails les qualités d'un bon équivalent terminologique : l'équivalent E correspond à la notion *n* dans l'esprit des spécialistes ; l'équivalent est conforme aux règles de la langue ; il est bref, transparent, productif ; ce peut être un emprunt interne ou un terme hybride.

¹⁶³ <http://www.sofotex.com> : "Crack Killer is a database application that enables software authors to track and report web sites hosting cracks, serials and pirated versions of their software [...]".

1.2.2.1. Les équivalences des composés anglais binaires

Tableau 72 : Équivalences du patron N + N

Rang	Structures françaises	Fréquence « type »	%
1	N + de + N	86	39,27%
2	N + N	73	33,33%
3	Emp2	25	11,42%
4	N	11	4,57%
5	N + A	10	5,02%
6	N + PS + N	3	1,37%
	PS + N	3	1,37%
7	N + de + PS	2	0,91%
8	N + de + N + de + N	1	0,46%
	N + sans + N	1	0,46%
	N + sur + N	1	0,46%
	V + N	1	0,46%
	N + à + N + A	1	0,46%
	N + à + N	1	0,46%
	Total	219	100,00%

Tableau 73 : Équivalences du patron A + N

Rang	Structures françaises	Fréquence « type »	%
1	N + A	67	79,76%
2	Emp	6	7,14%
3	N + de + N	4	4,76%
4	N + N	2	2,38%
	N	2	2,38%
	Fdiv	2	2,38%
5	N + à + N	1	1,19%
	Total	84	100,00%

Tableau 74 : Équivalences du patron PS + N

Rang	Structures françaises	Fréquence « type »	%
1	PS + N	61	64,21%
2	Emp	18	18,95%
3	N + PS	9	9,47%
4	N + de + N	2	2,11%
	N + N	2	2,11%
5	PS + A	1	1,05%
	N + en + N	1	1,05%
	N + de + PS	1	1,05%
	Total	95	100,00%

Tableau 75 : Équivalences du patron N+ PS

Rang	Structures françaises	Fréquence « type »	%
1	N + N	3	37,50%
2	Emp	3	37,50%
3	N + PS	1	12,50%
	PS + sans + N	1	12,50%
	Total	8	100,00%

Tableau 76 : Équivalences du patron A + PS

Rang	Structures françaises	Fréquence « type »	%
1	Emp	2	40,00%
	PS + A	2	40,00%
2	N + A	1	20,00%
	Total	5	100,00%

Tableau 77 : Équivalences des patrons anglais mineurs

Rang	Structures françaises	Fréquence « type »	%
in+N			
1	Emp	1	50,00%
	N + de + N	1	50,00%
up+N			
1	N + A	1	100,00%
N+up			
1	Emp	2	40,00%
2	A + N	1	20,00%
	N + A	1	20,00%
	N + N	1	20,00%
N+to			
1	Emp	1	100,00%
N+over			
1	emp	1	100,00%
down + V			
1	PS + N	3	60,00%
2	Emp	2	40,00%
up+V			
1	Emp	1	50,00%
	PS + N	1	50,00%
N + to + N			
1	Emp	4	100,00%
N + of + N			
1	N + de + N	3	75,00%
2	emp	1	25,00%
N + and + N			
1	emp	1	100,00%
Fdiv			
1	Fdiv	2	50,00%
2	N + N	1	25,00%
	Emp	1	25,00%

1.2.2.2. Les équivalences des surcomposés anglais à trois éléments

Tableau 78 : Équivalences du patron N + N + N

Les noms composés anglais français et espagnols du domaine d'Internet :

Rang	Structures françaises	Fréquence « type »	%
1	N+de+N+de+N	10	21,28%
2	N+de+N+N	7	14,89%
3	Emp	6	12,77%
	N+de+N	6	12,77%
4	N+N	5	10,64%
5	N+N+N	2	4,26%
	V+N+N	2	4,26%
	N	2	4,26%
6	N+A	1	2,13%
	N+de+N+A	1	2,13%
	N+de+N+en+N	1	2,13%
	N+de+N+sans+N	1	2,13%
	N+N+par+N	1	2,13%
	N+N+sans+N	1	2,13%
	N+PS	1	2,13%
	Total	47	100,00%

Tableau 79 : Équivalences du patron A + N + N

Rang	Structures françaises	Fréquence « type »	%
1	N+de+N+A	6	23,08%
	N+A+N	6	23,08%
2	N+N+A	5	19,23%
3	N+A	3	11,54%
4	N+N	2	7,69%
	Emp	2	7,69%
5	N+A+de+N	1	3,85%
	N+de+N+de+N	1	3,85%
	Total	26	100,00%

Tableau 80 : Équivalences du patron off + N + N

Rang	Structures françaises	Fréquence « type »	%
1	N+hors+N	1	25,00%
	N+de+N	1	25,00%
	N+PS	1	25,00%
	N+N	1	25,00%
	Total	4	100,00%

Tableau 81 : Équivalences du patron N + A + N

Rang	Structures françaises	Fréquence « type »	%
1	Emp	2	50,00%
2	N	1	25,00%
	N+de+N+A	1	25,00%
	Total	4	100,00%

Tableau 82 : Équivalences du patron N + up + N

Rang	Structures françaises	Fréquence « type »	%
1	N+N	3	30,00%
	N+A	3	30,00%
2	N+par+N+A	1	10,00%
	N+en+N	1	10,00%
	N+a+N+A	1	10,00%
	N+a+N+N	1	10,00%
	Total	10	100,00%

Tableau 83 : Équivalences du patron PS + N + N

Rang	Structures françaises	Fréquence « type »	%
1	N+de+N	13	35,14%
2	N+N	7	18,92%
3	N+de+N+A	4	10,81%
4	N+de+N+PS	2	5,41%
	N+de+PS+N	2	5,41%
5	N+de+N+de+N	1	2,70%
	PS+N+de+N	1	2,70%
	N+de+N+N	1	2,70%
	en+N+de+N	1	2,70%
	Emp	1	2,70%
	N+A	1	2,70%
	N+A+PS	1	2,70%
	N+PS	1	2,70%
	PS+N	1	2,70%
	Total	37	100,00%

Tableau 84 : Équivalences du patron A + PS + N

Les noms composés anglais français et espagnols du domaine d'Internet :

Rang	Structures françaises	Fréquence « type »	%
1	N+non+A	8	61,54%
2	N+PS+A	2	15,38%
	N+A	2	15,38%
3	N+A+A	1	7,69%
	Total	13	100,00%

Tableau 85 : Équivalences des patrons mineurs (3 éléments)

Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol

Rang	Structures françaises	Fréquence « type »	%
on + N + N			
1	N+en+N	1	100,00%
under +N+ N			
1	N +A	1	100,00%
N + under + N			
1	Emp	1	100,00%
N + down + N			
N + A + PS			
1	N + de + N	1	100,00%
down + V + N			
1	N+de+PS+N	5	100,00%
adv + A + N			
1	Emp	1	50,00%
	N+à+N	1	50,00%
PS + A + N			
1	N+PS+A	1	50,00%
	N+PS	1	50,00%
N + PS + N			
1	N+A	1	100,00%
PS + N + PS			
1	N+PS	1	100,00%
PS + PS + N			
1	N+N	1	100,00%
adv + on + N			
1	N+A	2	100,00%
N + of + N			
1	N+N	2	100,00%
N + to + N + N			
1	N+N	2	50,00%
2	N+N+N	1	25,00%
	N +N+par+N	1	25,00%
N + and + N + N			
1	N+A	1	100,00%

1.2.3. Les équivalences types anglais-espagnol

1.2.3.1. Les équivalences des composés espagnols binaires

Tableau 86 : Équivalences du patron N + N

Rang	Structures	Fréquence	%
------	------------	-----------	---

Les noms composés anglais français et espagnols du domaine d'Internet :

	espagnoles	« type »	
1	N de N	104	53,33%
2	N N	40	20,51%
3	Emp	26	13,33%
4	N A	9	4,62%
5	N a N	3	1,54%
	N	3	1,54%
	N PS	3	1,54%
6	N PS N	2	1,03%
	PS N	2	1,03%
7	Fdiv	1	0,51%
	N sobre N	1	0,51%
	PS PS	1	0,51%
	Total	195	100,00%

Tableau 87 : Équivalences du patron A + N

Rang	Structures espagnoles	Fréquence « type »	%
1	N + A	58	80,56%
2	N + de + N	4	5,56%
3	Emp	3	4,17%
4	Fdiv	2	2,78%
	N	2	2,78%
5	N + N	1	1,39%
	N + de + N + A	1	1,39%
	N + de + N + PS	1	1,39%
	Total	72	100,00%

Tableau 88 : Équivalences du patron PS + N

Rang	Structures espagnoles	Fréquence « type »	%
1	PS + N	52	57,78%
2	N + PS	16	17,78%
3	Emp	13	14,44%
4	N + a + N	3	3,33%
5	N + de + N	1	1,11%
	N + de + PS + N	1	1,11%
	N + de + N + PS	1	1,11%
	N + N	1	1,11%
	PS + PS	1	1,11%
	N + A	1	1,11%
	Total	90	100,00%

Tableau 89 : Équivalences du patron A + PS

Rang	Structure	Fréquence « type »	%
1	N + N	2	28,57%
	PS + A	2	28,57%
	Emp	2	28,57%
2	N+A+de+todo+N	1	14,29%
	Total	7	100,00%

Tableau 90 : Équivalences du patron N + PS

Rang	Structures espagnoles	Fréquence « type »	%
1	Emp	4	57,14%
2	N + A	2	28,57%
3	N + N	1	14,29%
	Total	7	100,00%

Tableau 91 : Équivalences des patrons mineurs

Rang	Structures espagnoles	Fréquence « type »	%
in + N			
1	Emp	1	100,00%
up + N			
1	PS + N	1	50,00%
	N+de+PS+N	1	50,00%
N + up			
1	N + A	2	50,00%
	Emp	2	50,00%
N + over			
1	Emp	1	100,00%
down + V			
1	N	2	100,00%
N + to + N			
1	Emp	4	80,00%
2	N + a + N	2	20,00%
N + and + N			
1	Emp	1	100,00%
N + of+ N			
1	N + de + N	3	75,00%
2	Emp	2	25,00%
Fdiv			
1	N + N	1	50,00%
2	Emp	1	50,00%

1.2.3.2. Les équivalences des surcomposés espagnols à trois éléments

Tableau 92 : Équivalences du patron N + N + N

Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol

Rang	Structures espagnoles	Fréquence « type »	%
1	N + de + N	9	22,50%
	N + de + N + de +N	9	22,50%
2	Emp	5	12,50%
3	N + N	4	10,00%
	N + de + N + a + N	4	10,00%
4	N+A+a+N	1	2,50%
	N+a+N+por+N	1	2,50%
	N+de+N+de+N+A	1	2,50%
	N+de+N+en+N	1	2,50%
	N+de+N+N	1	2,50%
	N+en+N	1	2,50%
	N+N+N	1	2,50%
	N+por+N+a+N	1	2,50%
	V+N+N	1	2,50%
	Total	40	100,00%

Tableau 93 : Équivalences du patron A + N + N

Rang	Structures espagnoles	Fréquence « type »	%
1	N+de+N+A	6	30,00%
2	N+de+A+N	4	20,00%
3	N+A	3	15,00%
	N+N+A	3	15,00%
4	N+de+N	2	10,00%
5	N+a+A+N	1	5,00%
	N+A+de+N	1	5,00%
	Total	20	100,00%

Tableau 94 : Équivalences du patron PS + N + N

Rang	Structures espagnoles	Fréquence « type »	%
1	N + de + N	20	62,5%
2	N + A	3	9,38%
3	N+de+N+A	2	6,25%
	N+N	2	6,25%
	N+de+N+PS	2	3,13%
4	N+de+N+a+N	1	3,13%
	Emp	1	3,13%
	PS+N	1	3,13%
	Total	32	100,00%

Les noms composés anglais français et espagnols du domaine d'Internet :

Tableau 95 : Équivalences du patron A+ PS + N

Rang	Structures espagnoles	Fréquence « type »	%
1	N+de+N	6	27,27%
	N+A	6	27,27%
2	N+A+A	3	13,64%
3	N+PS+A	1	4,55%
	PS+A	1	4,55%
	N+A+no+A	1	4,55%
	A+N	1	4,55%
	Emp	1	4,55%
	N+más+N	1	4,55%
	N+PS	1	4,55%
	Total	22	100,00%

Tableau 96 : Équivalences des patrons mineurs

Rang	Structures espagnoles	Fréquence « type »	%
on + N+ N			
1	N+N	2	75,00%
2	N+en+N	1	25,00%
off + N + N			
1	N+N	2	75,00%
2	N + fuera + de +N	1	25,00%
under + N + N			
1	N + N	1	100,00%
N + A + N			
1	N + A	1	50,00%
	Emp	1	50,00%
N + up + N			
1	N + N	2	28,57%
2	N + A	1	14,29%
	N + de + N	1	14,29%
	Emp	1	14,29%
	N + por + N	1	14,29%
	N + por + N + A	1	14,29%
	Total	7	100,00%
N + down + N			
1	Emp	1	100,00%
N + under + N			
1	Emp	1	100,00%
PS + A + N			
1	PS + N	1	50,00%
	N + PS + A	1	50,00%
N + of + N + N			
1	N+de+N+de+N	1	50,00%
	N+ N	1	50,00%
N + to + N + N			
1	N + N	2	50,00%
2	N + N + N	1	25,00%
	N + N + a + N	1	25,00%
	Total	4	100,00%

Pour ne pas fausser les statistiques des équivalences types, les trente-deux composés anglais exocentriques ne sont pas inclus dans la liste.¹⁶⁴

Les équivalences types des surcomposés de plus de trois éléments ne seront pas présentées car les données dont nous disposons ne sont pas suffisantes : les vingt et un surcomposés de quatre éléments sont issus de seize patrons syntaxiques, les cinq

¹⁶⁴ Ils seront étudiés séparément (Part. III, Ch. 2, Sec. 2.2).

surcomposés de cinq éléments ont été créés d'après cinq patrons syntaxiques et le corpus contient un seul surcomposé de six éléments.

2. Identification des exceptions

Pour évaluer le degré de pertinence de notre système d'équivalences types, nous allons appliquer les statistiques aux composés de notre corpus anglais formés par le type le plus fréquent : N + N.

2.1. Test du système d'équivalences types sur le patron compositionnel anglais N + N

À titre d'exemple, appliquons les données établies en 1.3.2.1. pour créer des équivalents français et espagnols au composé EN *address book*. Les cinq structures les plus fréquentes permettent de produire les équivalents suivants :

- Français
 - 1) *carnet d'adresses* (N + de + N),
 - 2) *carnet-adresses* (N + N),
 - 3) *address book* (Emp),
 - 4) *carnet* (N),
 - 5) *carnet relationnel* (N + A).
- Espagnol
 - 1) *libreta de direcciones* (N+de+N),
 - 2) *libreta direcciones* (N + N),
 - 3) *address book* (Emp),
 - 4) *libreta relacional* (N + A),
 - 5) *libreta* (N).

Les équivalences 1 correspondent aux termes français et espagnol implantés. Des occurrences de l'équivalent 2 français ont été relevées sur Google.fr. Les lexies simples (N) sont trop imprécises. L'équivalent 5 français et l'équivalent 4 espagnol auraient des chances se s'implanter si EN *address book* n'avait encore aucun équivalent français et espagnol.

La traduction de EN *book* par FR *carnet* et ES *libreta* suscite une question : comment un système d'équivalences types informatisé pourra-t-il traiter les éléments polysémiques et les métaphores ? Nous allons observer trois termes anglais (*bookmark*, *address book* et *guest book*) et leurs équivalents français (*marque-page*, *carnet d'adresses* et *livre d'or*). *Book* est l'élément déterminant (*bookmark*) ou l'élément déterminé (*address book*, *guest*

book). Il est traduit en français par trois équivalents (*page*, *carnet* et *livre*). La traduction ne dépend pas de la position de *book* puisque dans *address book* et *guest book*, il est le déterminé et est traduit de deux manières. Nous ne voyons pas comment paramétrier une machine pour qu'elle comprenne qu'un élément occupant la même position dans plusieurs composés n'a pas le même équivalent en langue cible. La meilleure solution est, selon nous, d'effectuer un emprunt interne en langue cible puisqu'il y en a un en langue source, et d'appliquer une des cinq équivalences types. Dans EN *navigation bar*, *bar* est un emprunt interne. En créant l'équivalent FR *barre de navigation*, on recourt au même emprunt interne qu'en anglais (idem dans EN *discussion board* et FR *tableau de discussion*).

Tableau 97 : Test du système d'équivalences sur les composés anglais N + N du corpus

¹⁶⁵ Les composés en caractères gras sont exocentriques.

¹⁶⁶ Les termes précédés du symbole # sont attestés (dans notre corpus ou sur Google.fr) et correspondent à une équivalence type. Les termes précédés du symbole * sont proposés mais non attestés (les termes utilisés par les spécialistes sont alors indiqués entre parenthèses).

Les noms composés anglais français et espagnols du domaine d'Internet :

Composés anglais ¹⁶⁵	Équivalents français ¹⁶⁶	Rang de la structure française
access control	#contrôle d'accès	1
ad blocker	#bloqueur de pub	1
address bar	#barre d'adresses	1
address book	#carnet d'adresses	1
ADSL modem	#modem ADSL	2
ADSL provider	#fournisseur d'ADSL	1
alias page	#page alias	2
antivirus provider	#fournisseur d'antivirus	1
archive manager	#gestionnaire d'archives	1
archive moderator	modérateur d'archives	1
attack program	#programme d'attaque	1
audience measurement	#mesure d'audience	1
bandwidth	*largeur de bande (#bande passante)	1
banner ad	#bannière de pub	1
banner advertising	#publicité de bannière	1
blocking access	#accès bloquant	5
bookmark	#signet	4
boot virus	#virus de boot	1
bottleneck	*goulot de bouteille (#goulot d'étranglement)	1
browser cache	#cache de/du navigateur	1
browser sniffer	#renifleur de navigateur	1
browsing habit	#habitude de navigation	1
bug hunter	#chasseur de bugs	1
cable modem	#modem câble	2
chat group	#groupe de chat	1
chat host	#hébergeur de chat	1
chat server	#serveur de chat	1
chat session	#session de chat	1
chat window	#fenêtre de chat	1
chatbox	#boîte de chat	1
chat-mate	*compagnon de chat	1
chatroom	#salon de chat	1
client interface	#interface client	2
comparison site	#site de comparaison	1

¹⁶⁵ Les composés en caractères gras sont exocentriques.

¹⁶⁶ Les termes précédés du symbole # sont attestés (dans notre corpus ou sur Google.fr) et correspondent à une équivalence type. Les termes précédés du symbole * sont proposés mais non attestés (les termes utilisés par les spécialistes sont alors indiqués entre parenthèses).

connection kit	#kit de connexion	1
connection speed	#vitesse de connexion	1
connection time	#temps de connexion	1
contact group	#groupe de contact	1
contact list	#liste de contact	1
cookie builder	#générateur de cookies	1
cookie file	#fichier cookie	1
cookie ID	#cookie d'ID	1
cookie management	#gestion de cookies	1
cookie manager	#gestionnaire de cookies	1
data packet	#paquet de données	1
discussion area	#espace de discussion	1
discussion board	#tableau de discussion	1
discussion forum	#forum de discussion	1
discussion group	#groupe de discussion	1
distribution list	#liste de distribution	1
DNS server	#serveur DNS	2
domain name	#nom de domaine	1
DoS attack	#attaque DoS	2
dot com	#point com	2
dot org	#point org	2
emoticon	#émoticône	2
encryption key	#clé de chiffrement	1
encryption program	#programme de chiffrement	1
encryption system	#système de chiffrement	1
file sharing	#partage de fichiers	1
file swapping	#échange de fichiers	1
filtering package	#logiciel de filtrage	1
filtering rule	#règle de filtrage	1
firewall	*mur de feu (#firewall, coupe-feu)	1
flamewar	#guerre d'insultes	1
forwarding address	#adresse de suivi	1
fraggle attack	#attaque fraggle	2
frame relay	#relais de trame	1
FTP access	#accès FTP	2
FTP client	#client FTP	2
FTP file	#fichier FTP	2
FTP server	#serveur FTP	2
FTP space	#espace FTP	2
guestbook	#livre des visiteurs	1
hacker attack	#attaque de hacker	1
hacker community	#communauté de pirates	1
hacking attack	#attaque de piratage	1
hacktivism	#hacktivisme	3
hacktivist	#hacktiviste	3

Les noms composés anglais français et espagnols du domaine d'Internet :

history folder	#fichier d'historique	1
hit counter	#compteur de visites	1
home page	*page maison (#page d'accueil)	2
hosting company	#entreprise d'hébergement	1
hosting package	#logiciel d'hébergement	1
hosting URL	#adresse (URL) d'hébergement	1
HTML editor	#éditeur HTML	2
HTML integrator	#intégrateur HTML	2
HTML page	#page HTML	2
HTML site	#site HTML	2
HTML server	#serveur HTML	2
indexing engine	#moteur d'indexation	1
Internet acces	#accès Internet	2
Internet cache	#cache Internet	2
Internet café	#café Internet	2
Internet community	#communauté Internet	2
Internet connection	#connexion Internet	2
Internet cookie	# cookie Internet	2
Internet name	*nom Internet (#adresse Internet)	2
Internet portal	#portail Internet	2
Internet protocol	#protocole Internet	2
Internet server	#serveur Internet	2
Internet tool	#outil Internet	2
Internet user	*utilisateur d'Internet (#internaute)	1
IP address	#adresse IP	2
IP network	#réseau IP	2
IP router	#routeur IP	2
IRC channel	#canal IRC	2
IRC client	#client IRC	2
IRC server	#serveur IRC	2
Java applet	#applet Java	2
keylogger	#enregistreur de touche (frappe)	1
log file	#fichier journal	2
log sniffer	*renifleur de log (#analyseur de log)	1
mail bombing	#bombardement de courrier	1
mail checker	#vérificateur de courrier	1
mail forwarding	#suivi de courrier	1
mail server	#serveur de courrier	1
mailbox	#boîte de courrier	1
mailing list	#liste d'envoi	1
message board	#panneau/tableau de messages	1
message filter	#filtre de message	1
mirror site	#site miroir	2
modem	#modem	3

MP3 file	#fichier MP3	2
MP3 site	#site MP3	2
navigation bar	#barre de navigation	2
netizen	*netizen (#citoyen du Net)	3
network administrator	#administrateur de réseau	1
network connection	#connexion réseau	2
newsgroup	#groupe de <i>nouvelles</i>	1
newsgroupie	*groupie de nouvelles	1
newsletter	#lettre d' <i>informations</i>	1
P2P network	#réseau P2P	2
page view	#page vue	5
parody site	#site de parodie	1
password	#mot de passe	1
ping sweep	#balayage de ping	1
POP account	#compte POP	2
POP server	#serveur POP	2
port blocking	#blocage de port	1
port scanning	#scan (balayage) de port	1
portal site	#site portail	2
program virus	#virus programme	2
proxy server	#serveur mandataire	5
PSTN connection	#connexion RTC	2
read notification	#accusé de lecture	1
receipt notification	#accusé de réception	1
rotation banner	#bannière rotative	5
satellite modem	#modem satellite	2
script blocking	#blocage de script	1
script file	#fichier script	1
script kiddy	*enfant du script (#script kiddy, apprenti pirate)	1
script virus	#virus de script	1
scrolling banner	#bannière défilante	5
search engine	#moteur de recherche	1
security hole	#trou de sécurité	1
server attack	#attaque de serveur	1
server connection	#connexion serveur	2
server package	#logiciel de serveur	1
shopbot	#shopbot	3
showroom	#salle de démonstration	1
site directory	#annuaire de sites	1
site owner	#propriétaire de site	1
site structure	#ossature de site	1
SMS messaging	#messagerie SMS	2
smurf attack	#attaque smurf	2
snail mail	#courrier escargot	2

Les noms composés anglais français et espagnols du domaine d'Internet :

source code	#code source	2
space provider	#fournisseur d'espace	1
spam protection	#protection spam	2
storage site	#site de stockage	1
streaming video	#contenu vidéo	2
style sheet	#feuille de style	1
surfing habit	#habitude de surf	1
techno-nerd	*crétin de technologie (#accro de technologie)	1
thumbnail	*ongle du pouce (#vignette)	1
transfer speed	#vitesse de transfert	1
URL history	#historique URL	2
Usenet group	#groupe Usenet	2
Usenet posting	#article Usenet	2
user account	#compte d'utilisateur	1
user authentication	#authentification d'utilisateur	1
user interface	#interface d'utilisateur	1
username	#nom d'utilisateur	1
video chat	#chat vidéo	2
video chatting	#chat vidéo	2
videostream	#flot vidéo	2
virus alert	#alerte de virus	1
virus attack	#attaque de virus	1
virus definition	#définition de virus	1
virus writer	#auteur de virus	1
Vmail	#courrier vocal	5
voice chat	#chat vocal	5
voice IP	#voix IP	2
WAP locking	#verrouillage WAP	2
WAP page	#page WAP	2
WAP portal	#portail WAP	2
wireless connection	#connexion mobile	5
worm alert	#alerte de ver	1
worm attack	#attaque de ver	1

Pour créer les équivalents français figurant dans le tableau 97, nous avons appliqué l'équivalence la plus fréquente en français, en calquant si possible. Si le terme obtenu est satisfaisant (Part. III, Ch. 1., Sec. 1.1.2.), la procédure s'arrête et s'il ne l'est pas, nous avons appliqué la deuxième équivalence la plus fréquente (*page WAP*), ou la troisième (*hacktiviste*), la quatrième ou la cinquième (*chat vocal*).

Le fait que quelques équivalents français soient des emprunts (*modem*, *émoticône*) et le nombre assez élevé de termes hybrides (*fenêtre de chat*, *attaque fraggle*) peuvent surprendre. Lorsque l'un des éléments d'un composé anglais est implanté en français (*chat*, *cookie*, *Internet*, *modem*, *hacker*, *boot*, *bug*, *log*, *ping*, *scan*, *script*, *spam*), nous le

réutilisons pour former ainsi un équivalent français conforme au lexique d'Internet existant. Dans deux cas, FR *attaque fraggle*, *attaque smurf*, le déterminant anglais a été emprunté puisqu'il y a une référence culturelle inconnue de l'internaute français.¹⁶⁷ Nous constatons que les éléments empruntés sont courts et connus des internautes français (sauf EN *smurf* et EN *fraggle*). Rares sont les cas où nous avons appliqué l'équivalence 3 (Emp) et ils concernent exclusivement des amalgames anglais : *Internet*, *netizen*, *modem*, *shopbot*, *émoticône*, *hacktivisme*, et *hacktiviste* (les trois derniers sont francisés).

Plusieurs questions naissent à la lecture du tableau 97 : les termes proposés grâce au système d'équivalences types correspondent-ils aux termes français attestés ?¹⁶⁸ Si oui, quelle proportion de termes proposés correspond à des termes attestés ? Les cinq équivalences types (N + de + N, N + N, Emp2, N + A et N) sont-elles également utilisées ?

Près de 93,50% (188 sur 201) des équivalents français proposés par le système d'équivalences types sont attestés. Ce chiffre montre que ce système est réellement exploitable et efficace et que nous sommes souvent fidèle au mode de formation des termes du domaine en ayant recours au calque. Les quelques équivalents français qui ne sont pas attestés correspondent aux termes anglais *bandwidth*, *bottleneck*, *thumbnail*, *firewall*, *home page*, *chate-mate*, *log sniffer*, *netizen*, *newsgroupie*, *script kiddy*, *Internet name*, *Internet user*, *techno nerd*. Les cinq premiers composés anglais sont exocentriques et les huit derniers endocentriques. Le faible taux de noms composés endocentriques peut s'expliquer par la concurrence terminologique. Ils ont généralement des concurrents plus fréquents en anglais (EN *Internet address*, *internaut*). Les termes attestés en français sont les calques de ces concurrents anglais (*adresse Internet*, *internaute* et *accro de la technologie*).

Quelques composés exocentriques créés par le système d'équivalences sont attestés. Outre leur exocentricité, ils présentent d'autres particularités. Ils sont des amalgames (*shopbot*), contiennent un emprunt implanté en français (*cookie*), un nom d'extension (*com* et *org*) ou un sigle (*IP*). Étant donné que seuls cinq équivalents de termes anglais exocentriques sur onze sont implantés, nous en déduisons qu'il était bénéfique de les exclure des statistiques lors de la création des équivalences types anglais-français et anglais-espagnol. Il est certainement moins chronophage de créer un équivalent français au cas par cas que de s'évertuer à appliquer un 'système d'équivalences'. En effet, ce système appliqué aux six autres composés exocentriques non attestés donne des équivalents ayant un potentiel d'implantation nul puisqu'ils sont opaques (**goulot de bouteille*, **biscuit d'identifiant*, **mur-feu*, **biscuit Internet*, **enfant script*, **ongle du pouce*). En revanche, il est toujours possible d'emprunter le terme anglais (troisième équivalence la plus fréquente de notre système). D'ailleurs, *script kiddy* et *shopbot* sont attestés et fréquents en français.

¹⁶⁷ EN *smurf* est le nom anglais équivalent de FR *schtroumpf* et FR *fraggle* est une allusion à aux marionnettes de Fraggle Rock.

¹⁶⁸ Cette question est inéluctable dans la mesure où le système d'équivalences types, mis en place pour la traduction des néologismes composés anglais ne possédant pas encore d'équivalents français et espagnols, est conçu d'après le vocabulaire existant.

Synthétiser la fréquence d'emploi des cinq équivalences types sous la forme d'un graphique permettra de répondre à notre troisième question : les équivalences types sont-elles également utilisées ?

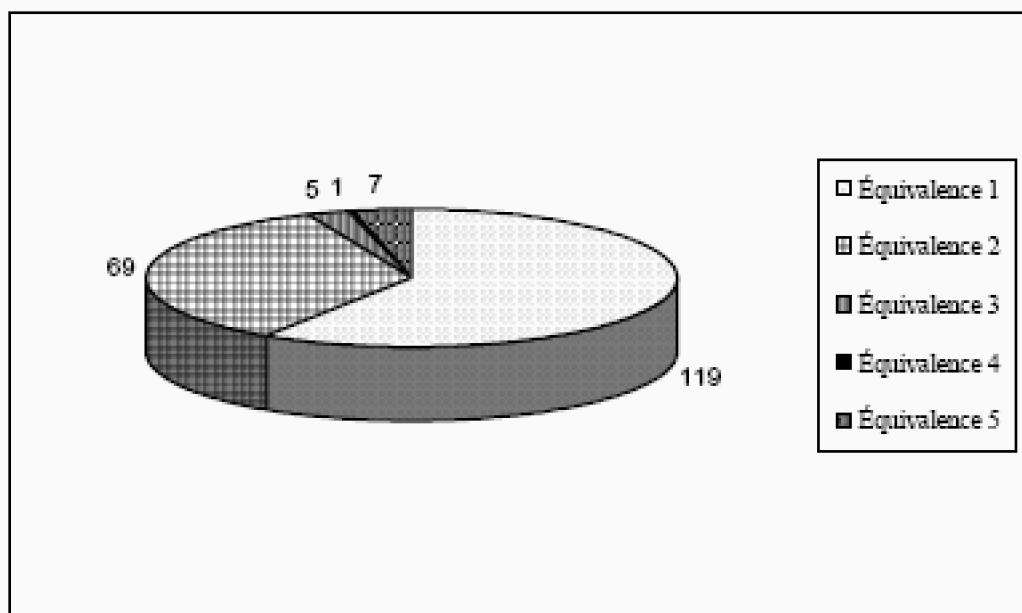


Figure 7 : Fréquence des équivalences types dans la traduction des N + N anglais

La figure 7 illustre les écarts d'utilisation des équivalences types (de 1 à 119). L'équivalence 1 (N + de + N) entrant dans la création de 60% des équivalents français se trouve en tête. Avec 34%, l'équivalence 2 (N + N) arrive en deuxième position. Cette équivalence s'applique généralement dans des cas précis. Un élément du composé anglais est :

- un sigle (*attaque DoS, portail WAP*),
- un amalgame (*Internet, modem, ping, applet* dans *café Internet, modem satellite, balayage ping, Java applet*),
- un nom propre (*article Usenet, applet Java*).

Pour les autres N + N français, l'existence d'équivalents déjà implantés n'est sûrement pas étrangère à l'utilisation du modèle N + N plutôt que N + de + N. Au troisième rangs se situe l'emprunt, utilisé avec parcimonie (termes anglais implantés en français, amalgames). Avec sept termes, l'équivalence 5 (N + A) est très peu représentée. L'équivalence 4 (lexie simple), utilisée une seule fois (*signet*), est en dernière position. Nous aurions sûrement proposé le calque FR *marque (de) page* si aucun terme français n'avait existé. Son potentiel d'implantation aurait été élevé. En effet, les sites Web sont constitués de pages, l'internaute aurait donc pu, pour mémoriser ses sites préférés, se servir d'un *marque-page* au lieu d'un *signet, bookmark* ou *favori*.

En somme, hormis les composés exocentriques, aucune autre propriété ne semble entraver la création de termes à l'aide du système d'équivalences types.

2.2. Étude des équivalences des composés exocentriques

Tableau 98 : Équivalents français des composés anglais exocentriques

¹⁶⁹ Les composés en caractères gras sont totalement métaphoriques. L'élément métaphorique des composés partiellement métaphoriques est souligné. Les termes suivis du symbole @ sont des métonymies.

Les noms composés anglais français et espagnols du domaine d'Internet :

Composés anglais exocentriques ¹⁶⁹	Équivalents français attestés	Méthodes de traduction employées	Équivalents français proposés	Équivalences types appliquées
backdoor	porte de derrière	Calque	porte de derrière	N+de+N
<u>bandwidth</u>	bande passante	Transposition	largeur de bande	N+de+N
black hat	pirate informatique	Transposition	chapeau noir	N+A
bottleneck	goulot d'étranglement	Transposition	cou de bouteille	N+de+N
click and mortar@	clic et brique/mortier	calque	click et mortar	Emp2
<u>cookie ID</u>	cookie d'identifiant	Calque + Emprunt	cookie d'ID	N+de+N
<u>cyberpunk</u>	cyberpunk	Emprunt	cyberpunk	PS+N
<u>cybersquatter</u>	cybersquatter	Emprunt	cybersquatteur	PS+N
<u>cybersquatting</u>	cybersquatting	Emprunt	cybersquat	PS+N
<u>dotcom@</u>	point com	Calque	point com	N+N
<u>dot org@</u>	point org	Calque	point org	N+N
firewall	firewall	Emprunt	mur de feu	N+de+N
	coupe-feu	Transposition		
<u>frame relay</u>	relais de trames	Calque	relais de trame	N+deN
<u>hyperlink</u>	hyperlien	Calque	hyperlien	PS+N
<u>intelligent agent</u>	agent intelligent	Calque	agent intelligent	N+A
<u>Internet cookie</u>	cookie Internet	Transposition	cookie Internet	N+N
mailto	mailto	Emprunt	mailto	Emp2
<u>permanent cookie</u>	cookie permanent	Emprunt + calque	cookie permanent	N+A
popup	fenêtre publicitaire	Transposition	popup	Emp2
rollover	effet de survol	Transposition	rollover	Emp2
<u>script kiddy</u>	Script kiddy	Emprunt	enfant de script	N+de+N
	pirate amateur	Transposition		
<u>search engine spider</u>	araignée de moteur de recherche	Calque	araignée de moteur de recherche	N+de+N+de+N
Secure Socket Layer	protocole SSL / sécurisé	Transposition	couche sécurisée	N+A
<u>shopbot</u>	robot logiciel /	Transposition	robot de	N+de+N

¹⁶⁹

Les composés en caractères gras sont totalement métaphoriques. L'élément métaphorique des composés partiellement métaphoriques est souligné. Les termes suivis du symbole @ sont des métonymies.

	assistant d'achat		magasin	
startup	startup jeune entreprise /jeune poussée	Emprunt Transposition	startup	Emp2
<u>temporary cookie</u>	cookie temporaire	Emprunt + calque	cookie temporaire	N+A
thumbnail	vignette	Transposition	ongle de pouce	N+de+N
Trojan horse	Cheval de Troie	Calque	Cheval de Troie	N+de+N
<u>uplink</u>	liaison montante	Transposition	lien montant	N+A
white hat	corsaire / mordu de l'info / fouineur	Transposition	chapeau blanc	N+A

Avant de commenter les termes réalisables (colonne ‘Équivalents français proposés’), nous analyserons les termes réalisés (colonne ‘Équivalents français attestés’). Les équivalents français des exocentriques anglais qui sont attestés auraient-ils pu être proposés par le système d’équivalences types ?

- Neuf équivalents attestés sur trente correspondent à une équivalence type. Ils auraient pu être formés en utilisant notre système. Il s’agit de *porte de derrière*, *clic et mortier*, *point com*, *point org*, *relais de trames*, *hyperlien*, *agent intelligent*, *araignée de moteur de recherche* et *cheval de Troie*. Ces équivalents représentent un peu plus de 28% des équivalents attestés. À cette liste, on pourrait ajouter les quatre équivalents créés au moyen du calque et de l’emprunt (*cookie d’ID*, *cookie Internet*, *cookie permanent*, *cookie temporaire*). En effet, *cookie* est implanté en français et nous l’avons inséré dans les équivalents du tableau 98.
- Les vingt-et-un équivalents restants n’auraient pas pu être proposés par une équivalence type puisqu’ils ont été créés par transposition¹⁷⁰ (*pirate informatique*, *jeune entreprise*) ou empruntés (*firewall*, *cyberpunk*). Ces cas d’« échec » représentent un peu moins de 70% des termes exocentriques.

En comparant avec les résultats du tableau 98, le système d’équivalences types est bien moins performant dans le cas des composés anglais exocentriques (28%) que dans celui des composés endocentriques (93%).

Autre question importante, combien de termes français proposés par une équivalence type ont un bon potentiel d’implantation ? Le système d’équivalences et le calque sont inapplicables à seize composés exocentriques : *black hat*, *bottleneck*, *cookie ID*, *firewall*, *Internet cookie*, *mailto*, *permanent cookie*, *popup*, *rollover*, *script kiddie*, *Secure Socket Layer*, *shopbot*, *startup*, *temporary cookie*, *thumbnail* et *white hat*. Les équivalents obtenus semblent avoir un très faible potentiel d’implantation car ils sont opaques, comme **chapeau noir*, **goulot de bouteille*, **mur (de) feu*, ou imprécis, comme **enfant du script*, **robot de magasin*. Ceci est certainement lié au fait que certains de ces termes ont été empruntés à d’autres domaines et que leurs équivalents français existaient avant Internet et ont été repris (*coupe-feu*, *goulot d’étranglement*).

¹⁷⁰ Par *transposition* nous entendons terme dont le signifiant a été traduit de façon non littérale.

En conclusion, lorsque le traducteur aura besoin de traduire un néologisme anglais qui est un composé exocentrique totalement ou partiellement métaphorique et qu'il est impossible de créer un terme satisfaisant en utilisant le système d'équivalences et en calquant, il devra utiliser la transposition,¹⁷¹ tout en essayant de respecter les structures préconisées par le système d'équivalences.

Nous allons maintenant opérer le croisement des typologies morphosémantiques pour savoir si la relation sémantique interne N1-N2 et la morphologie des composés exocentriques peuvent expliquer que le système d'équivalences ne soit pas performant dans la traduction de ces derniers.

Tableau 99 : Croisement des typologies morphosémantiques

¹⁷¹ Comme le montre le tableau 98, la plupart de ces termes sont rendus en français par transposition.

¹⁷² Les équivalents sont présentés dans la colonne 'Équivalents proposés' du tableau 97.

Patrons syntaxiques	Composés anglais exocentriques	Relations sémantiques	Équivalents satisfaisants ? ¹⁷²
A+N	backdoor	N1N2 est un accès secret	Oui
A+N	black hat	N1N2 est un pirate	Non
A+N	intelligent agent	N1N2 est un programme automatique	Oui
A+N	permanent cookie	N1 se distingue des autres N1 car il est N2 (nature)	Oui
A+N	temporary cookie	N1 se distingue des autres N1 car il est N2 (nature)	Oui
A+N	Trojan horse	N1N2 est un programme malveillant	Oui
A+N	white hat	N1N2 est un pirate	Non
N+N	bandwidth	N1N2 est une gamme de fréquence	Oui
N+N	bottleneck	N1N2 est un bouchon virtuel	Non
N+N	cookie ID	N1 a pour activité N2	Oui
N+N	dotcom	N1N2 est un acteur de la nouvelle économie	Oui
N+N	dot org	N1N2 est un acteur de la nouvelle économie	Oui
N+N	firewall	N1N2 est un dispositif informatique de sécurité	Non
N+N	frame relay	N1N2 est un mode de transmission de données	Oui
N+N	Internet cookie	N1 est un instrument, mode, moyen utilisé par N2	Oui
N+N	script kiddy	N1N2 est un acteur du Net	Non
N+N	shopbot	N1 (logiciel, technique) a pour activité, fonction N2	Robot de mag
N+N	thumbnail	N1N2 est une composante de page	Non
PS+N	cyberpunk	N1N2 est un fou des nouvelles technologies	Oui (emp)
PS+N	cybersquatter	N1N2 est une personne qui achète des noms de domaine	Oui
PS+N	cybersquatting	N1N2 est l'activité qui consiste à acheter des noms de domaine	Oui
PS+N	hyperlink	N1N2 est une connexion activable	Oui
N+N+N	search engine spider	N1 a N2 comme localisation, environnement spatial	Oui
A+N+N	Secure Socket Layer	N1N2 est un protocole permettant une transmission sécurisée	Oui
N+and+N	click and mortar	N1N2 est un acteur de la nouvelle économie	Oui (emp)

¹⁷² Les équivalents sont présentés dans la colonne ‘Équivalents proposés’ du tableau 97.

N+over	rollover	N1N2 est une technique de traitement de l'image	Oui (emp)
N+to	mailto	N1N2 est un mot clé	Oui (emp)
N+up	popup	N1N2 est une fenêtre publicitaire	Oui (emp)
N+up	startup	N1N2 est un acteur de la nouvelle économie	Oui (emp)
up+N	uplink	N1N2 est une transmission depuis une station terrienne	Oui

Pour chaque composé anglais exocentrique, sont extraits le patron syntaxique et la relation N1-N2. Ensuite, nous observons les équivalents français proposés et tentons de déterminer si les cas d'échec sont liés à une structure compositionnelle et/ou une relation N1-N2 en particulier.

- Patron syntaxique A + N. Le système d'équivalences est appliqué avec succès à cinq composés sur sept. Il y a deux relations sémantiques différentes : ‘*N1N2 n'est pas un N1 et/ou un N2’ et ‘N1 se distingue des autres N1 car il est N2 (nature)’. Les éléments des deux cas d'échec sont unis par la première relation qui a fonctionné dans les trois autres cas.
- Patron syntaxique N + N. Onze composés sont formés sur ce modèle. Dans six cas, une équivalence type a pu être appliquée. Il y a donc cinq cas d'échec. En faisant le lien avec les quatre relations N1-N2 présentes, on s'aperçoit que la relation ‘*N1N2 n'est pas un N1 et/ou un N2’ n'est pas stable puisque quatre équivalents sur huit ne sont pas satisfaisants (ceux correspondant à *bottleneck*, *firewall*, *script kiddie* et *thumbnail*). L'autre équivalent dont nous pensons qu'il ne s'implantera pas est *robot magasin*. La relation N1-N2 du composé EN *shopbot* est ‘N1 (logiciel, technique) a pour activité fonction N2’.
- Patron syntaxique PS + N. Il entre dans la formation de quatre composés dont la relation N1-N2 est ‘*N1N2 n'est pas un N1 et/ou un N2’. Comme pour les endocentriques PS + N, le pseudo-confixe est emprunté et le déterminé est calqué ou emprunté, s'il est déjà implanté en français comme *squat*, *squatteur* et *punk*.
- Patrons syntaxiques mineurs. Ils sont au nombre de cinq : N + and + N, N + to, N + up, N + over et up + N. Une équivalence type est toujours applicable puisque, dans tous les cas, il s'agit de l'emprunt (sauf pour up + N). Une relation sémantique est présente : ‘*N1N2 est un N1 et/ou un N2’.
- Patrons syntaxiques N + N + N et A + N + N. Une équivalence a pu être appliquée pour traduire les deux composés anglais concernés.

En somme, quatre relations N1-N2 ont été relevées chez les exocentriques anglais. Avec 24 composés sur 30, soit 80%, la relation la plus fréquente est ‘*N1N2 est un N1 et/ou un N2’. Elle concerne les composés pour lesquels la relation déterminant-déterminé n'est pas identifiable (*black hat*, *dotcom*, etc.). On ne peut pas prédire qu'un composé anglais exocentrique pourra ou ne pourra pas être traduit avec le système d'équivalences types puisque le croisement de typologies ne permet pas de déterminer qu'une relation N1-N2

met en cause la traduction de tel ou tel patron compositionnel. En effet, même lorsque la relation N1-N2 n'est pas identifiable, le système d'équivalences fonctionne parfois.

3. Les traducteurs automatiques et la traduction des noms composés du domaine d'internet

Les bases de notre système d'équivalences types sont jetées. Nous allons maintenant étudier l'utilité que présenterait son informatisation pour le traducteur spécialisé. Pour cela, nous observerons brièvement l'environnement de travail des traducteurs (3.1.) puis nous testerons plus particulièrement l'efficacité de quelques logiciels de TA et TAO (3.2.).

3.1. Traduction automatique (TA) et Traduction assistée par ordinateur (TAO)

La création d'un outil informatisé de traduction des noms composés répond à un besoin de traduction et de rédaction. Un outil informatique exploitant les statistiques élaborées (Part. III, Ch. 2.) serait un gain de temps considérable pour les traducteurs – et pour toutes les personnes – qui produisent des textes en français ou en espagnol à partir de textes anglais contenant des noms composés qui n'ont pas encore d'équivalent. Disposer d'un programme qui propose des composés a) répondant aux règles morphosyntaxiques des langues cibles et b) similaires aux termes existant dans le domaine et c) ayant un fort potentiel d'implantation serait un précieux avantage.

3.1.1. L'ordinateur : un outil du traducteur

La prise de conscience de l'importance d'informatiser le système d'équivalences est venue de l'observation de l'environnement de travail des traducteurs. Ces vingt dernières années ont été marquées par le développement de la communication et le développement de l'informatique. Les ordinateurs évoluent à pas de géant : ils gagnent sans cesse en performance et en accessibilité. Le corollaire de cette évolution technologique est un changement dans la méthode et les outils de travail du traducteur.¹⁷³ Comme l'écrit Gouadec (2002), la « Traduction Assistée par Gomme et Crayon (TAGC) » n'a pratiquement plus cours.

Les traducteurs renouvellent leurs ordinateurs environ tous les deux ans pour suivre les progrès techniques et répondre aux exigences des logiciels (Gouadec 2002 : 119). Le poste du traducteur est composé de programmes de traitement de textes, d'outils de recherche, d'aide au traducteur ou à la traduction, de traitement terminologique, de bases de données terminologiques – il permet également de se connecter à Internet, en intranet et de travailler à distance. Pour que le système d'équivalences soit utile au traducteur, nous devons donc l'adapter à son mode de travail. Il doit être accessible rapidement et facilement, c'est-à-dire être informatisé. Dans un contexte où les délais d'exécution sont décisifs, il permettrait au traducteur d'être plus productif.

¹⁷³ Par le biais de l'informatique, il dispose d'un moyen rapide, pratique et précieux d'accès au savoir, à l'information terminologique et aux aides au traducteur. Les outils du traducteur, désormais indispensables et omniprésents, ont révolutionné son métier.

3.1.2. Critères de qualité des outils de TA et TAO

Pour satisfaire les exigences des traducteurs en matière de terminologie, un système d'aide à la traduction doit répondre à certains critères. L'informatisation de nos modèles d'équivalences doit :

- permettre au traducteur de produire des composés répondant aux règles de formation morphologique des langues cibles.
- être robuste : proposer au moins une solution et intégrer des solutions de « rattrapage » au cas où le premier composé proposé n'ait pas un potentiel d'implantation élevé (Part. III, Ch. 1, Sec. 1.1.2.).
- être efficace : contenir un nombre suffisant de composés à offrir aux traducteurs. Il doit pour cela être évolutif, c'est-à-dire être régulièrement mis à jour pour intégrer les équivalents correspondants aux derniers termes anglais créés.
- être facile d'emploi : avoir une interface agréable, ergonomique et s'intégrer à la production de traductions.

3.2. Logiciels de TA : traduisent-ils correctement les noms composés du lexique d'Internet ?

Pour répondre à cette question, nous testerons des dizaines de noms composés anglais du domaine d'Internet sélectionnés dans notre corpus et saisis dans différents programmes pour obtenir leur(s) traduction(s) française(s) et espagnole(s). Pour augmenter les chances d'obtenir une (ou des) traduction(s), ces termes appartiennent à des champs notionnels variés (le courrier électronique, la navigation, les modes de connexion, le piratage entre autres) et les logiciels testés auront des niveaux de spécialisation différents. Voici les principales caractéristiques des logiciels observés :

- SYSTRAN¹⁷⁴ est le traducteur le plus utilisé par les particuliers et les professionnels pour l'anglais et le français. C'est un traducteur instantané généraliste qui contient une forte proportion de termes appartenant à des domaines techniques et scientifiques. Il intègre des fonctionnalités multilingues dans le traitement et l'échange d'informations pour le commerce électronique, le CRM, la gestion de contenu, les bases de données, l'Intranet d'entreprise, l'e-mail, la messagerie instantanée, le SMS, le WAP, etc. Il facilite la communication dans 36 paires de langues et dans 20 domaines spécialisés – seules les combinaisons anglais-français et anglais-espagnol ont été testées. Le développement de SYSTRAN s'étend sur plus de 30 ans de mise au point de solutions de traduction personnalisées pour de grandes sociétés, portails, ISP, administrations publiques et gouvernementales.
- Tradinfo v 4.11¹⁷⁵ est un logiciel bilingue qui – avec ses 40 000 termes, acronymes, expressions, extensions de fichiers, logiciel, matériel, langages, réseaux, bases de données, intelligence artificielle, parallélisme, objet, etc. – facilite l'interprétation, la

¹⁷⁴ www.systran.fr

¹⁷⁵ Version shareware disponible sur www.telecharger.01net.com.

décomposition et l'identification d'expressions et de termes techniques relatifs à l'informatique. Il affiche les traductions, les synonymes et fournit des informations complémentaires (genre, commentaires).

- Word Magic IT Dictionary v. 1¹⁷⁶ rassemble environ 35 000 références relatives aux domaines du commerce, de l'informatique, d'Internet, des IT, de la télévision, de la radio, des télécommunications, du management, de l'électronique, de la physique, des mathématiques. Des dictionnaires et glossaires sont la base de ce logiciel qui traduit de l'anglais vers l'espagnol et de l'espagnol vers l'anglais.
- Babylon-Pro v 4.0.1.6¹⁷⁷ contient vingt-deux dictionnaires qui totalisent 120 000 entrées, couvrent cinquante-cinq langues et donnent instantanément la traduction. Les thèmes couverts sont : affaires et économie, divertissements, santé, informatique et Internet, etc. Les versions anglais-français et anglais-espagnol ont été examinées.

Les logiciels ont tous été testés en décembre 2002. Les résultats obtenus sont résumés dans les tableaux 100 à 104.

3.2.1. SYSTRAN

Tableau 100 : Test de SYSTRAN

¹⁷⁶ Version shareware disponible sur www.telecharger.01net.com ou sur le site www.wordmagicsoft.com.

¹⁷⁷ Version shareware disponible sur www.telecharger.01net.com.

Les noms composés anglais français et espagnols du domaine d'Internet :

Composés anglais	Traductions françaises	Traductions espagnoles
address book	carnet d'adresses	libro de direcccionamiento
ad-blocker	annonce-dresseur	anuncio-ad-blocker
attack program	programme d'attaque	programa del ataque
audience measurement	mesure d'assistances	medida de las audiencias
backdoor	secret	
bandwidth	largeur de bande	anchura de banda
banner ad	annonce de drapeau	anuncio de la bandera
bookmark	signet	
bottleneck	goulot d'étranglement	embottellamiento
browsing habit	habitude de parcourir	hábito el hojear
bug hunter	chasseur d'anomalie	cazador del fallo de funcionamiento
bulletin board	tableau d'affichage	tablón de anuncios
cable Internet access	accès d'Internet de câble	acceso del Internet del cable
cable modem	modem câblé	módem de cable
chat session	session de causerie	sesión de la charla
chatroom		
cookie builder	constructeur de biscuit	constructor de la galleta
cyber space	cyberespace	
cyber-café	cybercafé	
cybercriminal		
dial-up connection	connexion commutée	conexión de marcado manual
discussion board	panneau de discussion	tarjeta de la discusión
dot com	COM de point	
emoticon	émoticône	
file sharing	partage de fichier	el compartir del fichero
file-sharing program	fichier-partage du programme	fichero-compartir a compañía
file-swapping	fichier-permutation	fichero-intercambio
firewall	mur à l'épreuve du feu	cortafuego
hacker access	accès d'intrus	acceso del hacker
hacking attack	entailier l'attaque	cortar ataque
history folder	chemise d'histoire	carpeta de la historia
home page		
hotline	ligne directe	teléfono directo
HTML page	HTML PAGE	HTML PAGE
HTML site	site de HTML	sitio del HTML
Internet content provider	Fournisseur de contenu d'Internet	Abastecedor conteno del Internet
Internet protocol		
IP address	IP ADDRESS	IP ADDRESS
mail server		

mailing list	liste d'expédition	lista que envía
Net access	Accès net	Acceso neto
Net economy	Economie nette	Economía neta
network connection	connexion de réseau	conexión de red
offline browsing	en différé parcourant	fueras de línea hojeando
parody site	site de parodie	sitio de la parodiar
peer-to-peer network	réseau de pair-à-pair	red del par-a-par
pop up	sautez vers le haut	haga estallar para arriba
price comparison site	site de comparaison des prix	sitio de la comparación del precio
satellite modem	modem satellite	módem basado en los satélites
script kiddy	séquence type kiddy	escritura kiddy
scrolling banner	drapeau de défilement	bandera del movimiento en sentido vertical
search engine		
search engine spider	araignée de Search Engine	araña del Serach Engine
site creation tool	outil de création de site	herramienta de la creación del sitio
space provider	fournisseur de l'espace	abastecedor del espacio
startup	mise en route	lanzamiento
storage site	dépôt	sitio de almacenaje
streaming video	video coulant	video que fluye
surfing habit	habitude surfante	hábito que practica surf
techno-nerd	maniaque de la technologie	nerd del techno
thumbnail		
video stream	jet visuel	secuencia video
virus writer	auteur de virus	programa de escritura del virus
virus-signature file	fichier de virus-signature	fichero de la virus-firma
voice chat	causerie de voix	charla de la voz
web access	accès de web	acceso del web
Web hosting company	Web accueillant la compagnie	Web que recibe a compañía
web surfer	webnaute	persona que practica surf de Web
worm alert	alerte de ver	alarma de gusano

La performance du logiciel est visiblement très aléatoire. SYSTRAN est globalement plus compétitif pour les traductions vers le français que vers l'espagnol. Quelques commentaires sont nécessaires :

- Le décroisement des éléments pour respecter l'ordre déterminé/déterminant des

langues d'arrivée est quasiment systématique pour les composés binaires (surtout pour les N + N) mais aléatoire pour les surcomposés. Exemples : EN *address book* > FR *carnet d'adresses*, EN *cable modem* > ES *módem de cable*, EN *virus-signature file* > FR **fichier de virus signature*¹⁷⁸, ES **fichero de la virus-firma*, EN *offline browsing* > FR **en différé parcourant*, ES **fueras de la línea hojeando*.

- De nombreux équivalents espagnols sont inintelligibles : **sitio de la parodiar*, **el compartir del fichero*, **bandera del movimiento en sentido vertical*, **video que fluye*, **hábito que practica surf*, **fichero de la virus-firma*, **fichero compartir a compañía*, **Web que recibe a compañía*, **abastecedor conteno del Internet*. À noter que la majorité de ces combinaisons contiennent un verbe. Les traductions françaises sont aussi parfois incompréhensibles : **jet visuel*, **habitude surfante*, **annonce-dresseur*, **séquence type kiddy*, **entailler l'attaque*, **en différé parcourant*.
- Même si SYSTRAN n'est pas un programme de TA dédié à la traduction de la terminologie d'Internet, les traductions de certains termes pourtant très employés sont erronées ou inusitées (cf. EN *chatroom*, *thumbnail*, *home page*, *firewall*, *search engine*, *Internet Protocol*, *HTML site* > FR **site de HTML*, ES **sitio del HTML*, EN *Net economy* > FR **Economie nette*, ES **Economía neta*, EN *cybercriminal*, etc.).
- Des composés ou des éléments de composés anglais ont une traduction française (bonne ou mauvaise) mais pas espagnole (cf. *address book*, *bookmark*, *voice chat*, *techno-nerd*, *backdoor*). L'inverse est également vrai mais rare (cf. *video stream*).
- Les versions françaises et espagnoles sont quelquefois discordantes (cf. EN *virus writer* > FR *auteur de virus*, ES **programa de escritura del virus* ; EN *web surfer* > FR *webnute*, ES **persona que practica surf de Web* ; EN *scrolling banner* > FR **drapeau de défilement*, ES **bandera del movimeinto en sentido vertical* ; EN *surfing habit* > FR **habitude surfante*, ES **hábito que practica surf*, etc.).
- Dans les combinaisons espagnoles, l'article défini est très souvent inséré après la préposition *de*, ce qui est exceptionnel dans les traductions françaises (ES **programa del ataque*, **acceso del Web*, **carpeta de la historia*, **acceso del hacker*, **charla de la voz*, **sesión de la charla*, etc ; FR **fournisseur de l'espace*).
- Les traductions données par SYSTRAN qui sont des termes implantés sont assez rares dans les deux langues cibles : FR *carnet d'adresses*, FR *programme d'attaque*, FR *signet*, FR *goulot d'étranglement*, ES *anchura de banda*, ES *cortafuego*, FR *site de parodie*, FR *auteur de virus*, FR *modem satellite*, ES *módem de cable*, FR *émoticône*, FR *partage de fichier*, FR *cyberespace*, FR *cybercafé*. SYSTRAN « calcule » le composé cible à partir des éléments constitutifs du composé source.

3.2.2. IT Dictionary

La traduction vers l'espagnol n'étant pas le point fort de SYSTRAN, étudier un logiciel consacré à la traduction de la combinaison anglais-espagnol semble judicieux. Le test porte sur la même liste de noms composés mais seulement seize des unités testées sont connues de IT Dictionary.

¹⁷⁸ L'astérisque signale que l'équivalent proposé n'est pas employé par les internautes.

Tableau 101 : Test de IT Dictionary

Composés anglais	Traductions espagnoles
email address	dirección de correo electrónico, dirección de e-mail, dirección de email, dirección email
address book	libro de direcciones, agenda, libro de señas
backdoor	puerta trasera
bookmark	señal de lectura, marcador de libro, señal de libro
bottleneck	cuello de botella, embotellamiento, punto de reducción del pase
bulletin board	tablero de anuncios, tablero, pizarra donde se colocan los anuncios del día, tablilla de anuncios, tablón, tablón de anuncios, tablón donde se colocan los anuncios del día
cyberspace	espacio cibernetico, ciberespacio
dot com	de la Internet
home page	página principal del website
mail server	servidor de correo
mailing list	lista de correo, lista de direcciones, lista de direcciones de correo
popup	mensaje de aparición automática
search engine	motor de búsqueda, directorio de búsqueda en Internet
startup	empresa recién fundada
thumbnail	uña del dedo pulgar, uña del pulgar
Web browser	navegador de Internet, navegador

La lecture de ce tableau appelle un certain nombre de remarques :

- Cinquante-quatre composés sur les soixante-dix de notre échantillon – dont de nombreux termes clés – ne sont pas répertoriés dans IT dictionary qui est pourtant spécialisé dans les nouvelles technologies : *attack program, voice chat, Web access, chat session, chatroom, firewall, history folder, video stream, parody site, script kiddie, worm alert, audience measurement, network connection, discussion board, bug hunter, cookie builder, virus writer, web surfer, storage site, space provider, Internet protocol, cable modem, satellite modem, banner ad, HTML site, IP address, HTML page, Net access, Net economy, techno-nerd, ad-blocker, emoticon, file sharing, file swapping, browsing habit, hacking attack, scrolling banner, streaming video, surfing habit, hacker access, hotline, cybercriminal, cyber-café, search engine spider, virus-signature file, price comparison site, site creation tool, cable Internet access, file-sharing program, Web hosting company, Internet content provider, offline*

browsing, dial-up connection, peer-to-peer network.

- IT Dictionary fonctionne à partir d'une base de données. Par conséquent, aucune traduction n'est proposée pour les composés qui n'y figurent pas. Contrairement à SYSTRAN, il ne dissocie pas les éléments des composés sources pour 'fabriquer' le composé cible à partir de chacun des éléments constitutifs.
- Plusieurs traductions pertinentes sont parfois proposées pour un même composé (cf. *cyberspace, email address, Web browser, bookmark*, etc.).
- Certaines traductions comme **uña del dedo pulgar* et **embotellamiento* appartiennent à la langue générale et non au domaine d'Internet.
- Dans de nombreux cas le logiciel fournit plus une explication qu'une traduction (cf. EN *bulletin board* > ES **pizarra donde se colocan los anuncios del día, * tablón donde se colocan los anuncios del día*, EN *home page* > ES **página principal del website*, EN *dot com*, ES **de la Internet*, EN *startup* > ES **empresa recién fundada*, EN *popup* > ES **mensaje de aparición automática*).

3.2.3. TRADINFO

Ce traducteur instantané est spécialisé dans le domaine informatique. Il inclut peu de composés d'Internet, ce qui peut surprendre puisque les dictionnaires d'informatique comptent généralement une bonne proportion de termes d'Internet.

Tableau 102 : Test de Tradinfo

Composé anglais	Traductions françaises
bandwidth	bande passante ; largeur de canal ; largeur de bande
bookmark	signet, marque page
bottleneck	goulot d'étranglement
cyberspace	cyberespace, espace virtuel
firewall	coupe-feu ; bastion ; écluse
hotline	assistance téléphonique
IP address	adresse ip
mailing list	liste de diffusion ; liste de distribution
startup	démarrage
thumbnail	croquis minuscule

Tradinfo fournit des équivalents pour dix composés sur les soixante-dix sélectionnés.

- Aucun des composés de plus de deux éléments n'est recensé. Or, si on regarde la liste des composés connus de Tradinfo, les surcomposés de trois éléments sont très nombreux, ce qui n'est pas le cas des autres logiciels.
- La structure de Tradinfo est similaire à celle de IT Dictionary : les équivalents français et espagnols des termes anglais sont regroupés dans un glossaire.

- Pourquoi des notions-clés comme *home page*, *address book*, *mail server*, *startup* ne figurent-elles pas dans Tradinfo alors que d'autres – *mailing list*, *IP adresse*, *cyberspace*, *bookmark* – sont répertoriées ? On peut se demander si les concepteurs de Tradinfo ont effectué une taxinomie du domaine avant de créer leur base de données.
- Certains équivalents français sont peu ou pas employés par les internautes : **croquis minuscule*, **bastion*, **écluse*, **largeur de bande*, **largeur de canal*, **marque page*.
- Contrairement aux autres outils de TA testés, Tradinfo ne donne ni définition, ni contexte définitoire.

3.2.4. Babylon-Pro

Si deux combinaisons (anglais-français et anglais-espagnol ici) sont installées, Babylon-Pro présente l'avantage d'afficher simultanément les traductions françaises et espagnoles après validation de la saisie du terme anglais.

Tableau 103 : Test de Babylon-Pro : anglais-espagnol.

Composés anglais	Traductions espagnoles
address book	Libro de direcciones ; Lista de direcciones del correo electrónico
backdoor	Puerta trasera; Brecha, Modo no documentado de penetrar en un sistema de ordenadores o en una base de datos
bandwidth	(inform) la cantidad de datos que pueden ser transmitidos en una cantidad de tiempo fija, a través de una red de comunicación
bookmark	(inform) Señal de libro; (inform) Dirección de Internet en la lista de sitios preferidos por el usuario de WWW
bulletin board	cartel de anuncios
chatroom	cuarto de conversación en tiempo concreto real (internet)
cyberspace	espacio virtual, no físico que contiene todo lo que ocurre en un sistema de ordenadores y su alrededor
emoticon	(internet) especie de icono, símbolo en forma de cara con expresión como el smile
file sharing	ficheros compartidos ; la capacidad de permitir a un número de fichero en un mismo tiempo
firewall	(inform) sistema diseñado para prevenir el acceso no autorizado a o desde una red privada, en general para intrusos desde el internet
home page	Sitio familiar, Página de hogar, Documento de hipertexto en el internet consagrado al tema a la compañía o a una persona en particular
hotline	teléfono rojo
IP address	dirección el el internet, dirección numérica que se refiere a un nuevo computador en la red
mail server	servidor de correo, programación que reacciona automáticamente a pedidos enviados por correo electrónico
mailing list	Lista de envío, Lista de correo, Lista de operadores que reciben letras de un tema específico a través del servidor de listas de envío
search engine	Motor de búsqueda, Mecanismo de búsqueda, Mecanismo en el internet que permite la búsqueda de una información específica
startup	Carga de inicio del ordenador, Recarga del ordenador, Grupo que trabaja en el desarrollo de una nueva idea para un nuevo producto en una empresa
thumbnail	una miniatura visualizada en una página que va a ser impresa ; miniatura de imagen que permite ver el diseño de varias páginas en la pantalla al mismo tiempo tiempo ; uña del dedo pulgar, uña del pulgar

Tableau 104 : Test de Babylon Pro : anglais-français.

Troisième partie. La traduction des noms composés anglais du domaine d'Internet en français et en espagnol

Composés anglais	Traductions françaises
address book	dictionnaire d'adresses, Liste d'adresses de courrier électronique
backdoor	porte dérobée ; secret ; façon non documentée d'accéder à un système ou aux données (inform)
bandwidth	largeur de bande, rythme maximal de données transmissibles sur un réseau (inform)
bookmark	signet, marque ; adresse internet d'un fichier comprenant la liste des sites préférés par un utilisateur de WWW
bulletin board	tableau d'affichage
chatroom	salle de discussion
cyberspace	espace cybernétique, espace virtuel englobant tout ce qui se passe dans les systèmes informatiques et autour de ceux-ci (inform), espace cybernétique, espace virtuel, le lieu théorique où se trouvent les ordinateurs lorsqu'ils se branchent sur réseau informatique ou internet (inform, internet)
emoticon	(sur internet) sorte de logo ou de dessin en forme de visage reproduisant les sentiments ainsi que l'expression du visage (comme smile)
file sharing	partage de fichier, possibilité d'offrir à plusieurs ordinateurs un accès simultané au même fichier
firewall	système pour la protection des réseaux locaux reliés à Internet contre l'accès non autorisé (inform)
home page	site domestique, site d'origine, page domestique ; document hypertexte souvent consacré à un sujet, à une société ou à un individu
hotline	ligne directe, téléphone rouge
IP address	adresse Internet, adresse numérique indiquant un ordinateur particulier sur le réseau
mail server	programme réagissant automatiquement aux demandes envoyées par messagerie électronique
mailing list	liste de publipostage, liste de courrier, liste d'utilisateurs recevant des lettres sur un sujet déterminé à partir d'un serveur de publipostage
search engine	moteur de recherche, dispositif de recherche, dispositif sur internet permettant de rechercher une information spécifique
startup	groupe de mise en œuvre, groupe constitutif d'une société, démarrage, mise en route de l'ordinateur ; réinitialisation de l'ordinateur
thumbnail	image réduite, petite image représentant une image plus grande, plus détaillée ; ongle de pouce

Pour l'espagnol comme pour le français, Babylon-Pro recense dix-huit composés

anglais sur les soixante-dix testés (pas beaucoup plus que les deux traducteurs précédents). De manière générale :

- Quelques erreurs commises dans les traductions et gloses espagnoles montrent un manque de rigueur : EN *IP address* > ES *dirección el el internet [...]*, EN *thumbnail* > ES [...] *al mismo tiempo tiempo*.
- Ce logiciel a plus tendance à fournir une explication qu'une traduction : EN *emoticon* > FR *sorte de logo ou de dessin en forme de visage reproduisant les sentiments ainsi que l'expression du visage, EN *firewall* > FR *système pour la protection des réseaux locaux reliés à Internet contre l'accès non autorisé, EN *mail server* > FR *programme réagissant automatiquement aux demandes envoyées par messagerie électronique, EN *bandwidth* > ES *la cantidad de datos que pueden ser transmitidos en una cantidad de tiempo fija, a través de una red de comunicación, EN *firewall* > ES *sistema diseñado para prevenir el acceso no autorizado a o desde una red privada, en general para intrusos desde el internet, EN *chatroom* > ES *cuarto de conversación en tiempo concreto real, etc..
- Babylon-Pro accompagne généralement les traductions françaises et espagnoles de contextes (cf. : *backdoor*, *bandwidth*, *bookmark*, *cyberspace*, *file sharing*, *home page*, *IP address*, *mail server*, *mailing list*, *search engine*, *startup* pour l'espagnol et *backdoor*, *bandwidth*, *bookmark*, *cyberspace*, *file sharing*, *home page*, *IP address*, *mailing list*, *search engine*, *thumbnail* pour le français).
- Les résultats obtenus sont parfois maladroits ou difficilement intelligibles : EN *backdoor* > ES *modo no documentado de penetrar en un sistema de ordenadores o en una base de datos, EN *file sharing* > ES *la capacidad de permitir a un número de fichero en un mismo tiempo, EN *home page* > ES *documento de hipertexto en el internet consagrado al tema a la compañía o a una persona en particular, EN *backdoor* > FR *façon non documentée d'accéder à un système ou aux données.
- Certaines traductions ou définitions sont imprécises, confuses voire incorrectes : EN *bulletin board* > FR *tableau d'affichage, EN *cyberspace* > FR *espace cybernétique, EN *home page* > FR *site domestique, *site d'origine, *page domestique, EN *mail server* > FR *programme réagissant automatiquement aux demandes envoyées par messagerie électronique, EN *mailing list* > FR *liste de courrier.
- Quelques traductions données par Babylon-Pro sont peu employées par les internautes ou ne le sont pas du tout : FR *dictionnaire d'adresses, *liste d'adresses de courrier électronique, *secret, *largeur de bande, *marque, *espace cybernétique, *site domestique, *site d'origine, *page domestique, *ligne directe, *téléphone rouge, *liste de publipostage, liste de courrier, *dispositif de recherche, image réduite ; ES *brecha, *señal de libro, *sitio familiar, *página de hogar, *teléfono rojo, *mecanismo de búsqueda.

3.2.5. Bilan

Étant donné la piètre qualité générale des traductions des composés offertes par les

traducteurs instantanés, l'automatisation du système d'équivalences répond à un réel besoin. En résumé, les programmes de TA ont des fonctionnements variés. SYSTRAN « calcule » généralement les composés français et espagnols à partir des éléments des composés anglais saisis. Il recherche le sens des éléments dans son répertoire pour fabriquer les équivalents alors que IT Dictionary, Tradinfo et Babylon-Pro recherchent les termes saisis dans leur base de données. Ils ne traduisent que les mots répertoriés, les autres ne sont pas reconnus. SYSTRAN permet d'obtenir un plus grand nombre de traductions que les trois autres logiciels mais leur qualité laisse parfois à désirer. Ces logiciels de TA permettent de rendre globalement le sens d'un texte dans une langue cible mais ils sont inefficaces pour effectuer un travail terminologique. Les termes composés ne sont pas traduits avec régularité et précision.

Chapitre 3. Vers l'automatisation du système d'équivalences types anglais-français, anglais-espagnol

1. Future base du système d'équivalences types automatisé

Le système d'équivalences types va maintenant prendre la forme d'une base de données terminologiques trilingue (anglais, français, espagnol) de création néologique. L'outil informatisé sera appelé Générateur d'équivalents (GE). Sa conception sera développée dans les sections 1.1., 1.2. et 1.3. au moyen de schémas explicatifs simples brièvement commentés. Les données figurant dans ce chapitre sont nécessaires à l'utilisation du GE. Pour l'exploiter pleinement, l'utilisateur doit cerner certains aspects de son organisation interne. Les données et les étapes plus techniques, non indispensables à l'utilisation du GE, seront placées en Annexe. Les schémas retraceront les étapes de création guidées par des contraintes de pertinence d'utilisation (par rapport aux exigences du traducteur) et de faisabilité (sur le plan informatique). Les schémas conçus (Part. III, Ch. 3, Sec. 1.) constitueront une maquette. Lors de la conception et réalisation définitives et lors des essais, le GE sera sujet à des modifications (améliorations pour mieux répondre aux attentes du traducteur, rectifications imposées par des contraintes informatiques).

1.1. L'ossature générale

Plan. 30 – LA TRADUCTRICE DES NOMS COMPOSÉS ANGLAIS DU DOMAINE D'INTERNET EN FRANÇAIS ET EN ESPAGNOL.
UN LOGICIEL POUR FAIRE POSSIBLE L'ACCÈS À DES DOCUMENTS DANS DIFFÉRENTES LANGUES SUR LE WEB.

Schéma 1 : Ossature générale du GE

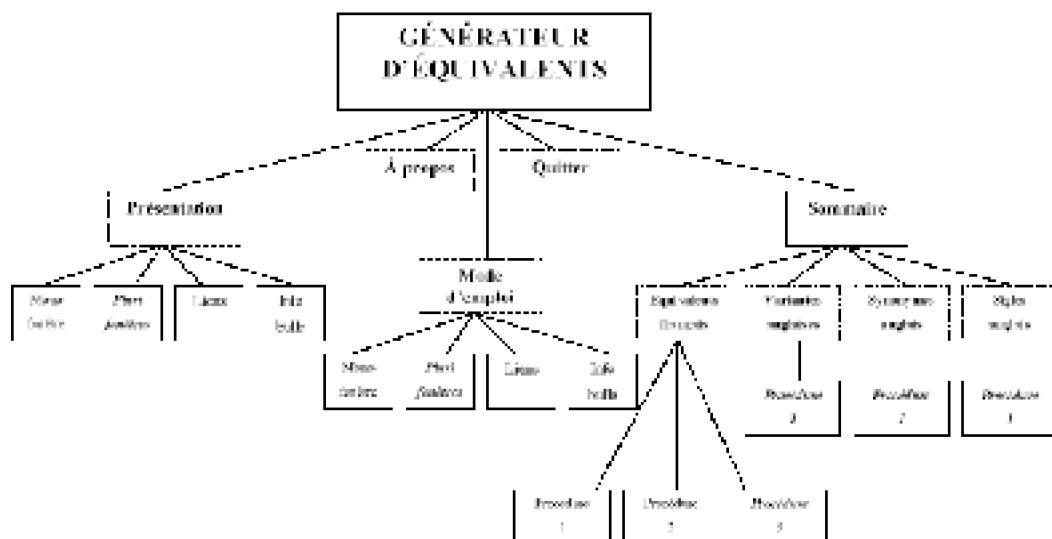


Schéma 1 : Ossature générale du GE

Légende : Cadre épais : application Cadre en petits points : fenêtre d'information
 Cadres en petits pointillés : fenêtres de présentation Cadres en pointillés irréguliers : boutons
 Cadres en lignes continues: options possibles pour la réalisation de la fenêtre
 Caractères gras + italique : option retenue parmi les options possibles

Le GE sera une application constituée de fenêtres de présentation et surtout d'information sur les composés anglais et leurs équivalents en langues cibles. La fenêtre dite d'information (Sommaire) donnera accès, grâce à des boutons de commande, à d'autres fenêtres d'information affichant les résultats des requêtes demandées. Les fenêtres dites de présentation (Présentation, Mode d'emploi, À propos) afficheront simplement des informations relatives à l'utilisation du GE.

Le schéma 1 retrace la première étape de la conception du GE : la constitution et l'organisation des options proposées au traducteur. Ses fonctions principales seront regroupées dans une fenêtre « Sommaire » à l'intérieur de laquelle des boutons de

commande permettront de naviguer dans les menus :

- Recherche des équivalents français et espagnols d'un composé anglais. Plusieurs procédures sont envisageables pour le traitement des composés et des équivalents. Elles seront étudiées en détail (Part. III, Ch. 3, Sec. 1.4.). Le GE sera construit à partir d'une base de données constituée de nombreuses tables, des requêtes et des formulaires.
- Recherche d'informations sur la vedette anglaise : existence de synonymes, de variantes et de sigles.

Un deuxième schéma, relatif à l'interface utilisateur, donnera une idée claire du fonctionnement du GE. Il précisera, d'une part, la fonction de chaque bouton et déterminera, d'autre part, l'ordre d'affichage des informations.

1.2. L'interface utilisateur

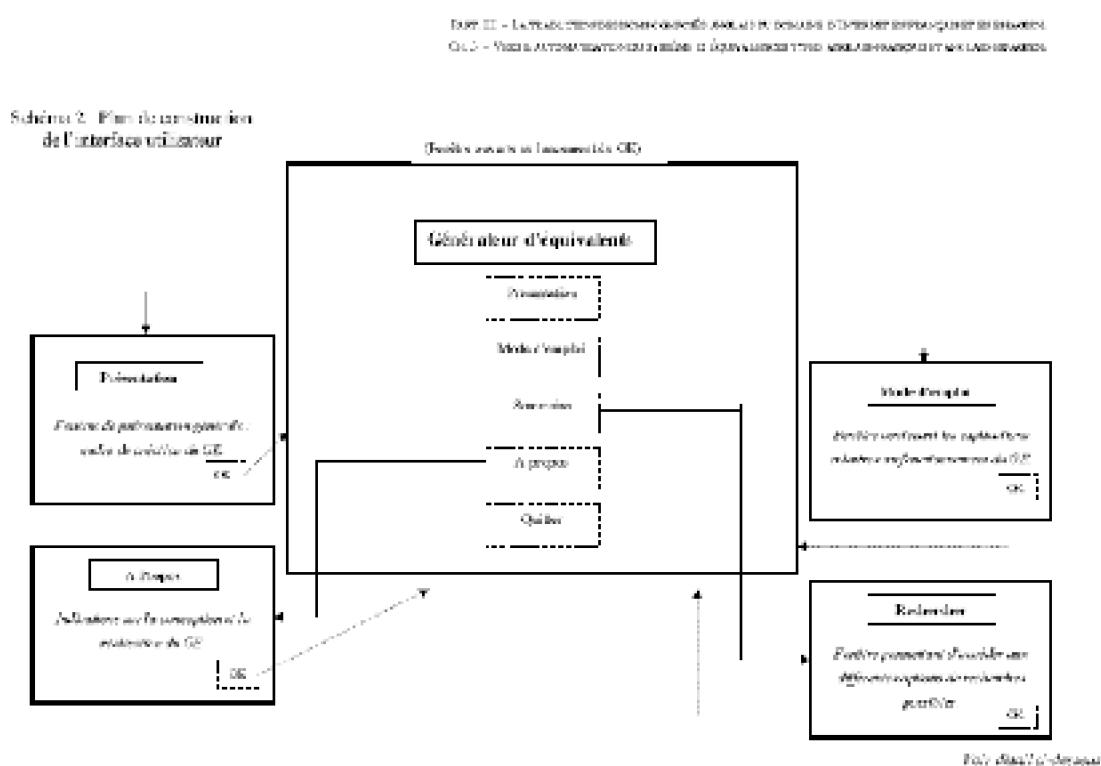
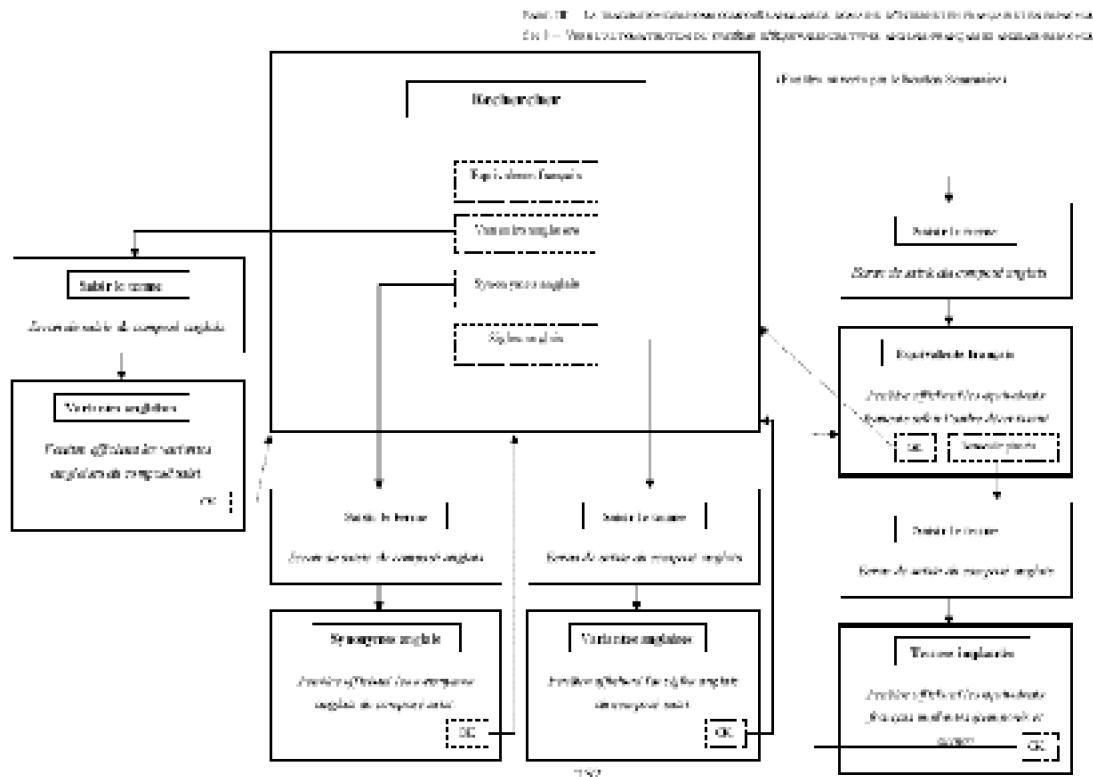


Schéma 2 : Plan de construction de l'interface utilisateur



Légende : Cadres épais : fenêtres explicatives ou d'affichage Cadres fins : boîtes de dialogue Cadres en pointillés : boutons Flèches pleines : ouverture d'une fenêtre Flèches en pointillés : retour à l'écran précédent

Texte en italique : résumés des informations contenues dans la fenêtre
 Cette maquette prévoit que l'ouverture du GE s'effectue sur une fenêtre de menu général, intitulée « Générateur d'équivalents », comportant cinq boutons dont voici les fonctions :

Boutons	Effets
1) Présentation	Ouvrira une fenêtre d'explications sur le cadre de création du GE, son but et son principe. OK permettra un retour au menu général.
2) Mode d'emploi	Affichera les informations nécessaires à l'utilisation du GE. Plusieurs fenêtres seront reliées les unes aux autres et l'utilisateur naviguera aux moyens de boutons Page suivante et Page précédente. OK permettra un retour au menu général.
3) Sommaire	Ouvrira la fenêtre « Rechercher » qui comportera quatre fonctions renfermant des informations terminologiques : Équivalents français, Variantes anglaises, Synonymes anglais et Sigles anglais.
A) Équivalents français	Ouvrira une boîte de dialogue dans laquelle l'utilisateur saisira le composé dont il souhaite connaître les équivalents. OK entraînera l'affichage des équivalents. L'utilisateur pourra ensuite retourner au sommaire de la fenêtre « Rechercher » ou bien consulter les termes implantés pour cette même vedette en cliquant sur Termes implantés. ¹⁷⁹
B) Variantes anglaises	Ouvrira une boîte de dialogue dans laquelle l'utilisateur saisira le composé dont il souhaitera connaître les variantes éventuelles. OK retournera au sommaire.
C) Synonymes anglais	Ouvrira une boîte de dialogue dans laquelle l'utilisateur saisira le composé dont il souhaitera connaître les synonymes éventuels. OK retournera au sommaire.
D) Sigles anglais	Ouvrira la boîte de dialogue dans laquelle l'utilisateur saisira le composé dont il souhaitera connaître les sigles éventuels. OK retournera au sommaire.
4) À propos	Ouvrira une fenêtre relative à la conception et la réalisation du GE. OK retournera au menu général.
5) Quitter	Mettra fin à la session ouverte dans le GE.

Un troisième schéma, consacré à l'option de recherche des équivalents en langue cible, nous permettra d'affiner la fonction la plus importante du GE.

¹⁷⁹ Le système d'équivalences est construit à partir du lexique d'Internet existant. Certains des termes proposés par le GE ne correspondent pas aux termes employés par les internautes. Il est possible d'accéder aux termes attestés en cliquant sur le bouton « Termes implantés ».

1.3. La fonction « Recherche des Équivalents français »¹⁸⁰

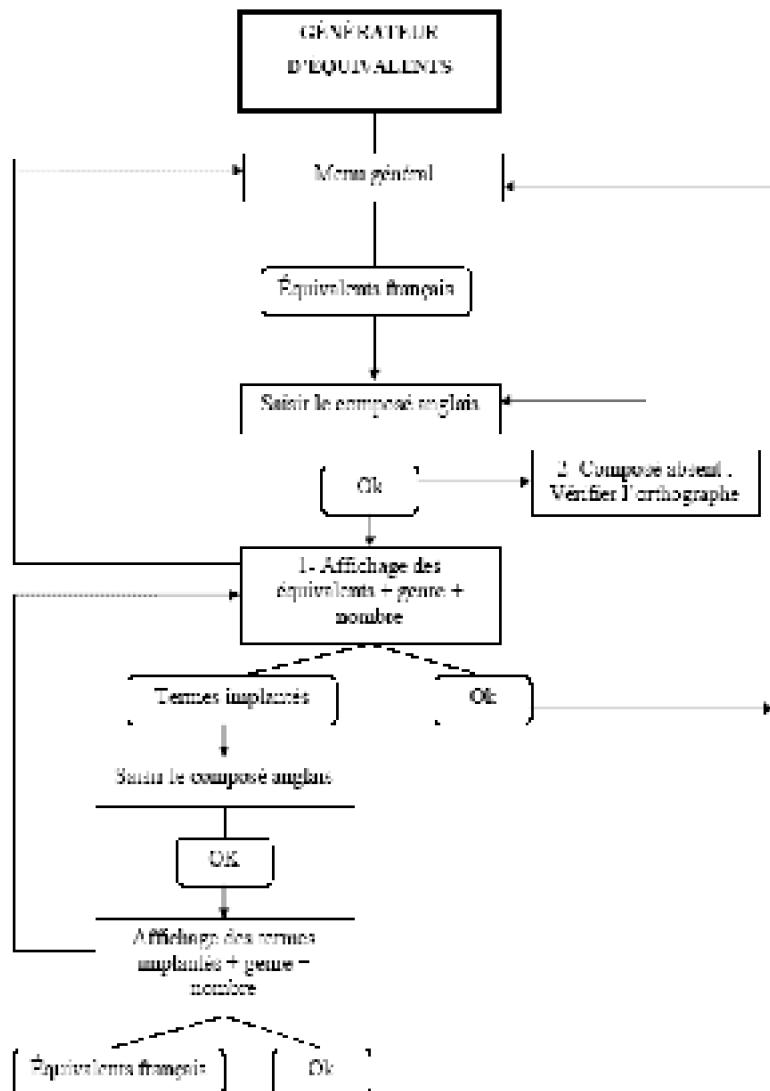


Schéma 3 : Fonction « Équivalents français »

Légende : Cadres rectangulaires : fenêtres Cadres à bords arrondis : boutons (présents dans la fenêtre située au dessous du cadre à bords arrondis) Flèches pleines : ouverture d'une fenêtre Flèches en pointillés : possibilité de retour Ok : validation

Pour obtenir l'affichage des équivalents d'un composé anglais, plusieurs étapes sont nécessaires :

¹⁸⁰

Une fonction similaire permettra l'accès aux équivalents espagnols. Le temps que nécessite la création de la base de données et du GE nous contraint à nous consacrer à une langue cible, dans un premier temps. Nous optons donc pour notre langue maternelle.

Bouton	Effet
Équivalents français	Ce bouton du Menu général ouvrira une boîte de dialogue dans laquelle l'utilisateur saisira le composé anglais dont il souhaitera connaître les équivalents.
OK	Validera la saisie du composé. Après la validation, il existera une alternative : le GE reconnaîtra le terme et affichera les équivalents français ou le GE ne reconnaîtra pas le terme saisi et affichera une boîte de dialogue « Terme absent » dans laquelle il faudra ressaisir la vedette (suppression des fautes, des erreurs de frappe). ¹⁸¹ Ensuite, l'utilisateur pourra accéder aux termes implantés ou fermer la fenêtre.
Termes implantés	(option de la fenêtre d'affichage des équivalents) Ouvrira une boîte de dialogue dans laquelle l'utilisateur saisira le composé anglais pour lequel il souhaitera connaître les termes implantés. Affichera les termes implantés autres que ceux correspondant aux modèles compositionnels les plus fréquents (le GE est fondé sur le lexique existant). Après consultation, l'utilisateur pourra retourner à la fenêtre d'affichage des équivalents pour effectuer une nouvelle recherche ou retourner au Menu général.

L'élaboration de schémas permet de déterminer les informations qui devront figurer dans le GE et d'évaluer le travail à effectuer pour informatiser le système d'équivalences. Le GE sera évolutif, il sera possible d'actualiser la base de données (insertion de nouvelles entrées). Si les tendances de traductions des noms composés anglais en français et en espagnol sont modifiées, le GE sera adaptable et le traducteur n'aura pas besoin de créer lui-même des néologismes.

1.4. Un SGBD pour exploiter les équivalences types

Avant de créer une base de données, nous devons considérer l'objectif qu'elle doit permettre d'atteindre. Pour cela il faut :

- recenser les informations à stocker dans la base en les répartissant dans différentes tables,
- définir les éventuelles liaisons entre les différentes tables,
- éviter la redondance des informations pour ne pas alourdir la base et éviter les erreurs de saisie et d'extraction.

Plusieurs procédures de réalisation ont été étudiées pour déterminer la plus appropriée au GE. Seule la structure retenue, « base de données avec requêtes Union », est

¹⁸¹ Nous envisageons d'ajouter un bouton « Équivalences types » qui permettra au traducteur de consulter les statistiques pour former lui-même un composé en langue cible si le terme anglais saisi est inconnu du GE.

présentée ici. Les deux autres procédures analysées figurent en Annexe (Annexe 11).

1.4.1. Travail préliminaire

- Calculer le nombre de structures utilisées en langue cible pour rendre une structure anglaise.
- Construire des tables identiques contenant les équivalents en langue cible. Chacun des patrons compositionnels a sa propre table découpée de la manière suivante :
 - Enregistrement 1
 - Champ n°1 : n°auto
 - Champ n°2 : Vedette anglaise
 - Champ n°3 : Équivalent
 - Champ n°4 : Fréquence
 - Champ n°5 : Genre
 - Champ n°6 : Nombre

Note : Dans les tables, les champs se répètent pour chacune des 55 vedettes anglaises.

- Construire une table contenant les informations relatives à la vedette anglaise, c'est-à-dire les éventuels synonymes, variantes et sigles, est utile pour fournir plus de renseignements terminologiques au traducteur. La table est organisée de la manière suivante :
 - Enregistrement 1
 - Champ n°1 : n°auto
 - Champ n°2 : Vedette anglaise
 - Champ n°3 : Synonyme 1
 - Champ n°4 : Synonyme 2
 - Champ n°5 : Synonyme 3
 - Champ n°6 : Variante 1
 - Champ n°7 : Variante 2
 - Champ n°8 : Variante 3
 - Champ n°9 : Sigles
- Construction de tables identiques pour les termes implantés (nombre fixé à deux) :
 - Enregistrement 1
 - Champ n°1 : n° auto
 - Champ n°2 : Vedette anglaise

- Champ n°3 : Terme implanté
- Champ n°4 : Fréquence
- Champ n°5 : Genre
- Champ n°6 : Nombre
- Créer un champ avec la liste de toutes les fréquences qu'un équivalent en langue cible peut avoir (de 0,01 à 100%).
- Créer des requêtes de types différents : paramétrées, Sélection, Union.
- Créer des formulaires d'entrée de données pour saisir les enregistrements dans les tables, d'affichage des résultats, de menu général pour ouvrir d'autres formulaires et des boîtes de dialogue pour exécuter des actions.

1.4.2. Fonctionnement du GE : Recherche des équivalents (fonction principale)

- Lancer l'application ¹⁸²
- La fenêtre *Menu général* du GE s'ouvre
- Cliquer sur le bouton *Équivalents français*
- La boîte de dialogue *Saisie du composé anglais* s'ouvre
- Saisir le composé dans la boîte de dialogue
- Cliquer sur le bouton OK
 - Une requête paramétrée isole la vedette dans les tables et champs
 - Une requête Union :
 - * réunit les champs « équivalents », « fréquence », « genre » et « nombre » issus des tables « équivalents » dans un champ.
 - * classe les équivalents selon leur fréquence (ordre décroissant).
- La fenêtre Affichage des équivalents français s'ouvre
- Cliquer sur le bouton Termes implantés de la fenêtre d'affichage des équivalents (option facultative)

Ou

- Fermer la fenêtre pour effectuer un retour au *Menu général*

Ou

- Quitter l'application

¹⁸² Caractères gras = opérations visibles pour l'utilisateur. Caractères normaux = opérations sous-jacentes

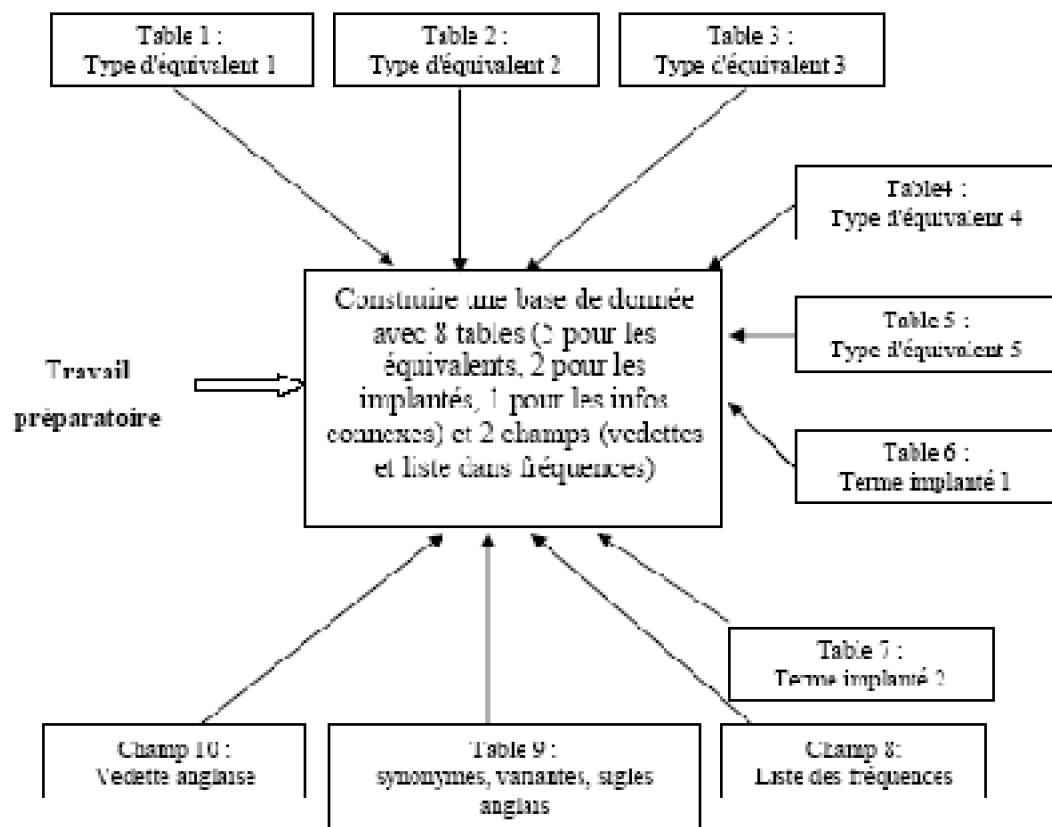
1.4.3. Avantages

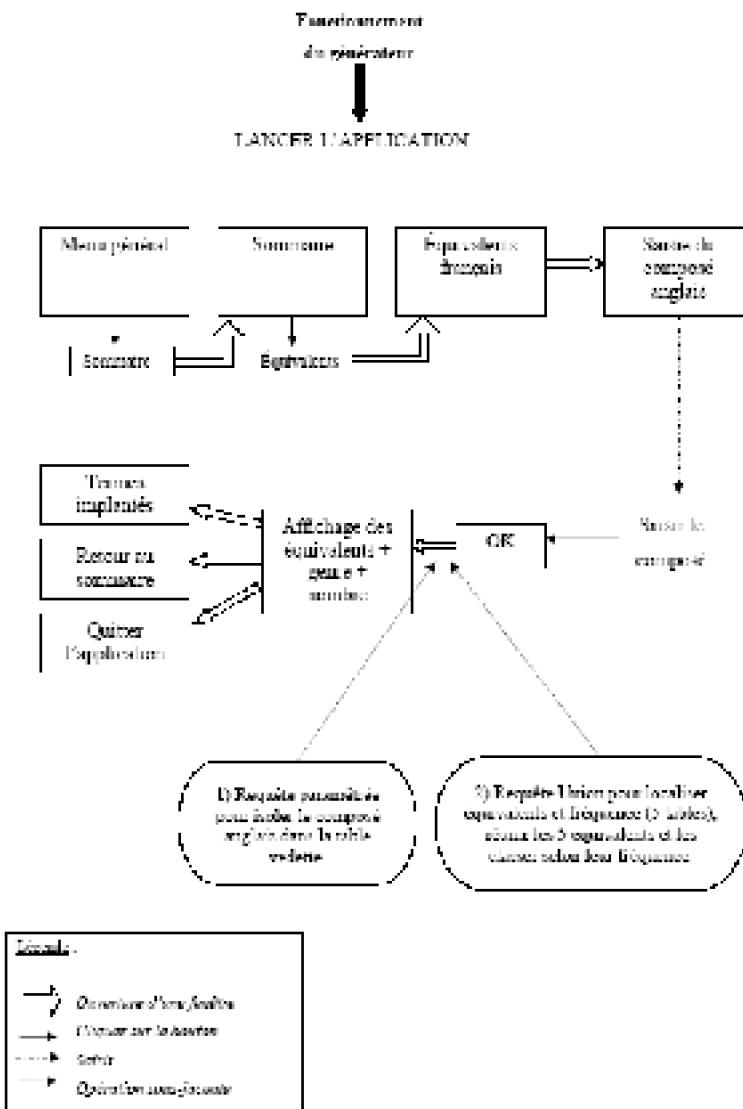
- Les tables et les champs pourront être actualisés simplement (requête Action/Ajout/Mise à jour). Aucune modification profonde de la base de données ne sera nécessaire pour l'ajout de nouveaux enregistrements dans les tables et pour les éventuelles modifications de fréquence des équivalences. La structure est donc souple et évolutive.
- Le GE ainsi construit est transposable à d'autres domaines.
- La vitesse de fonctionnement du GE est optimale.
- La programmation est légère.
- Les données peuvent être importées d'une table (ou de n'importe quel fichier) à une autre, ce qui allège la phase de saisie.
- Le principe de fonctionnement du GE est original. Il se différencie de celui des dictionnaires électroniques puisqu'il est construit sur la fréquence d'utilisation des équivalences anglais-français et anglais-espagnol.

1.4.4. Inconvénients

Pour donner x informations il faut x champs. Il est impossible d'avoir moins de champs. Le GE contient huit tables et champs différents qui sont à remplir lors de la saisie : équivalents 1, 2, 3, 4 et 5, termes implantés 1 et 2, table principale (comportant les données relatives à la vedette anglaise).

1.4.5. Schémas





1.5. La base de données

Comme l'illustrent les trois procédures (Part. III, Ch. 3, Sec. 1.4. et Annexe 11), dans un fichier de base de données des objets de diverses natures – dont nous ne donnerons ici que de brèves indications – permettent d'exploiter les informations. Les requêtes concernent l'extraction des données. Les formulaires concernent l'entrée et la restitution des données extraites. Quelques détails de la création des tables, les formulaires et les requêtes du GE figurent en Annexe (Annexe 12).

1.5.1. Les tables

Chaque table contient les données se rapportant à un sujet particulier. Dans le cadre du GE, une table ‘Équivalent français’, par exemple, contiendra la liste des vedettes anglaises et de leurs équivalents respectifs, la fréquence de la structure des équivalents, leur genre et leur nombre.

La base de données du GE contiendra dix-neuf tables (Annexe 12) : cinq tables ‘équivalents français’, cinq tables ‘structures compositionnelles’, une table ‘exceptions’, deux tables ‘termes implantés’, une table ‘champs notionnels’, une table ‘fréquence’, une table ‘générale’, une table ‘vedettes anglaises’. Les requêtes et les formulaires créés ultérieurement reposeront sur ces tables.

1.5.2. Les requêtes

Les requêtes seront utilisées pour interroger les données d'une ou plusieurs tables. Les informations sont présentées sous la forme d'une feuille de données ou, dans le GE, à l'intérieur d'un formulaire. Par exemple, lorsque l'utilisateur saisira la vedette anglaise, la requête sélectionnera les équivalents français dans les cinq tables ‘Équivalents français’ et les classera par ordre décroissant de leur fréquence. Lorsqu'une requête sera exécutée, les enregistrements de la table source seront sélectionnés selon les critères établis (Annexe 12).

1.5.3. Les formulaires

Dans le cadre du GE, trois types de formulaires seront utilisés pour :

- Saisir les données dans les tables,
- Afficher les informations (Guide de l'utilisateur, Présentation, A propos, Menu général, les sommaires comme ‘Informations connexes’),
- Afficher les résultats (équivalents français, variantes anglaises, champs notionnels, etc.).

2. Présentation du Générateur d'équivalents

Rappelons les faits les plus importants concernant la gestion des équivalences types par le GE :

- Le GE propose les traductions des composés anglais selon l'ordre décroissant de la fréquence des structures compositionnelles anglais-français.
- La fréquence des structures a été mesurée à partir des formants des composés du vocabulaire d'Internet existant. Les propositions de traduction n'ont pas été fabriquées arbitrairement mais d'après les données observées dans des textes spécialisés (corpus trilingue).
- Le GE proposant plusieurs traductions pour chaque composé anglais, il y a 99% de chance que la bonne traduction y figure.
- Le GE propose des équivalents même dans le cas des composés anglais exocentriques. Il gère les exceptions.
- Dans quelques cas, des particularités comme l'implantation d'un emprunt (*cookie*, par exemple) entravent l'apparition d'une traduction 100% endogène. Cependant, étant

donné le nombre considérable de termes hybrides relevés dans le corpus, on l'a vu les traductions hybrides sont satisfaisantes (cohérence avec le vocabulaire existant utilisé).

Au lancement du GE¹⁸³ apparaît le Menu général qui est composé de sept boutons :¹⁸⁴ **Présentation**, **Guide de l'utilisateur**, **À propos**, **Équivalents français**, **Champs notionnels**, **Équivalences types** et **Quitter** (Figure 8).

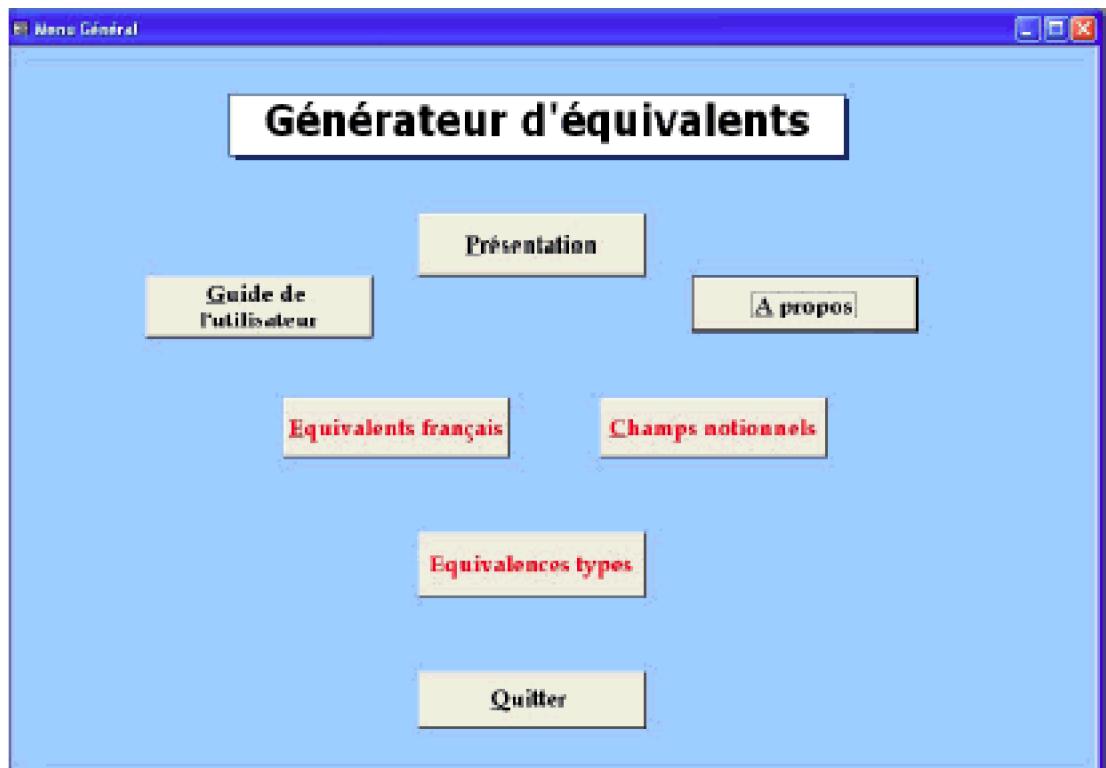


Figure 8 : Menu général

Présentation, **Guide de l'utilisateur** et **À propos** ouvrent des fenêtres fournissant des indications générales sur le GE (Figures 9 à 11).

¹⁸³ La découverte du fonctionnement du GE se fait grâce à la présente fiche explicative qui offre des exemples variés et est illustrée de captures d'écran.

¹⁸⁴ Dans le corps du texte, les boutons sont signalés par l'utilisation de caractères gras.

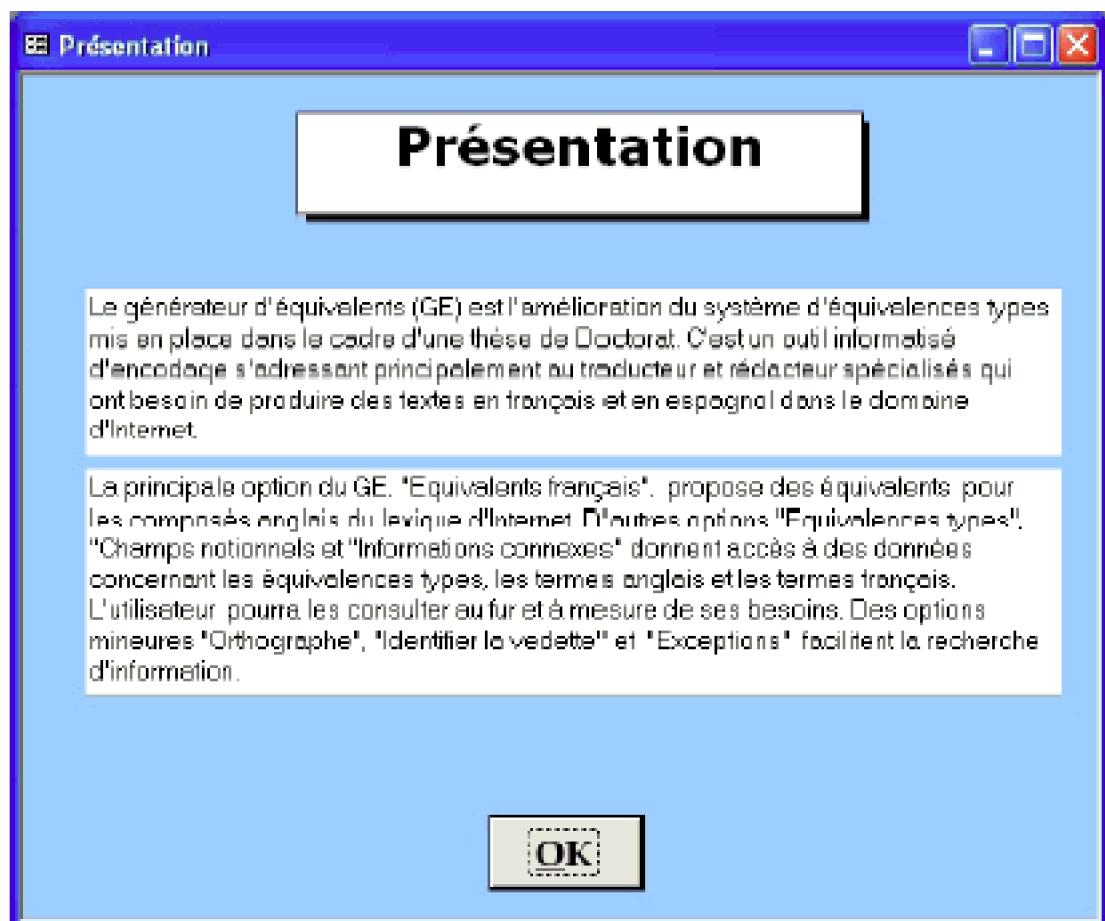


Figure 9 : Fenêtre Présentation

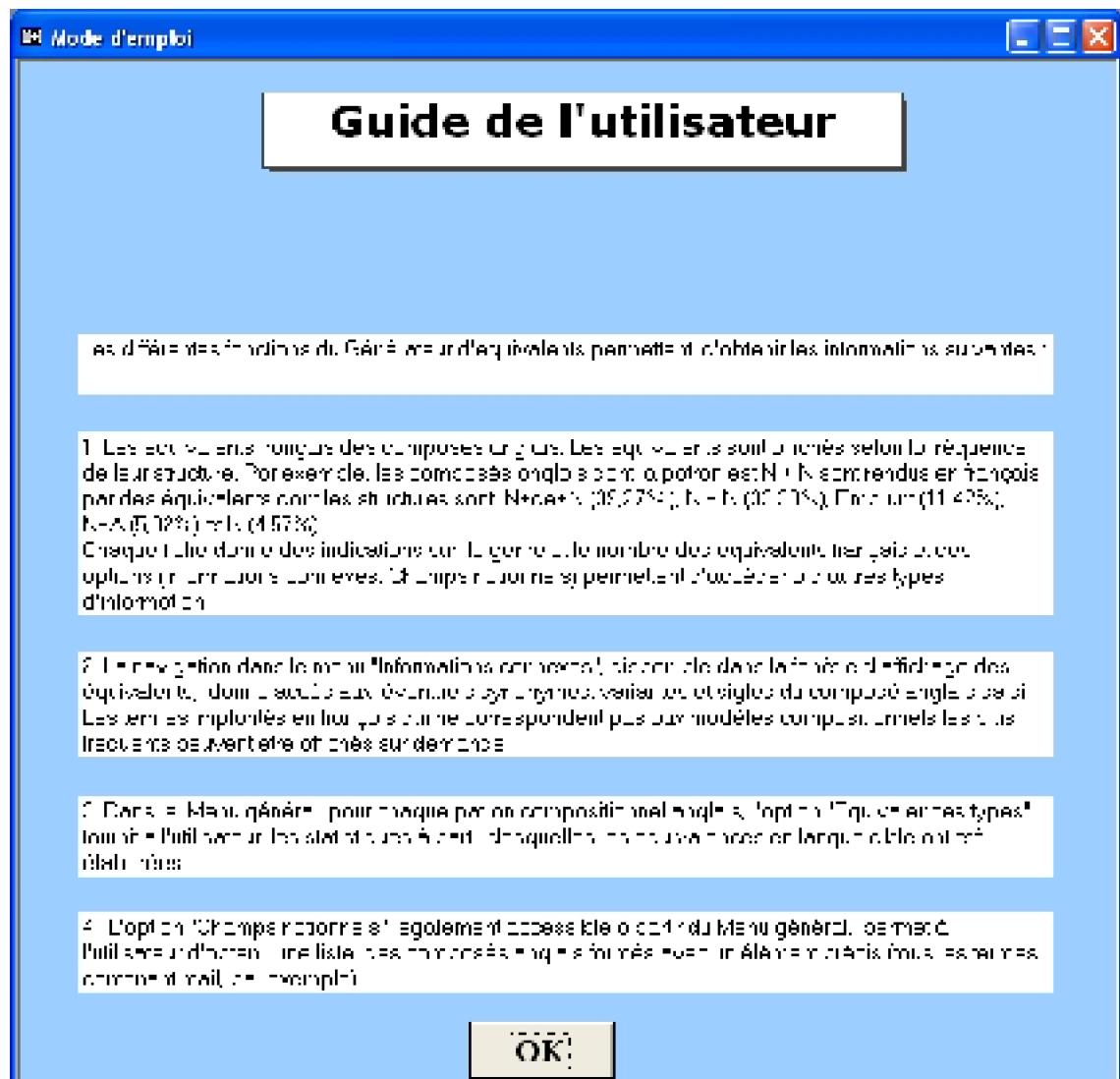


Figure 10 : Fenêtre Guide de l'utilisateur

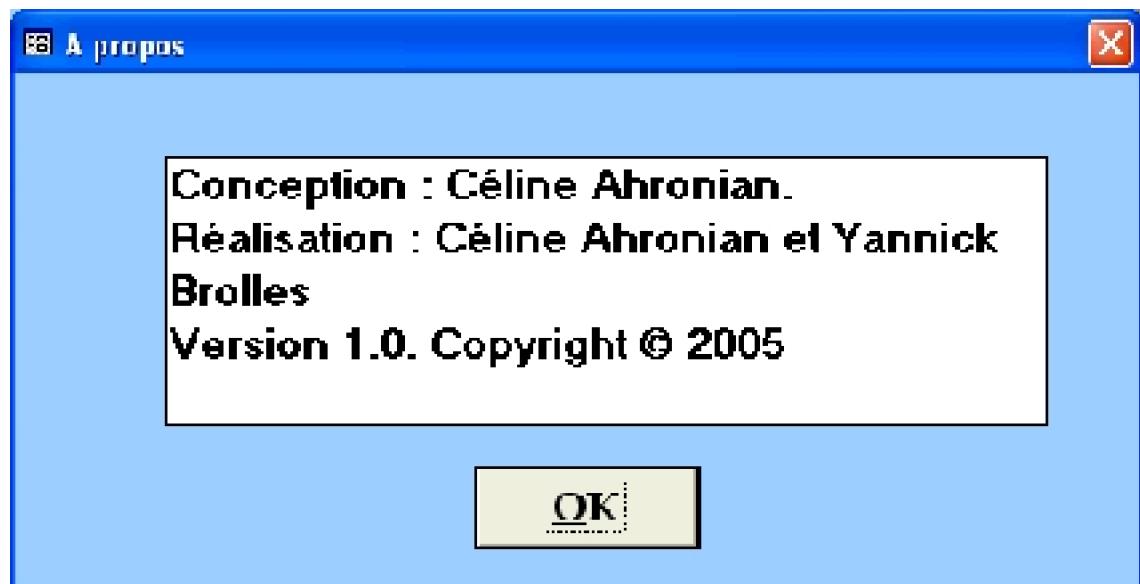


Figure 11 : Fenêtre À Propos

Les fonctions **Équivalents français**, **Équivalences types** et **Champs notionnels**, que nous allons présenter conjointement aux options liées à la recherche des équivalents, permettent la recherche d'informations sur les composés anglais et les termes français. Seront ensuite introduites les options relatives au menu Informations connexes, dont l'accès est réalisé à partir de la fenêtre d'affichage des équivalents.

Cliquer sur **Équivalents français** ouvre une boîte de dialogue dans laquelle l'utilisateur saisit le composé¹⁸⁵ dont il souhaite connaître les équivalents (Figure 12).



Figure 12 : Saisie du composé

Après saisie et validation de la vedette anglaise, *address book* dans l'exemple, la fenêtre d'affichage des équivalents français s'ouvre en dévoilant cinq propositions :¹⁸⁶ #carnet d'adresses, #carnet(-)adresses, *address book*, *carnet* et *carnet relationnel* (Figure 13).

¹⁸⁵ Si le composé saisi et validé est reconnu par le GE, il s'agit alors d'une 'vedette', c'est-à-dire d'une entrée du GE. Si aucun équivalent ne s'affiche après saisie et validation d'un composé bien orthographié, le concept à traduire est désigné par une autre dénomination. Une procédure permet de trouver la vedette désignant le concept (Figures 17 à 19).

¹⁸⁶ Les équivalents précédés du symbole # sont attestés en français.

Équivalents	Genre	Nombre
# carnets d'adresses	masc	sg
# carnet(-)d'adresses	masc	sg
address book	masc	sg
carnet	masc.	sg
carnet relationnel	masc	sg

INFORMATIONS :

- Si aucun équivalent n'apparaît :
- Le composé est peut-être mal orthographié, vérifiez en cliquant sur "Orthographe".
- Le composé saisi n'est peut-être pas une vedette, testez-le en cliquant sur "Identifier la vedette".
- Une fois les équivalents consultés, vous pouvez :
 - Recommander l'information complémentaire en cliquant sur "Informations connexes".
 - Retourner au Menu général en fermant le fenêtre.
- Si le bouton "Exceptions" apparaît dans la colonne "Équivalents", cliquez sur le bouton "Exceptions" pour connaître les équivalents.

Orthographe **Informations connexes** **Exceptions** **Identifier la vedette**

Figure 13 : Fenêtre d'affichage des équivalents français

Ces équivalents sont présentés dans l'ordre décroissant de la fréquence des structures françaises équivalentes au patron compositionnel du composé anglais saisi (N+N pour *address book*).

Si après saisie et validation de la vedette anglaise dans la boîte de dialogue, aucun équivalent ne figure dans la fenêtre d'affichage, l'utilisateur dispose de deux recours (commentés dans le cadre « Information ») (Figure 14).

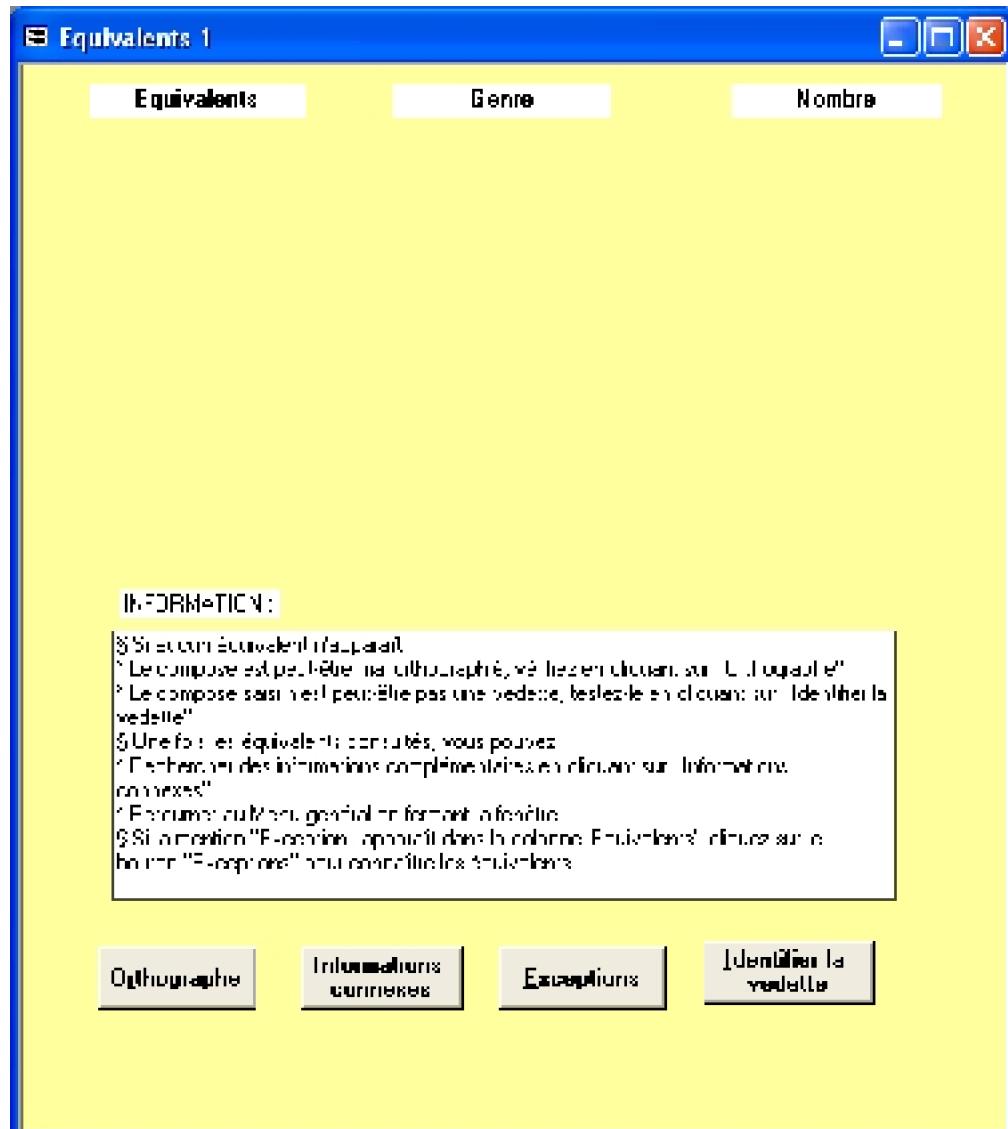


Figure 14 : Fenêtre d'échec d'affichage des équivalents

Le composé a pu être mal orthographié ou n'est pas peut-être une vedette. Pour procéder à une vérification orthographique, cliquer sur **Orthographe** (Figure 15).

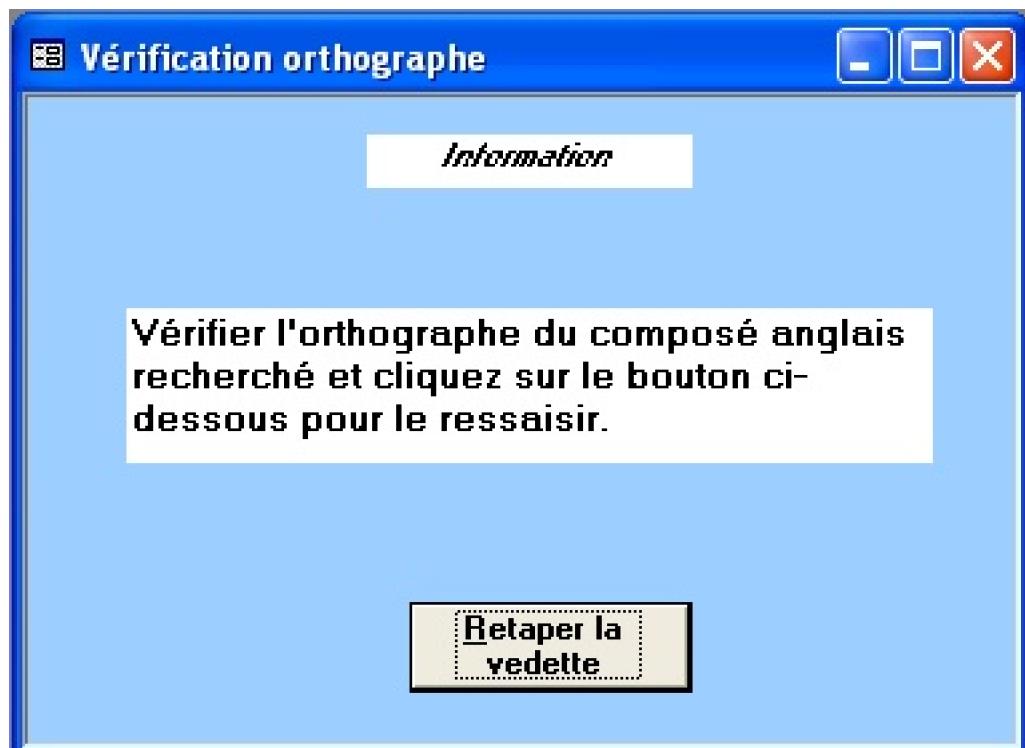


Figure 15 : Fenêtre de vérification orthographique

Après contrôle de l'orthographe, cliquer sur **Retaper la vedette** et saisir le composé recherché. Si la fenêtre d'affichage des équivalents reste vide (Figure 16), le composé saisi n'est peut être pas une vedette.

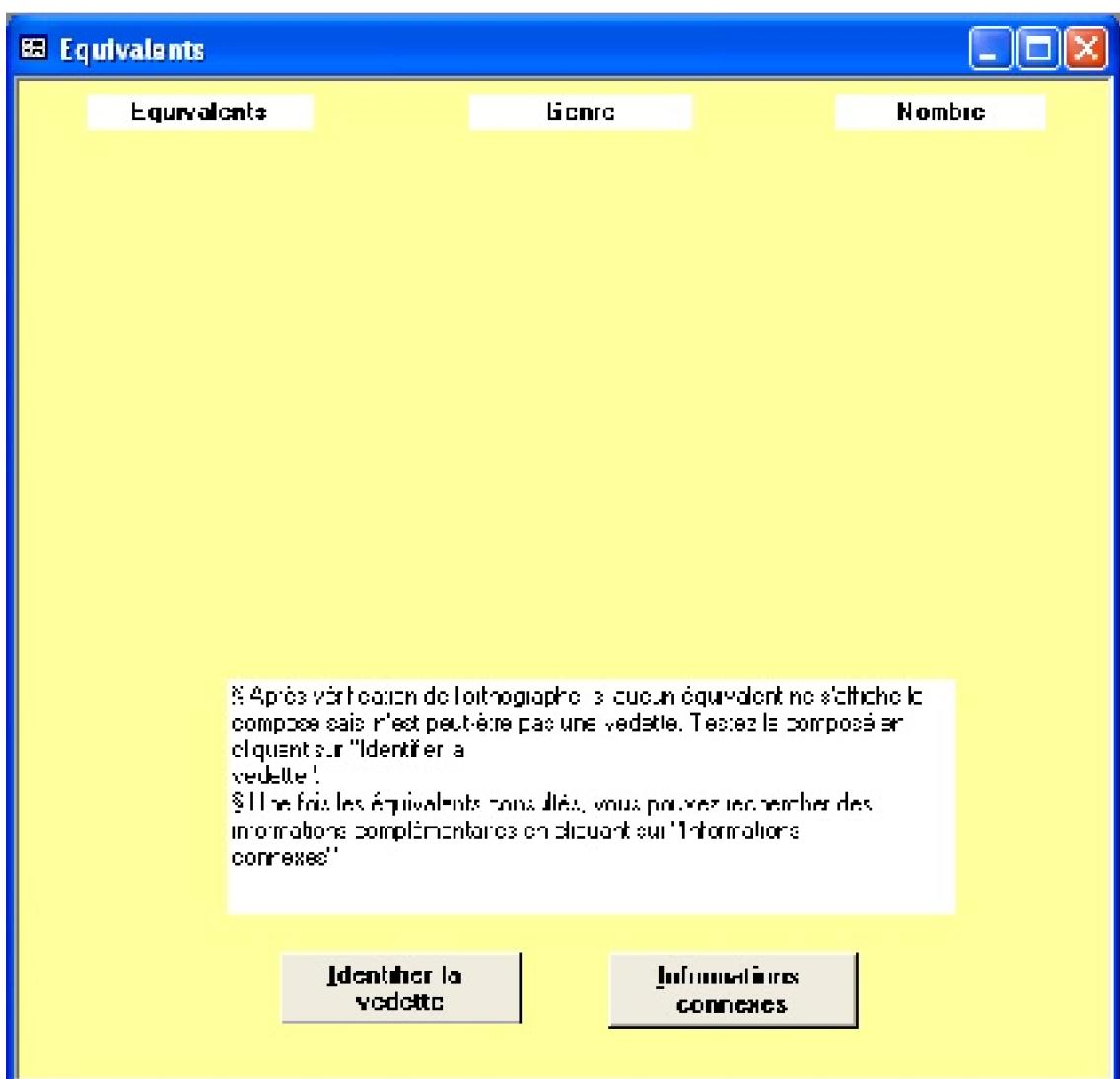


Figure 16 : Fenêtre d'échec d'affichage des équivalents après vérification orthographique

Le concept correspondant au terme recherché est connu sous une autre dénomination par le GE si la fenêtre des équivalents est vide bien que le nom composé soit correctement orthographié. Pour identifier la vedette correspondante, cliquer sur **Identifier la vedette** (Figure 17).

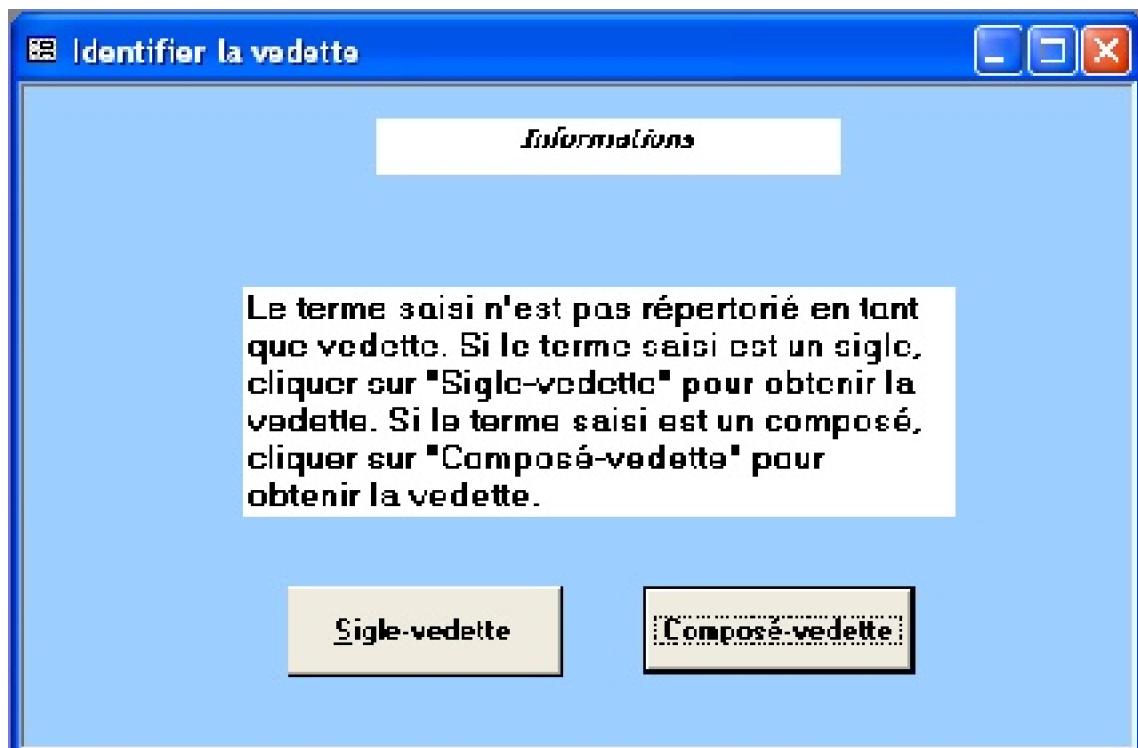


Figure 17 : Fenêtre d'identification de la vedette

Si le terme anglais recherché est un composé, utiliser l'option **Composé-vedette** et s'il s'agit d'un sigle opter pour **Sigle-vedette**. Nous illustrerons la fonction composé-vedette avec *Bulletin Board System* que le GE ne reconnaît pas comme vedette. Cliquer sur **Composé-vedette** et saisir *Bulletin Board System* dans la boîte de dialogue (Figure 18).



Figure 18 : Boîte de dialogue de saisie du composé

Après validation, il s'avère que pour ce concept, la vedette connue par le GE est *message board* (Figure 19).

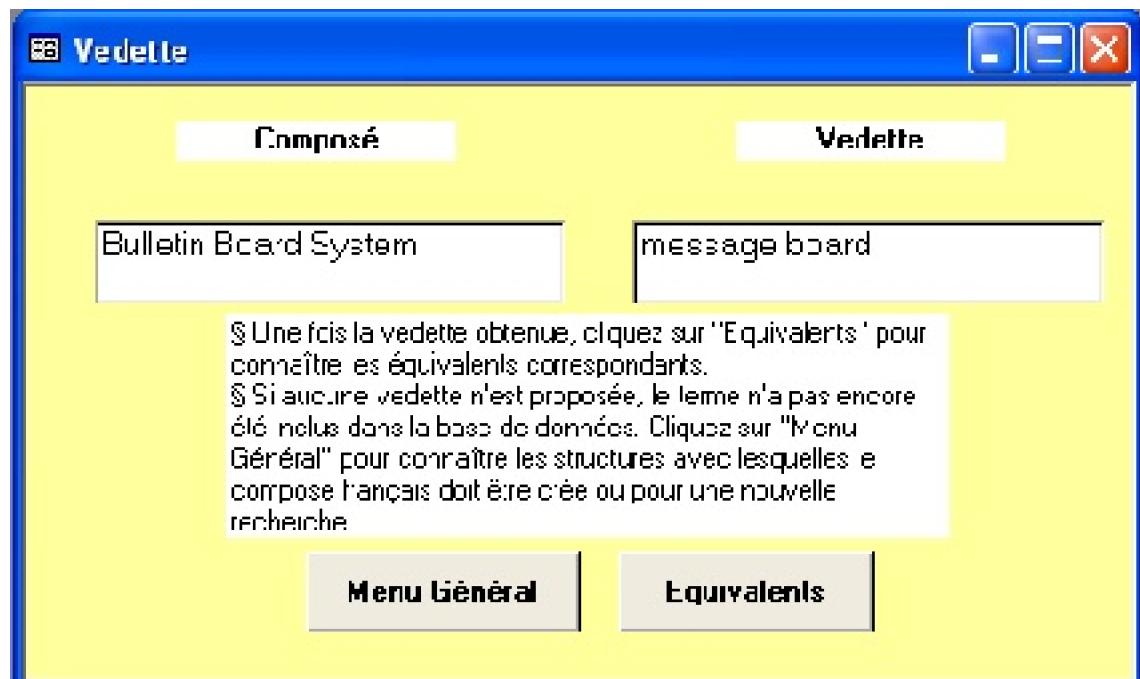


Figure 19 : Fenêtre d'affichage de la vedette anglaise

Pour obtenir les équivalents français désignant le concept, cliquer sur **Équivalents** et saisir *message board* dans la boîte de dialogue (Figure 20).



Figure 20 : Boîte de dialogue de saisie du composé

Après validation, la fenêtre contenant cinq équivalents français s'affiche (Figure 21).

187 Si aucune vedette n'apparaît, cliquer sur **Menu général** puis sur **Structures compositionnelles** pour connaître les équivalences types à partir desquelles construire l'équivalent français (Figures 26 et 27).

Les noms composés anglais français et espagnols du domaine d'Internet :

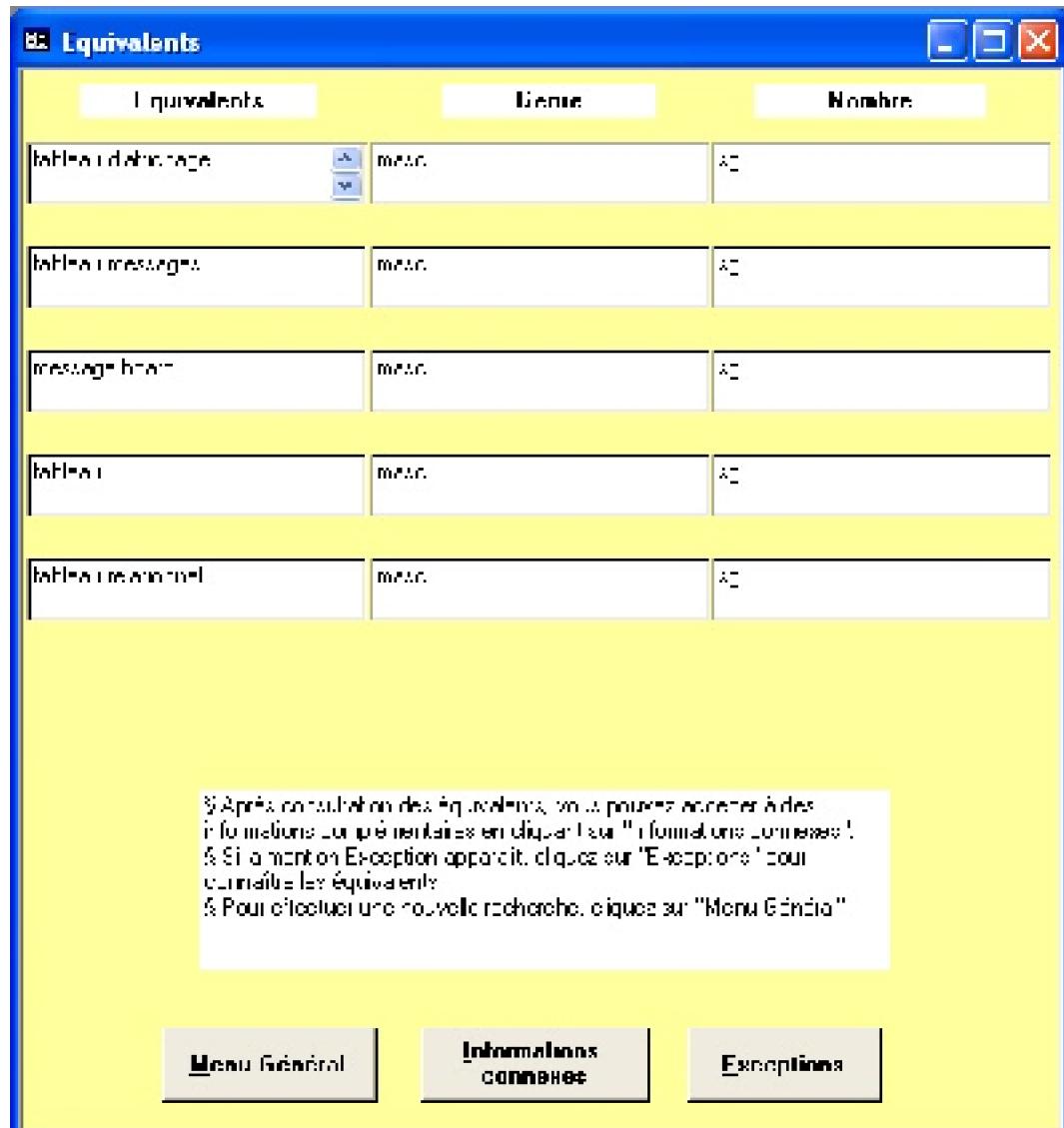


Figure 21 : Fenêtre d'affichage des équivalents

Si le terme anglais qui n'est pas une vedette du GE est un sigle, cliquer sur **Sigle-vedette**. Saisir *BtoB* ou *B2B* dans la boîte de dialogue (Figure 22).



Figure 22 : Boîte de dialogue de saisie du sigle

La validation fait apparaître que la vedette *Business-to-Business* correspond au concept recherché (Figure 23).

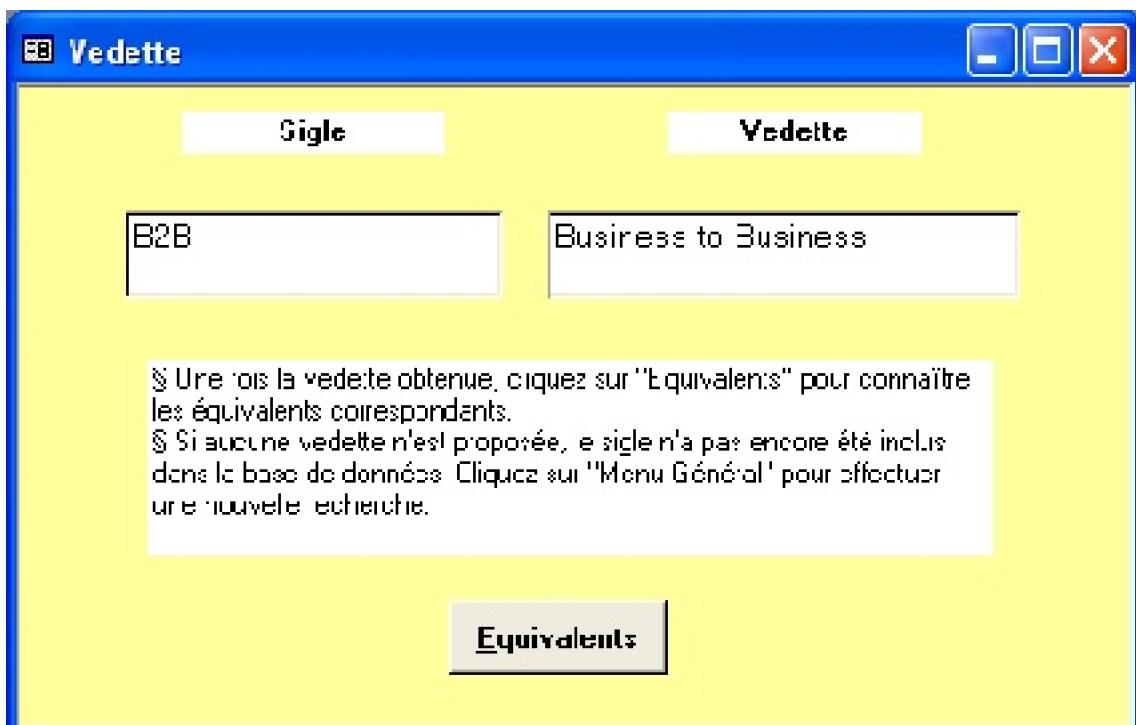


Figure 23 : Fenêtre d'affichage de la vedette anglaise

Pour obtenir les équivalents français, cliquer sur **Équivalents** et saisir *Business-to-Business* dans la boîte de dialogue (Figure 24).



Figure 24 : Boîte de dialogue de saisie du composé

Après validation, la fenêtre contenant un équivalent français (#*Business-to-Business*) s'affiche (Figure 25).

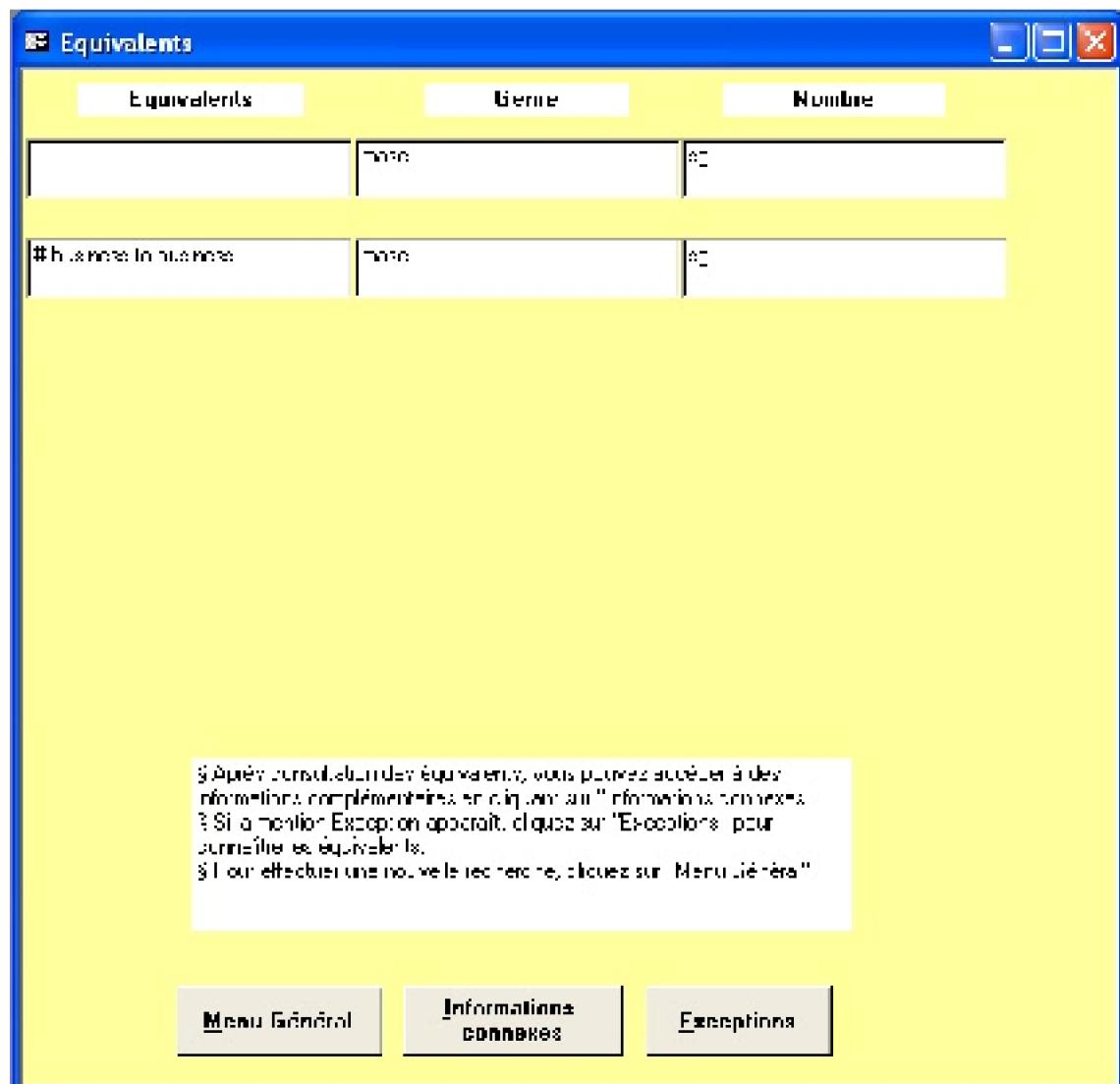


Figure 25 : Fenêtre d'affichage des équivalents

Après consultation des équivalents français de *Bulletin Board system* ou *Business-to-Business*, dans les exemples ci-dessus, cliquer sur **Informations connexes** donnera accès au menu du même nom qui propose cinq options grâce auxquelles l'utilisateur peut rechercher les éventuels :¹⁸⁸

- termes français implantés autres que ceux proposés par le GE (**Termes français implantés**),
- variantes du composé anglais saisi (**Variantes anglaises**),
- synonymes du composé anglais saisi (**Synonymes anglais**),
- sigles du composé anglais saisi (**Sigles anglais**),
- composés contentant le même déterminé ou déterminant (**Champs notionnels**) que

¹⁸⁸

Avant de découvrir les options de ce menu secondaire, il est préférable de finir la présentation du Menu général.

le composé anglais saisi.

Si les procédures de vérification orthographique (**Orthographe**) et d'identification de la vedette (**Composé-vedette**, **Composé-sigle**) n'ont pas permis à l'utilisateur de connaître les équivalents français alors la séquence saisie n'est pas un terme composé ou la mise à jour de la base de données n'a pas encore eu lieu. Dans ce dernier cas, retourner au Menu général et cliquer sur **Équivalences types** donnera accès aux structures françaises utilisées pour former les équivalents (en fonction du patron du terme anglais). Le GE ouvrira une boîte de dialogue dans laquelle il faudra saisir la structure compositionnelle anglaise du composé recherché¹⁸⁹ (Figure 26).



Figure 26 : Saisie de la structure compositionnelle

Après saisie et validation de la structure compositionnelle, N+N dans l'exemple, la fenêtre d'affichage des équivalences types présente les cinq structures françaises intervenant pour rendre les composés anglais N+N (Figure 27).

¹⁸⁹ Un guide apportant les connaissances nécessaires au découpage des composés sera élaboré.

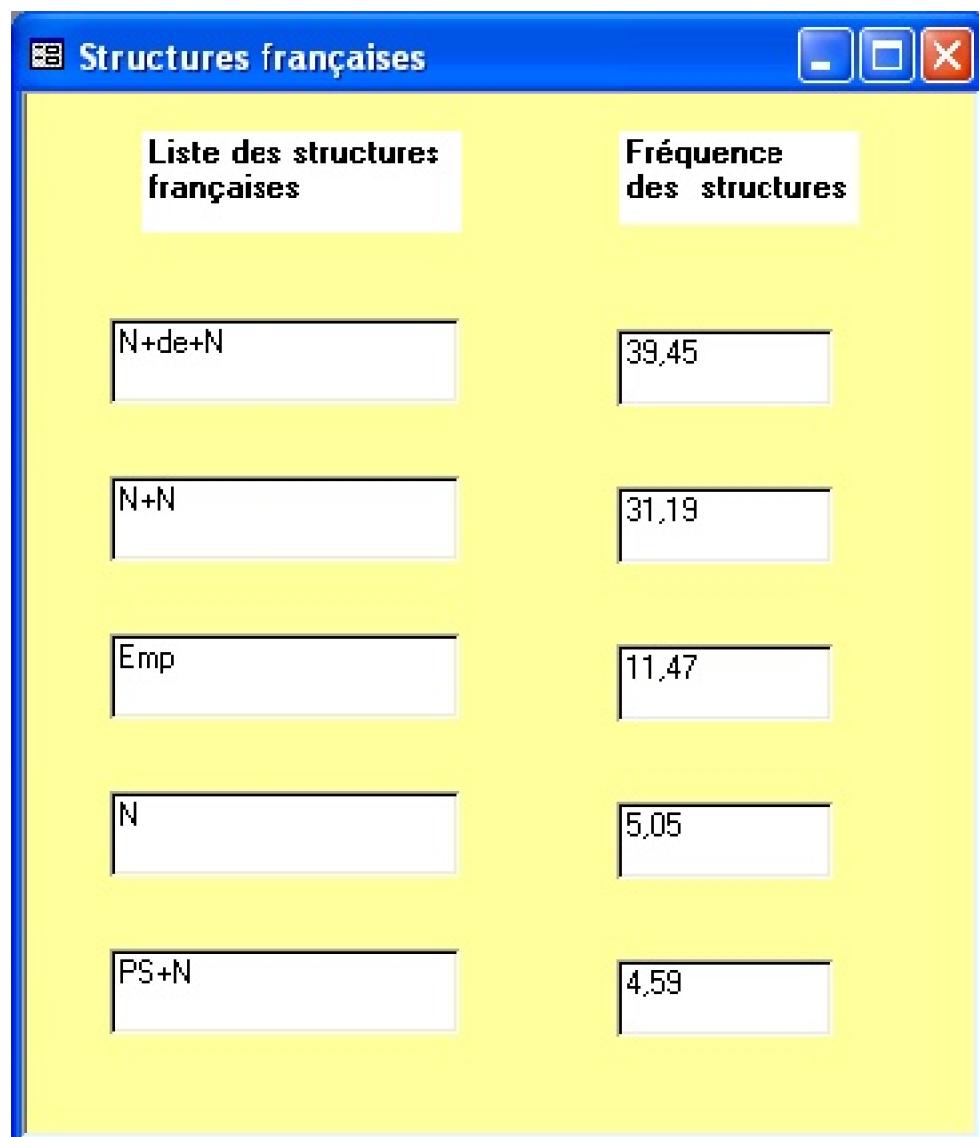


Figure 27 : Fenêtre d'affichage des équivalences types

L'utilisateur formera alors un terme selon la méthode préconisée (Part. III, Ch. 2, Sec. 1.2.1.).

Il se peut qu'après saisie d'une vedette anglaise dans la boîte de dialogue de recherche des équivalents, la mention Exception apparaisse dans la fenêtre des équivalents (Figure 28). C'est le cas pour *firewall*, par exemple.

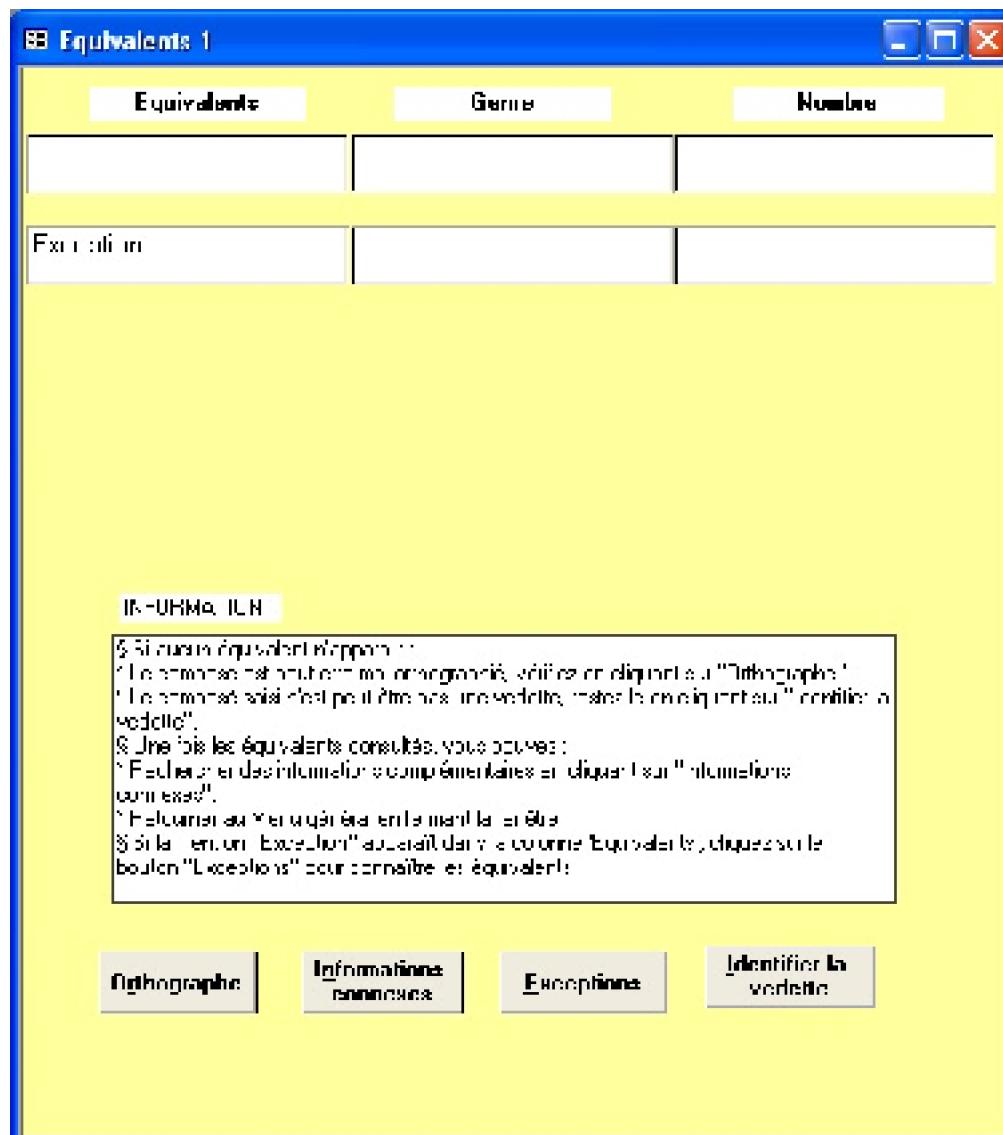


Figure 28 : Fenêtre d'affichage des équivalents français

Aucun équivalent n'est proposé par le système d'équivalences car la vedette *firewall* est un composé exocentrique. Suivre les instructions et cliquer sur **Exceptions**. Saisir *firewall* dans la boîte de dialogue (Figure 29).



Figure 29 : Boîte de dialogue

Après validation, deux traductions françaises, qui sont des emprunts internes à un

autre domaine, apparaissent : *coupe-feu* et *pare-feu*.

Equivalent	Genre	Nombre
coupe-feu	m	sg
pare-feu	m	sg
firewall		

Figure 30 : Fenêtre d'affichage des exceptions

Après obtention des équivalents français ou des équivalences types, l'utilisateur désirera peut-être savoir si des termes français autres que ceux proposés par le GE sont attestés, s'il existe une variante, un synonyme, un sigle pour la vedette anglaise recherchée ou si d'autres termes anglais ont le même déterminé ou déterminant. Ces cinq options sont offertes par le menu Informations connexes (Figure 31).

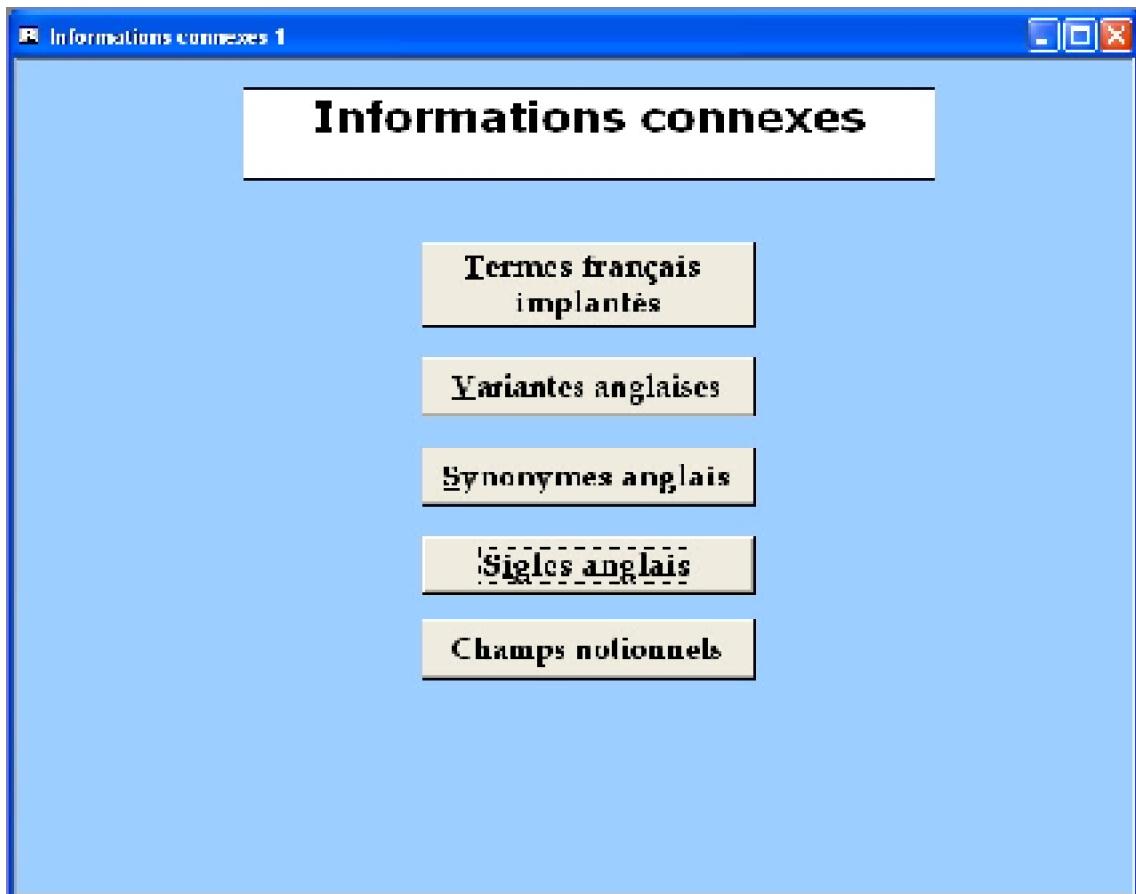


Figure 31 : Menu Informations connexes

Termes français implantés ouvre la boîte de dialogue de saisie du composé anglais pour lequel l'utilisateur veut connaître les éventuels termes français implantés (Figure 32). Prenons l'exemple de *log file*. Le GE propose cinq équivalents (*#fichier de connexion*, *#fichier journal*, *#log file*, *journal* et *fichier connectif*), dont trois correspondent à des termes attestés.



Les noms composés anglais français et espagnols du domaine d'Internet :

Figure 32 : Boîte de dialogue de saisie du composé

Après validation il apparaît que deux autres termes français sont attestés : *fichier de logs* et *journal des connexions* (Figure 33).



Figure 33 : Fenêtre d'affichage des termes français implantés

Pour connaître les variantes anglaises d'un composé, cliquer sur le bouton correspondant, toujours dans le menu Informations connexes, et saisir la vedette *cybercafé*, par exemple, dans la boîte de dialogue (Figure 34).



Figure 34 : Boîte de dialogue

La validation ouvre une fenêtre présentant une variante anglaise : *cyber-café* (Figure 35).



Figure 35 : Fenêtre d'affichage des variantes anglaises

Certains composés anglais peuvent avoir un synonyme. Cliquer sur **Synonyme anglais** et saisir la vedette anglaise *cybercafé* dans la boîte de dialogue (Figure 36).



Figure 36 : Boîte de dialogue

Suite à la validation, la fenêtre d'affichage des synonymes s'ouvre et Web bar apparaît (Figure 37).

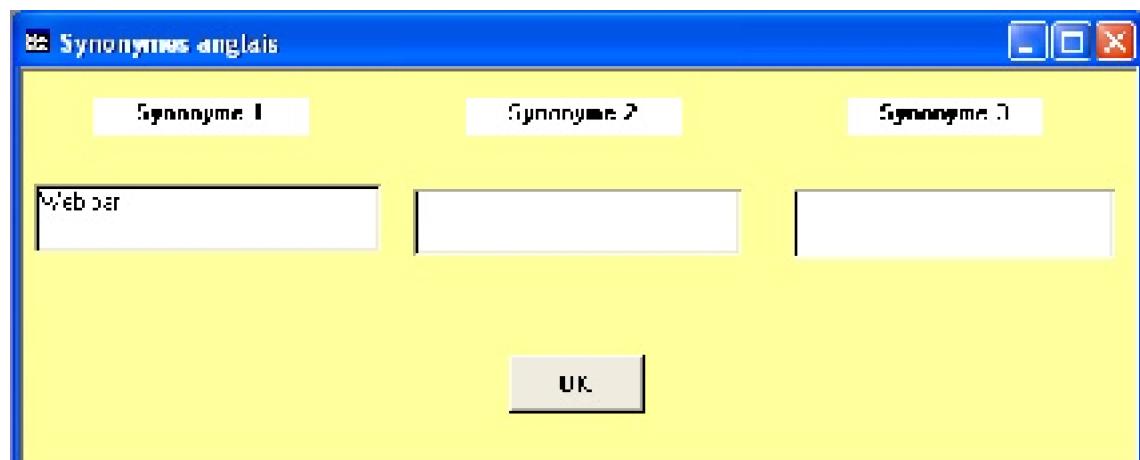


Figure 37 : Fenêtre d'affichage des synonymes anglais

Parfois un composé anglais peut être raccourci en sigle. Pour le découvrir, cliquer sur **Sigle anglais** et saisir une vedette anglaise, *peer-to-peer* par exemple, dans la boîte de dialogue (Figure 38).



Figure 38 : Boîte de dialogue

Il s'avère que deux sigles désignent ce concept : *P2P* et *PtoP* (Figure 39).



Figure 39 : Fenêtre d'affichage des sigles anglais

La dernière option offerte par le menu Informations connexes concerne la recherche des composés anglais formés à partir du même déterminé ou déterminant. Cliquer sur **Champs notionnels** et saisir *mail*, par exemple, dans la boîte de dialogue (Figure 40).



Figure 40 : Boîte de dialogue

Trente-neuf composés du corpus contiennent l'élément *mail* en position déterminé ou déterminant (Figure 41).

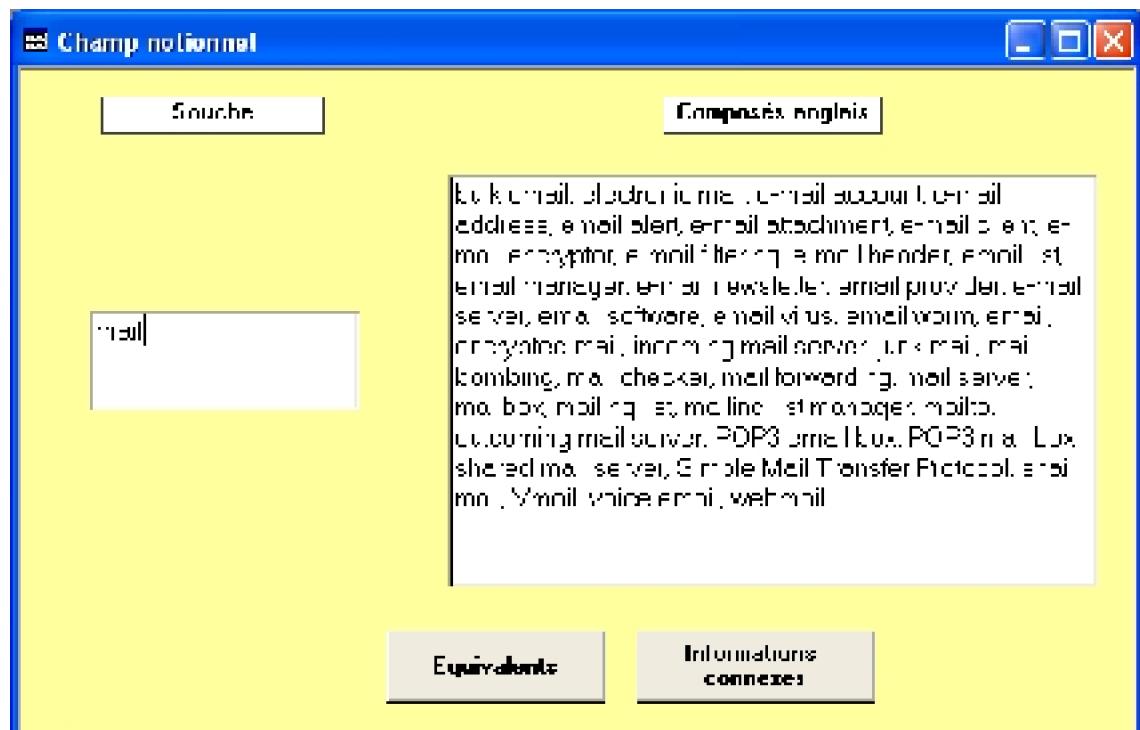


Figure 41 : Fenêtre d'affichage par champ notionnel

L'utilisateur peut rechercher les équivalents (**Équivalents**) ou retourner au menu secondaire (**Informations connexes**).

Comme le montre la présentation des différentes options, le GE est plus qu'une amélioration du système d'équivalences types. Non seulement la consultation des équivalents et des équivalences est aisée et rapide mais de nombreuses fonctions supplémentaires sont proposés : connaissance des éventuels synonymes, variantes et sigles anglais, termes français implantés et champs notionnels.

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Conclusion

En introduction, nous avons formulé deux questions relatives aux noms composés du lexique d'Internet :

- Quelles sont les particularités morphosyntaxiques et sémantiques des noms composés anglais, français et espagnols ?
- Quel est le devenir des noms composés anglais en français et en espagnol ?

Notre objet scientifique étant de fonder une méthode pour traduire les noms composés anglais à venir d'Internet en français et en espagnol – après examen des composés anglais attestés et de leurs équivalents de traduction –, nous nous inscrivons dans une perspective d'aide à la traduction et de création néologique. Plusieurs éléments ont suscité la volonté d'élaborer un système d'équivalences types anglais-français et anglais-espagnol pour assister le traducteur spécialisé : les caractéristiques des noms composés anglais, français et espagnols, les difficultés de traduction que posent les composés anglais d'Internet, le manque de fiabilité des outils de TA dans la traduction des composés du domaine et la confirmation de la possibilité de créer des équivalences entre les composés anglais et français et les composés anglais et espagnols.

Pour répondre à la première question, il fallait commencer par cerner le phénomène de la composition et du nom composé, c'est-à-dire savoir comment isoler les composés parmi toutes les unités lexicales et analyser les caractéristiques morphosyntaxiques et sémantiques des unités identifiées comme des noms composés. Nous avons étudié les définitions de toutes les unités lexicales de Tournier puis examiné des critères de

compositionnalité (Part. I). Aucun de ces critères n'est fiable à 100%. Pour recenser nos termes, nous avons donc eu recours à une définition peu restrictive, utilisé les critères réunissant un faisceau de présomptions et vérifié que les candidats-composés étaient répertoriés dans des ouvrages de référence et/ou qu'un moteur de recherche présentait cinq occurrences dans des contexte Internet différents. Cette démarche permet de tenir compte des catégories mineures. D'autre part, s'il est impossible de réunir toutes les particularités du nom composé ou de la séquence fortuite, alors les différences entre les deux sont extrêmement faibles. L'analyse des critères démontre que le type fondamental de construction est le même. Un composé a été une séquence libre avant d'être répertorié et sa fréquence peut diminuer au point qu'il peut perdre son statut de composé et redevenir une séquence libre. Par conséquent, les difficultés de traduction sont identiques.

La première partie s'est achevée avec l'étude de typologies des noms composés anglais, français et espagnols de la langue générale et de langues de spécialité. Nos buts étaient de disposer de modèles de construction où puiser les informations indispensables à la création de nos typologies et d'évaluer le degré d'affinage des catégories de notre classification. Le système d'équivalences types allait être basé sur un classement des composés anglais mais, les trois langues étant proches, étudier des classifications des composés français et espagnols élargissait notre choix.

Pour dresser le profil morphosyntaxique et sémantique des noms composés anglais, français et espagnols d'Internet nous avons réalisé divers classements (Part. II).¹⁹⁰ Pour recueillir les noms composés à étudier, le choix de la constitution d'un corpus s'est imposé. Pour garantir la pertinence d'un système d'équivalences s'adressant à des traducteurs et rédacteurs spécialisés, les termes observés devaient être attestés et réellement utilisés par les internautes. Les composés ont été collectés dans des revues d'Internet écrites par les internautes, véritables experts du domaine. Les magazines et les articles constituant le corpus ont été sélectionnés car leurs thèmes variés – navigation, moteur de recherche, sécurité, piratage, e-commerce, réseau, messagerie instantanée, courrier électronique, chat, etc. – permettent de regrouper l'essentiel des termes du domaine et de toucher un très vaste public.

Les termes extraits du corpus ont d'abord servi à la création d'une typologie morphosyntaxique qui allait constituer la toile de fond du système d'équivalences types. L'organisation des noms composés nous a confronté à des obstacles principalement liés à la réalisation du système d'équivalences. En effet, nous avions prévu de créer les équivalents en langue cible en fonction du patron syntaxique des composés sources, c'est-à-dire à partir d'une typologie des noms composés anglais existants mettant en parallèle leurs équivalents de traduction français et espagnols pour dégager les tendances de traduction. L'insertion d'une structure secondaire, basée sur la nature précise de chaque élément constitutif des composés (*N-ion*, *N-ment*, par exemple), pour affiner notre typologie a été écartée. Elle engendrait des complications pour la création des équivalents français et espagnols.

¹⁹⁰ Nous présenterons nos conclusions sur ce point uniquement après le rappel des principales particularités morphosyntaxiques et sémantiques.

L'analyse des noms composés a débuté par cette typologie morphosyntaxique. Au cours de nos observations, nous avons enregistré les caractéristiques suivantes :

- 82,10% des composés du corpus anglais sont formés d'après six structures compositionnelles (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. Quant aux composés espagnols, 87,85% d'entre eux ont été formés par 44 patrons endogènes. Avec 72,74% des termes, cinq structures se démarquent (N+de+N, N+A, N+N, PS+N et N+PS). Les composés exogènes comptent pour 12,15% du total.
- Le faible pourcentage de composés français et espagnols exogènes est aussi surprenant qu'intéressant étant donné la dépendance technologique des pays francophones et hispanophones en matière d'Internet. La brièveté généralement constatée des emprunts (or les unités considérées dans cette étude sont longues), le fait que le corpus exploité soit un corpus écrit excluant 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 ont été relevés (44,26% des binaires français et 46,04% des binaires espagnols). Ces termes hybrides semblent être une particularité du lexique d'Internet. Ils le différencient des autres terminologies plus soutenues.
- Des comparaisons ont révélé que les patrons compositionnels les plus fréquents dans le lexique d'Internet sont souvent les mêmes que ceux employés dans la langue générale et dans quelques terminologies spécialisées comme la physicochimie, l'informatique et la science fiction.
- La formation de composés à partir de pseudo-confixes semble constituer une grande nouveauté. Les premiers termes pseudo-confixés du domaine ont été créés en anglais mais la pseudo-confixation n'est pas une particularité du lexique anglais. Le corpus anglais compte sept pseudo-confixes (*cyber-*, *e-*, *web-*, *-ware*, *net-*, *hyper-* et *meta-*) et les corpus français et espagnols en comptent neuf (FR *cyber-*, *web-*, *e-*, *télé-*, *net-*, *-ware*, *méta-*, *hyper-* et *-nauta* ; ES *ciber-*, *web-*, *e-*, *-ware*, *hiper-*, *meta-*, *-nauta*, *tele-* et *-net*). L'hypothèse selon laquelle chacun des pseudo-confixes a un sens précis a été confirmée. Par exemple, *cyber-* entre dans la formation de composés désignant des activités, choses et personnes du monde réel reproduites sur la Toile alors que *meta-* intervient dans la création de termes désignant un type d'application. Les pseudo-confixes sont employés dans des champs notionnels précis. Leur fréquence très disparate reflète l'importance des champs notionnels.

Pour finir de dresser le profil des composés anglais, français et espagnols, une analyse de la centricité et des relations sémantiques est venue compléter utilement l'analyse morphosyntaxique. L'étude sémantique était également entreprise pour tenter d'expliquer, au moyen d'un croisement de typologies, les exceptions du système d'équivalences

types. Nos observations ont dévoilé les particularités suivantes des noms composés :

- Dans les trois langues, les composés endocentriques sont massivement représentés : 94,21% du corpus anglais, 93,68% du corpus français et 94,79% du corpus espagnol. Cette prépondérance s'explique par le fait que l'endocentricité facilite la compréhension du nom composé. Les combinaisons endocentriques présentent un degré de motivation (elles exposent dans leur signifiant un rapport non arbitraire avec leur signifié). De nombreux composés ont une double motivation morphologique et sémantique.
- En ce qui concerne les relations sémantiques, treize relations principales ont été identifiées en anglais, et quatorze en français et en espagnol. Pour les trois langues, les relations les plus fréquentes sont (ordre décroissant) : la fonction (N1 actif), le principe de fonctionnement, la nature et la fonction (N1 passif). En français et en espagnol, la relation de localisation est fortement représentée.

En raison de leurs particularités sémantiques (non compositionnalité due à la métaphore ou à la métonymie) et de leur origine (emprunts internes à la langue générale et à d'autres domaines), nous avions émis l'hypothèse que les composés exocentriques anglais entraîneraient le système d'équivalences types. En effet, les termes français et espagnols équivalents des termes anglais exocentriques ne sont que rarement des calques contrairement aux endocentriques. Un nouveau sens est ajouté, probablement sous influence de l'anglais, à une lexie existante ou un emprunt externe partiel ou total est adopté.

Cette analyse morphosyntaxique et sémantique des noms composés anglais, français et espagnols permet de tirer d'enrichissantes conclusions :

- Les lexiques français et espagnol sont dépendants du lexique anglais. L'influence de l'anglais est très marquée. Outre les emprunts, les composés français et espagnols sont massivement des calques des composés anglais. « Les calques sont une [...] forme de trace d'une langue dans une autre langue » (Arnaud, 2004 : 346). Ils passent inaperçus car ils respectent la formation des composés en langue cible mais ils augmentent peut-être le nombre de composés. Les langues cibles auraient pu utiliser d'autres matrices lexicogéniques pour dénommer ces concepts rendus par des composés. Les particularités des composés anglais se retrouvent dans les composés français et espagnols : a) les patrons compositionnels sont identiques, b) plus les composés sont longs moins ils sont nombreux, c) les pseudo-confixes anglais ont des équivalents français et espagnols, eux aussi fréquents, d) les proportions d'endocentricité et d'exocentricité sont similaires et e) les relations déterminant-déterminé sont les mêmes. Cette dépendance terminologique est sans doute le reflet d'une 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. Les internautes francophones et hispanophones semblent disposer d'une culture leur permettant de s'approprier la

majorité des composés. Cependant, seul un noyau d'internautes plus qualifiés (webmasters et programmateurs, par exemple) maîtrise réellement le domaine. En effet, notre corpus, constitué de revues grand public, contient peu de composés longs (trois éléments et plus) et d'emprunts qui sont la marque des termes techniques.

La conclusion la plus marquante est l'absence de frontière langue populaire-langue spécialisée dans le domaine d'Internet. Pour nommer une même réalité, le webmaster et l'internaute néophyte utilisent le même terme. Si ce lexique est loin des terminologies spécialisées généralement soutenues, c'est parce que ce domaine fait partie du quotidien des locuteurs francophones et hispanophones. 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, médecine, par exemple) est peut-être également liée à l'utilisation abondante d'Internet par le grand public.

La troisième partie répond à la deuxième question. Avant de mettre en place un outil d'aide à la traduction des noms composés, il fallait en déterminer la nécessité. Nous avons procédé à une pré-enquête qui a permis a) d'étudier les difficultés liées à la traduction des noms composés d'Internet (identification, compréhension des noms composés et mise en correspondance avec un équivalent de traduction de la langue cible), et à l'acquisition des connaissances nécessaires à leur traduction (pertinence des sources terminologiques), b) de déterminer le sens de circulation des termes du domaine et c) de vérifier la faisabilité des équivalences types grâce à un échantillon de composés.

Nous nous sommes ensuite employée à la conception des équivalences. Cinq étapes ont présidé à la mise en place des modèles d'équivalences morphosyntaxiques anglais-français et anglais-espagnol : a) les composés anglais ont été regroupés en fonction de leur patron compositionnel (la typologie morphosyntaxique de l'anglais joue pleinement son rôle ici) ; b) le (les) terme(s) français et espagnol(s) équivalent(s) de traduction de chaque composé anglais ont été identifiés ; c) une base de données mettant en parallèle les termes anglais et leurs équivalents français et espagnols et faisant ressortir la structure de ces derniers a été créée ; d) la fréquence « type » des modèles morphosyntaxiques français et espagnols a été calculée et e) les structures françaises et espagnoles ont été classées selon leur fréquence « type », et ce pour chaque modèle compositionnel anglais. Ces étapes ont permis de savoir si un patron compositionnel anglais est traduit en langue d'arrivée par un ou plusieurs patrons compositionnels, ou par d'autres matrices lexicogéniques, et de connaître leur fréquence. Seuls les composés endocentriques anglais ont été mis en correspondance avec leurs équivalents de traduction en langue cible. L'exclusion des trente composés anglais exocentriques devait éviter de fausser les statistiques. Des conseils d'utilisation du système d'équivalences ont ensuite été donnés. Lorsqu'on crée un terme, il faut veiller à lui conférer les qualités et les propriétés requises dans le domaine pour garantir son implantation : breveté, transparence, productivité, emprunt interne et hybridation.

Pour identifier les exceptions, nous avons appliqué notre système d'équivalences types aux composés anglais du corpus formés par le patron le plus fréquent ($N + N$). Notre outil s'est révélé être véritablement exploitable et efficace tout en étant fidèle au

mode de formation des termes du domaine, c'est-à-dire en créant par calque. En ce qui concerne les composés exocentriques, la création d'un équivalent français ou espagnol au cas par cas est préférable à l'application d'un système d'équivalences. En effet, le potentiel d'implantation est nul pour un certain nombre des équivalents de composés exocentriques. Hormis l'exocentricité, aucune autre propriété ne semble gêner la formation de termes par calque à l'aide du système. Lorsqu'un néologisme anglais sera un composé exocentrique totalement ou partiellement métaphorique, il devra être transposé (avec respect des structures préconisées). Un croisement des typologies morphosémantiques n'a pas montré que des relations N1-N2 mettaient en cause la traduction de tel ou tel patron compositionnel anglais.

Le système d'équivalences étant opérationnel, les atouts de son informatisation pour le traducteur spécialisé ont été évalués grâce à une brève observation de son environnement de travail et par un test de l'efficacité de logiciels de TA et TAO. Ce dernier a révélé la piètre qualité générale de ces outils de TA. L'automatisation du système d'équivalences répondait donc à un réel besoin. Le système d'équivalences est peu à peu devenu une base de données terminologique multilingue de création néologique et a été nommé Générateur d'équivalents (GE). Le GE a pris forme grâce à des schémas et des plans retraçant les projets d'organisation. Une ossature générale, une interface utilisateur, le détail de la fonction principale « Recherche des équivalents », l'utilisation d'un SGDB pour exploiter les équivalences et la création de la base de données ont été présentés. Le produit fini a été dévoilé et agrémenté de commentaires de captures d'écran exemplifiées pour expliquer la navigation entre les options offertes à l'utilisateur.

En élaborant un système d'équivalences types anglais-français et anglais-espagnol, nous sommes parvenue à concevoir un outil d'aide à la traduction des noms composés visant un public francophone et hispanophone de traducteurs et rédacteurs spécialisés dans le domaine d'Internet et nous avons gagné le pari de son informatisation. Le GE est un outil évolutif. L'évolutivité est le maître mot garantissant la pertinence et la longévité des outils terminologiques informatisés. Pour ne pas perdre en efficacité, les mises à jour, qui devront être effectuées au rythme d'apparition des concepts et des créations néologiques les dénommant, demanderont rapidité et facilité d'exécution.

Quel est l'apport de la présente étude ? Comment améliorer le GE ?

- Notre travail met en relief le profil morphosyntaxique et sémantique des noms composés anglais, français et espagnols, ce qui ouvre une porte pour l'élaboration de stratégies de traduction.
- La méthode de création des équivalences types et le GE pourront être étendus à d'autres lexiques spécialisés.
- Le principe novateur du GE pourra servir à l'amélioration d'outils terminologiques existants, voire être repris dans le développement de nouvelles applications.

Pour être plus performant, le GE devra être généralisé à toutes les unités lexicales et ne pas proposer uniquement des équivalents pour les noms composés anglais. Le traducteur disposera ainsi d'un outil complet couvrant tous les termes du domaine. Ensuite, il serait bénéfique d'élargir notre corpus aux revues underground et plus techniques. Le GE

regrouperait ainsi tous les termes dénommant des concepts utilisés quotidiennement par les internautes débutants, expérimentés, les programmeurs, etc. De plus, dans les revues techniques nous pourrions collecter d'autres surcomposés (notre corpus est trop peu riche en termes très longs), ce qui permettrait d'établir des statistiques sur un nombre raisonnable de termes longs et donc d'inclure les surcomposés dans le GE. Enfin, le GE devra être soumis à un public de traducteurs pour voir en quoi il répond à leurs attentes et leurs besoins et quelles fonctions méritent d'être modifiées, complétées, ou ajoutées.

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Bibliographie

* ouvrages traitant d'Internet

ADAMS Valérie, 1973, *An Introduction to Modern English Word-Formation*, Harlow, Longman, 230 p.

ADAMS Valérie 2001, *Complex Words in English*, Harlow, Harlow, 173 p.

AHRONIAN Céline, 2001, *Les anglicismes dans la langue espagnole, Etude de deux domaines : l'informatique et la génitique*, Mémoire de DEA, Université Lumière Lyon II, Dir. BÉJOINT Henri, 125 p.

ALCINA FRANCH Juan, BLECUA José Manuel, 1988 (1975), *Gramática española*, Barcelona, Ariel, 1274 p.

ALEXIEV Boyan, 2000, « Aspects of Terminological Metaphor », *Chuzhdoezikovo Obuchenie [Foreign Language Teaching Journal]*, vol. 3-4, pp. 25-31.

ALMELA PEREZ Ramón, 1999, *Procedimientos de formación de palabras en español*, Barcelona, Editorial Ariel, 253 p.

ALVAR EZQUERRA Manuel, POTTIER Bernard, 1983, *Morfología histórica del español*, Madrid, Gredos, 533 p.

ALVAR EZQUERRA Manuel, 1999, *La Formación de palabras en español*, Madrid, Arco Libros, coll. Cuadernos de la lengua española, 77 p.

*ANDRIEU Olivier, 1995, *Internet, guide de connexion*, Paris, Eyrolles, 239 p.

- ANSCOMBRE Jean-Claude, 1999, « Le Jeu de la prédication dans certains composés nominaux », *Langue française*, pp. 52-69.
- ARNAUD Pierre, 2003, *Les Composés timbre-poste*, Lyon, PUL, 168 p.
- ARNAUD Pierre, 2004, « Problématique du nom composé », in ARNAUD Pierre, dir., *Le Nom composé, Données sur 16 langues*, Lyon, PUL, pp 329-353.
- ARNAUD René, 1972, « Les Marques de fabrique, quelques réflexions », *La Banque des mots*, vol. IV, pp. 123-133.
- ARNTZ Reiner, 1993, « Terminological Equivalence and Translation », in SONNEVELD Helmi, LOENING Kurt, dir., *Terminology: Applications in Interdisciplinary Communication*, Amesterdam, John Benjamins, pp. 5-19.
- ATTAL Jean-Pierre, 1987, *Grammaire et usage de l'anglais*, Paris, Duculot, 991 p.
- AYTO John, 1996, « Lexical Life Expectancy, a Pronostic Guide », *Words*, pp. 181- 188
- *BALLE Francis, COHEN-TANUGI Laurent, 2001, *Dictionnaire du Web*, Paris, Dalloz, 322 p.
- BAUER Laurie, 1983, *English Word-Formation*, Cambridge, CUP, 311 p.
- BAUER Laurie, 1998, « When is a sequence of two nouns a compound in English? », *English Language and Linguistics*, 2-1, pp.65-86.
- BAUER Laurie, 2001, *Morphological Productivity*, Cambridge, CUP, 245 p.
- BAUER Laurie, HUDDLESTON Rodney, 2002, « Word-formation », in HUDDLESTON R., PULLUM Geoffrey K, dir., *The Cambridge Grammar of the English Language*, Cambridge, CUP, pp. 1644-1666.
- BÉCIRI Hélène, 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, Dir. CORTES, 577 p.
- BÉCIRI Hélène, 2000, « Néologie générale, néologie spécialisée : quelques observations dans le domaine de la micro-informatique », *La Banque des mots*, 60, pp. 19-36.
- BÉJOINT Henri, 1988, « Scientific and Technical Words in General Dictionaries », *International Journal of Lexicography*, 1-4, pp. 354-368.
- BÉJOINT Henri, THOIRON Philippe, 1989, « Pour un index évolutif et cumulatif de cooccurrences en langue techno-scientifique sectorielle », *Meta*, 34-4, pp. 661-671.
- BÉJOINT Henri, 1994, *Tradition and Innovation in Modern English Dictionaries*, Oxford, Oxford University Press, 276 p.
- BÉJOINT Henri, THOIRON Philippe, 1997, « Modèle relationnel, définition et dénomination », in BOISSON Claude et THOIRON Philippe dir., *Travaux CRTT 2 : Autour de la dénomination*, Lyon, PUL, pp. 187-203.
- BELOT Albert, 1987, *L'Espagnol d'aujourd'hui, Aspects de la créativité lexicale en espagnol contemporain*, Perpignan, Ed. du Castillet, 127 p.
- BENVENISTE Émile, 1974, *Problèmes de linguistique générale*, Paris, Gallimard, vol. 2, 286 p.
- BIDARD Josseline, 1999, *Le Nom et le groupe nominal*, Paris, Ellipses, 143 p.

- BLANCO Xavier, 1997, « Noms composés et traduction français-espagnol », *Lingvisticæ Investigationes*, 21-1, pp. 1-16.
- BLANCO Xavier, BONELL Carmen, 1998, « Vers une structuration syntactico-sémantique de la terminologie médicale, Application à la traduction espagnol-français », *Cahiers de Grammaire*, 23, pp. 1-11.
- BOISSON Claude, 1980, *L'Accentuation des composés en anglais contemporain, avec quelques contributions à l'accentologie générale*, 3 tomes, Université Paris VII, Lille, Atelier national de reproduction des thèses, 911 p.
- BOISSY J., 1988, « Tendances linguistiques de la néologie en terminologie », *La Banque des mots*, numéro spécial, pp. 77-83.
- BOSQUE Ignacio, DEMONTE Violeta, dir., 1999, *Gramática descriptiva de la lengua española*, vol. III, *Entre la oración y el discurso, morfología*, Madrid, Real Academia Española, RAE, pp. 3503-5351.
- BOULANGER Jean-Claude, 1984, « Quelques observations sur l'innovation lexicale spontanée et sur l'innovation lexicale planifiée », *La Banque des mots*, 27, pp 3-29.
- BOULANGER Jean-Claude, 1989, « Le Statut du syntagme dans les dictionnaires généraux monolingues », *Meta*, 34-3, pp. 360-369.
- BOULANGER Jean-Claude, 1989, « La Place du syntagme dans le dictionnaire de langue », *Meta*, 34- 3, pp. 516-526.
- BOURIGAULT Didier, JACQUEMIN Christian, 2000, « Construction de ressources terminologiques », in PIERREL Jean-Marie, *Ingénierie des langues*, Chapitre 9, Paris, Hermès, pp. 215-233.
- BOWKER Lynne, PEARSON Jennifer, 2002, *Working with Specialized Language, A practical guide to using corpora*, London and New York, Routledge, 241 p.
- BROCARD Valérie, 1998, « Étude lexicologique des termes techniques du cinéma américain », in TOURNIER Jean, dir., *Lexicologie anglaise et comparée*, Paris, Annales littéraires de l'Université de Franche Comté, Belles Lettres, pp. 15-30.
- BROLLES Yannick, 2001, *Terminologie française de l'Internet, Y a-t-il une alternative aux calques et aux emprunts ?*, Mémoire de DEA, Université Lumière Lyon II, Dir. BÉJOINT Henri, 107 p.
- BROWNLIE Siobhan, 2002, « La Traduction de la terminologie philosophique », *Meta*, 47-3, pp. 295-310
- BUREAU Claude, 1992, « Télécommunication et micro-informatique, les alliés du traducteur d'aujourd'hui », *Meta*, 37-4, pp. 595-609.
- BUSTOS GISBERT Eugenio (De), 1986, *La Composición nominal en español*, Salamanca, Ediciones Universidad de Salamanca, 469 p.
- CABRÉ et al., 2000, « Evaluación de la vitalidad de una lengua a través de la neología : A propósito de la neología espontánea y de la neología planificada », in CALLEBAUT Bruno, 1983, « Les Vocabulaires scientifiques et techniques dans les dictionnaires », *Cahiers de lexicologie*, 43, pp. 33-52.
- CANDEL Danielle, 2000, « Français scientifique et technique », in ANTOINE Gérald. et CERQUIGLINI Bernard, dir., *Histoire de la langue française (1945-2000)*, Paris, CNRS éd., pp. 339-369.

- CATCHAT Nina, 1981, *Orthographe et lexicographie, Les mots composés*, Paris, Nathan, 350 p.
- CHEVALIER Jean Claude, BLANCHE-BENVENISTE, ARRIVE, PEYTARD, 1997 (1964), *Grammaire du français contemporain*, Paris, Larousse, 495 p.
- CHUKWU Uzoma, 1996, « Optimisation du rapport coûts-avantages dans la réalisation de dictionnaires terminologiques informatisés (DTI) sur PC », *Meta*, 41-3, pp. 317-327.
- CLAS André, 1985, « Composés lourds et créations brachygraphiques terminologiques », *La Banque des mots*, 30, pp. 135-145.
- CLAS André, 1994, « Collocation et langues de spécialités », *Meta*, 39-4, pp. 576-580.
- CLAS André, 1998, « Quelques réflexions sur les matrices de néologie morpho-sémantique », in PAULIN Catherine, dir., *Recherche en linguistique étrangère, Mélanges Jean Tournier*, vol. 19, Paris, Belles Lettres, pp. 111-118.
- *Commission générale de terminologie et de néologie, 2001, « Terminologie de l'Internet », *La Banque des mots*, 62, pp. 73-83.
- Comisión de gramática, 1973, *Ezbozo de una nueva gramática de la lengua española*, Real Academia Española, dir., Madrid, Espasa Calpe, 592 p.
- *COLOMBAIN Jérôme, 1998, *Le Dico du multimédia*, Toulouse, Editions Milan, col. Les dicos essentiels, 255 p.
- *CONDAT Jean-Bernard, PIOCH Nicolas, 1994, *Internet*, Montréal, J.C. Inc, 224 p.
- CORBIN Danièle, TEMPLE Martine, 1994, « Le Monde des mots et des sens construits : catégories sémantiques, catégories référentielles », *Cahiers de lexicologie*, 65-2, pp. 5-28.
- CORMIER Monique C., RIOUX Louis-Paul, 1991, « Procédés de formation et matrices terminogéniques en terminologie des systèmes experts », *Meta*, 36-1, pp. 248-263.
- COSQUER Catherine, 1996, « Terminologie d'Internet », *La Banque des mots*, 51, pp. 89-115.
- CRUSE David Allan, 1986, *Lexical Semantics*, Cambridge, CUP, 310 p.
- DARMESTETER Arsène, 1967 (1872), *Traité de la formation des mots composés de la langue française comparée aux autres langues romanes et au latin*, Paris, Honoré Champion, 365 p.
- DEPECKER Loïc, 2000, « Quelques aspects des vocabulaires spécialisés de la fin du XX^e siècle : entre créativité lexicale et parallélisation sur l'anglais », *La Banque des mots*, 60, pp. 103-134.
- DEPIERRE Amélie, 2001, *La Composition en anglais scientifique : application au domaine de la physico-chimie*, Thèse de doctorat, Université du Littoral Côte d'Opale, Dir. GACHELIN, 344 p.
- DEROY Louis, 1971, « Néologie et néologismes : essai de typologie générale », *La Banque des mots*, Paris, 1, pp. 5-12.
- *DUFOUR Arnaud, 1995, *Internet*, Paris, PUF, 127 p.
- DUGAS André, 1997, « Les Variantes graphiques et leur source de variation », *La Mémoire des mots*, Actualités scientifiques, Actes du colloque de Tunis 25-26, 27

- septembre 1997, Montréal, Serviced, AUPELF-UREF, pp. 527-534.
- DURY Pascaline, 1997, *Étude comparative et diachronique de l'évolution de dix dénominations fondamentales du domaine de l'écologie en anglais et en français*, Thèse de doctorat, Université Lumière Lyon II, Dir. THOIRON Philippe, 425 p.
- EEK Elisabeth, 1998, « La Langue française de l'informatique envisagée depuis une perspective américaine », *Meta*, 43-3, pp. 240-247.
- ELUERD Roland, 2000, *La Lexicologie*, Paris, P.U.F., Coll. *Que sais-je ?*, PUF, 127 p.
- GAUDIN François, 1995, « Champs, clôtures et domaines : des langues de spécialités à la culture scientifique », *Meta*, 40-2, pp. 229-237.
- GAUDIN François, GUESPIN Louis, 2000, *Initiation à la lexicologie française, De la néologie aux dictionnaires*, Bruxelles, Editions Duculot, 355 p.
- GILE Daniel, 1995, *Basic Concepts and Models for Interpreter and Translator Training*, Amsterdam/Philadelphia, John Benjamins, 277 p.
- GINDRE Philippe, 1998a, « Le Lexique de la science-fiction », in TOURNIER Jean, dir., *Lexicologie anglaise et comparée*, Paris, Annales littéraires de l'Université de Franche Comté, Belles Lettres, pp.31-58.
- GINDRE Philippe, 1998b, « L'intuition lexicale au secours de la lexicogénèse », in PAULIN Catherine, dir., *Recherche en linguistique étrangère, Mélange Jean Tournier*, vol. 19, Paris, Belles Lettres, pp. 225-235.
- GINDRE Philippe, 1999, *La formation des néologismes de la littérature de Science-Fiction d'expression anglaise contemporaine*, Thèse de doctorat, Université de Franche Comté, Dir. TOURNIER Jean, 630 p.
- GOUADEC Daniel, 1997, « Dictionnaires terminologiques, l'impact des nouvelles technologies », *Meta*, 42-1, pp. 24-32.
- GOUADEC Daniel, 2002, *Profession : traducteur*, Paris, La Maison du Dictionnaire, 432 p.
- GREVISSE, 2001 (1936), *Le Bon usage*, Paris, Duculot, 1762 p
- GROSS Gaston, 1986, *Typologie des noms composés*, Villetaneuse, Rapport A.T.P. « Nouvelles recherches sur le langage », Université Paris XIII, 150 p.
- GROSS Gaston, 1988, « Degré de figement des noms composés », *Langages*, 90, pp. 57-72.
- GROSS Gaston, 1990, « Définition des noms composés dans un lexique-grammaire », *Langue française*, 87, pp. 84-90.
- GROSS Gaston, 1996, *Les Expressions figées en français, noms composés et autres locutions*, Paris, Ophrys, 161 p.
- GRUNIG Blanche, 1967, « Contribution à l'étude des composés et dérivés », *Words*, 23-1, 2, 3, pp. 239-253.
- GUERRERO RAMOS, Gloria, 1997, *Neologismos en el español actual*, Madrid, Arco Libros, coll. *Cuadernos de lengua española*, 52 p.
- GUILBERT Louis, 1971a, « La Néologie scientifique et technique », *La Banque des mots*, Paris, 1, pp. 45-54.
- GUILBERT Louis, 1971b, « De la Formation des unités lexicales », in GUILBERT Louis,

- LAGANE René, NIOBEY Georges, dir., *Grand Larousse de la langue française*, vol. I, Paris, Larousse, pp. IX-LXXX.
- GUILBERT Louis, 1973, « La Spécificité du terme scientifique et technique », *Langue française*, 17, pp. 5-17.
- GUILBERT Louis, 1975, *La Créativité lexicale*, Paris, Larousse, 285 p.
- *GUSDORF Florent, WISDOM John, 1998, *Guide anglais-français du Cybermonde*, Paris, Ellipses, 287 p.
- HEID Ulrich, FREIBOTT Gerhard, 1991, « Collocations dans une base de données terminologique et lexicale », *Meta*, 36-1, pp.77-91.
- HELIOS Jaime, 1997, *Linguistique espagnole*, Paris, col. Les Essentiels-Espagnol, Ellipses, 128 p.
- HERMANS Adrien, VANSTEELANDT Andrée, (1999), « Néologie traductive », Bruxelles, *Terminologies nouvelles*, 20, pp. 37-42.
- HUART Ruth, 1989, « Nouveau regard sur les composés » in GAUTHIER André, dir., *Exploration en linguistique anglaise*, Bern, Peter Lang, pp. 25-90.
- HUMBLEY John, 1987, « L'Emprunt sémantique dans la terminologie de l'informatique », *Meta*, 32-2, pp. 46-70.
- HUMBLEY John, 1988, « Comment le français et l'allemand aménagent la terminologie de l'informatique », *La Banque des mots*, numéro spécial, pp. 85-148.
- HUMBLEY John, 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 John, 1998, « À Propos de l'instabilité désignationnelle en néologie », PAULIN Catherine, dir., *Recherche en linguistique étrangère, Mélanges Jean Tournier*, vol. 19, Paris, Belles Lettres, pp. 267-276.
- HUMBLEY John, 2000, « Évolution du lexique », in ANTOINE Gérald et CERQUIGLINI Bernard, dir., *Histoire de la langue française (1945-2000)*, Paris, CNRS éd., pp. 71-106.
- JASTRAB DE SAINT-ROBERT Marie-Josée, 1987, « Les Syntagmes nominaux complexes en anglais et en français : éléments de réflexion », *Meta*, 32-3, pp. 260-266.
- JESPERSEN Otto, 1965, *A Modern English Grammar on Historical Principles*, part VI Morphology, London, George Allen & Unwin Ltd, 570 p.
- JOOS Martin, 1967 (1961), *The Five Clocks*, New York, Harcourt, Brace and World, 108 p.
- JOUNEAU Isabelle, 1997, *La Composition nominale par juxtaposition en anglais contemporain*, Thèse de doctorat, Université de Franche-Comté, 219 p.
- JOUNEAU Isabelle, 1998, « La Métasémie dans la composition nominale par juxtaposition en anglais contemporain », in TOURNIER Jean, dir., *Lexicologie anglaise et comparée*, Annales littéraires de l'Université de Franche Comté n°652, Belles Lettres, pp. 59-71.
- JOSSELIN Amélie, 2001, *L'Emprunt lexical en France et au Canada : le cas particulier des anglicismes et des gallicismes et leur traitement lexicographique*, mémoire de

- DEA, Université Lumière Lyon II, 2 vol.
- JUNG René, 1990, « Remarques sur la constitution du lexique des noms composés », *Langue française*, 87, pp. 91-97.
- KING Margaret, 1992, « L'Evaluation des systèmes de traduction automatique dans le cadre d'un service de traduction », *Meta*, 37-4, pp.817-727.
- KOCOUREK Rotislav, 1991a, « Textes et termes », *Meta*, 36-1, pp. 71-76.
- KOCOUREK Rotislav, 1991b, *La Langue française de la technique et de la science*, Wiesbaden, Brandstetter Verlag, 2^e éd. (1982), 327 p.
- LAKOFF George, JOHNSON Mark, 1980, *Metaphors We Live By*, Chicago, University of Chicago Press.
- *LE DIBERDER Alain, 2000, *Histoire d'@, L'abécédaire du cyber*, Paris, Editions La Découverte, 175 p.
- LEHMANN Alise, MARTIN-BERTHET Françoise, 1998, *Introduction à la lexicologie, Sémantique et morphologie*, Paris, Dunod, 201 p.
- LEON Monique, 2000 (1997), *La Prononciation du français*, Paris, Nathan, 127 p.
- LERAT Pierre, 1994, « Composé syntagmatique, dénomination, terminologie », *Cahiers de lexicologie*, 65-2, pp. 151-158.
- LERAT Pierre, 1995, *Les Langues spécialisées*, Paris, PUF linguistique nouvelle, 201 p.
- MANIEZ François, 2002, « Distinguer les termes des collocations : étude sur corpus du patron <adjectif-nom> en anglais médical », *Actes du colloque de TALN de Nancy*, 24-27 juin 2002, pp. 345-350.
- MANIEZ François, 2004, « Extraction d'équivalents de traduction à partir d'un corpus bilingue aligné : étude du patron syntaxique Adj N en anglais médical », *Actes des 2èmes journées de la Linguistique de Corpus*, septembre 2002.
- MATHIEU-COLAS Michel, 1994, *Les Mots à traits d'union, Problèmes de lexicographie informatique*, Paris, Didier Eruditio, 351 p.
- MATHIEU-COLAS Michel, 1996, « Essai de typologie des noms composés français », *Cahiers de lexicologie*, 69-2, pp. 71-125.
- MARCHAND Hans, 1967, « On the Description of Compounds », *Words*, 23-1,2,3, pp.379-387.
- MARCHAND Hans, 1969 (1960), *The Categories and Types of Present-Day English Word-Formation*, München, C.H. Beck'sche Verlagsbuchhandlung, 545 p.
- MARCHELLO-NIZIA Christiane, 1995, *L'évolution du français, Ordre des mots, démonstratifs, accent tonique*, Paris, Armand Colin, 213 p.
- MARTINET André, 1967, *Eléments de linguistique générale*, Paris, Armand Colin, 221 p.
- MELKA Francine, 2002, « La Création morphologique dans le dictionnaire », in MELKA Francine, AUGUSTO M. Celeste, dir., *De la Lexicologie à la lexicographie*, Utrecht, pp. 100-115.
- Merriam Webster's Collegiate, 10th edition*, 2001 (1993), Springfield, Massachussets, 1555 p.
- MERILLOU-BOUFFAULT Catherine, 1997, *Anglicismes et traduction*, Thèse de

- doctorat en science du langage, Villeneuve d'Ascq, Presses Universitaires du Septentrion, 351 p.
- MEYER Ingrid, ZALUSKI Victoria, MACKINTOSH Kristen, FOZ Clara, 1997, « La Terminologie métaphorique d'Internet : aspects notionnels, structurels et traductionnels », *Actes des V^e journées scientifiques AUPELF/UREF*, pp. 1-14.
- MEYER Ingrid, ZALUSKI Victoria, MACKINTOSH Kristen, 1998, « Metaphorical Internet Terms : A Conceptual and Structural analysis », *Terminology*, vol. 4-1, pp. 1-34.
- MITTERAND Henri, 2000, *Les Mots français*, Paris, PUF, coll. *Que sais-je ?*, 127 p.
- MOLLER Bernt, 1994, « Où va la terminologie technique française ? », *La Banque des mots*, 6, pp. 69-98.
- MOPOHO Raymond, 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, AUPELF-UREF, pp. 551-561.
- NAKOS Dorothy, 1989, « Etude comparée des modes de formation des lexies complexes dans deux domaines différents », *Meta*, 34-3, pp. 352-359.
- *NEWTON Harry, 2000, *Le Newton Nouveau Dictionnaire des télécommunications de l'informatique et de l'Internet*, Paris, Osman Eyrolles Multimédia, 846 p. Trad. par Zina Charles.
- NICOLAS Christian, 1994, « Le Procédé du calque sémantique », Paris, *Cahiers de Lexicologie*, 65-2, pp. 75-101.
- NYMANSSON Karin, 1995, « Le Genre grammatical des anglicismes contemporains en français », *Cahiers de lexicologie*, 66-1, pp. 95-113.
- OTMAN Gabriel, 2000, « Le Vocabulaire de l'informatique et des technologies de l'information et de la communication », in ANTOINE Gérald et CERQUIGLINI Bernard, dir., *Histoire de la langue française (1945-2000)*, Paris, CNRS éd., pp. 371-396.
- PAILLARD Michel, 2000, *Lexicologie contrastive anglais-français, formation des mots et construction du sens*, Paris, Ophrys, 199 p.
- *PELIKS Gérard, 1995, *Le World-Wide Web*, Paris, Editions Addison-Wesley France, 226 p.
- PERCEBOIS Jacqueline, 2001, « Fonctions et vie des sigles et acronymes en contextes de langues anglaise et française de spécialité », *Meta*, 46-4, pp. 627-645.
- PERGNIER Maurice, 1989, *Les Anglicismes*, Paris, PUF, 215 p.
- PHAL André, 1969, « La Recherche au Crédif : la part du lexique commun dans les vocabulaires scientifiques et techniques », *Langue Française*, 2, pp. 73-81.
- PHAL André, 1964, « Les Groupes de mots et les problèmes qu'ils posent dans la préédition de textes scientifiques destinés à l'analyse mécanographique », *Cahiers de lexicologie*, Paris, Larousse, 4, pp. 45-60.
- PICOCHE Jacqueline, 1986, *Structures sémantiques du lexique français*, Paris, Nathan, 143 p.
- PLUVINET Patricia, 2001, « Vocabulaire de l'escalade, pour en finir avec l'escalade à

- mains nues », *La Banque des mots*, 62, pp. 13-33.
- PORTELANCE Christine, 1986, « A Propos du rôle des matrices terminologiques », *Terminogramme*, 39-40.
- PORTELANCE Christine, 1987, « Fertilisation terminologique ou insémination terminologique artificielle », *Meta*, 32-3, pp. 356-360.
- POTTIER Bernard, DARBORD Bernard, CHARAUDEAU Patrick, 1994, *Grammaire explicative de l'espagnol*, Paris, Nathan Université, 318 p.
- PRATT Chris, 1980, *El Anglicismo en el español peninsular contemporáneo*, Madrid, Gredos.
- QUIRK Randolph et al., 1972, *A Grammar of Contemporary English*, London, Longman, 1120 p.
- REY Alain, 1979, *La Terminologie, noms et notions*, Paris, PUF, coll. *Que sais-je ?*, 127 p.
- REY Alain, 1980, *La Lexicologie*, Paris, PUF, coll. *Que sais-je ?*, 127 p.
- RAUS Rachele, 2001, « Productivité de *cyber* et *hyper* dans le lexique français d'Internet », *La Linguistique*, 37-2, pp. 71-86.
- RODRIGUEZ Nadia, SCNELL Bettina, SCMHIT Christine, LECOT Vanesa, « Terminologie de la sécurisation des paiements par internet », *La Banque des mots*, pp. 24-30.
- RONDEAU Guy, 1980, « Terminologie et documentation », *Meta*, 25-1, pp. 152-170.
- ROBERTS Roda P., PERGNIER Maurice, 1987, « L'Equivalence en traduction », *Meta*, 32-4, pp. 392-1987.
- ROUDAUD Brigitte, 1992, « La Traduction automatique : l'ordinateur au service des traducteurs », *Meta*, 37-4, pp. 828-845.
- ROYAUTE Jean, 1999, *Les Groupes nominaux complexes et leurs propriétés : application à l'analyse de l'information*, Thèse de doctorat en informatique, Université de Nancy I, 228 p.
- SABLAYROLLES Jean-François, 2000, *La Néologie en français contemporain, Examen du concept et analyse de productions néologiques récentes*, Paris, Honoré Champion, 588 p.
- *SAINT-LEGER Olivier, PIGOT Thierry, 1995, *Clés de contact pour Internet*, Paris, Sybex, 298 p.
- SCHAETZEN C. de, 1993, « Diachronie et libéralisme en terminologie », *La Banque des mots*, 45, pp. 33-39.
- SCHAPIRA Charlotte, 1987, « Comment rendre en français les termes anglais dérivés et composés à la fois ? », *Meta*, 32-3, pp. 342-346.
- SILBERZTEIN Max, 1990, « Le Dictionnaire électronique des mots composés », *Langue française*, Paris, Larousse, 87, pp. 71-83.
- SILBERZTEIN Max, 1993, « Les Groupes nominaux productifs et les noms composés lexicalisés », *Lingvisticæ Inverstigationes*, Amsterdam, John Benjamins, 27-2, pp. 405-425.
- SINCLAIR John, 1991, *Corpus Concordance Collocation*, Oxford, OUP, 179 p.

- SLODZIAN Monique, 1995, « Transfert de connaissances nouvelles et aménagement terminologique », *Meta*, 40-2, pp. 238-243.
- SOUBRIER Jean, 1997, « Néologismes et termes d'emprunt dans le français des affaires, Une étude de la terminologie en usage dans les offres d'emploi de la presse française », *La Mémoire des mots, Actualités scientifiques, Actes du colloque de Tunis 25-26, 27 septembre 1997*, Montréal, Serviced, AUPELF-UREF, pp. 403-413
- SOURNIA Jean-Charles, 1994, « Les Phases évolutives du vocabulaire médical français », *Meta*, 39-4, pp. 692-700.
- SZABO Levente, 1996, *La Composition nominale, Etude comparative sur le français, l'anglais et le hongrois*, Mémoire de DEA, Université Lumière Lyon II, Dir. ARNAUD Pierre, 49 p.
- TETET Christiane, 1994, « La Terminologie de l'alpinisme dans les dictionnaires », *Meta*, 39-4, pp. 651-661.
- THIELE Johannes, 1987, *La Formation des mots en français moderne*, Montréal, Presses Universitaires de Montréal, Trad. André CLAS, 180 p.
- THOIRON Philippe, MANIEZ Dominique, 1989, « Recherche d'utilisation optimale d'un SGBD en traduction avec aides informatiques : le dictionnaire informatisé bilingue d'écologie des eaux continentales (DIBEC) », *Meta*, 34-3, pp. 509-515.
- THOIRON Philippe, BÉJOINT Henri, 1991, « La Place des reformulations dans les textes scientifiques », *Meta*, 36-1, pp. 101-110.
- THOIRON Philippe, 1994, « La Terminologie multilingue : une aide à la maîtrise des concepts », *Meta*, 34-4, pp. 765-773.
- THOIRON Philippe, ARNAUD Pierre, BEJOINT Henri, BOISSON Claude, 1996, « Notion d'archi-concept et dénomination », *Meta*, 41-4, pp. 512-524.
- THOIRON Philippe, 2000, « La Traduction des termes scientifiques : jeu entre concept et termes », in VERINE Henri, dir., *Les Sciences et leur langages*, Paris, Edition du CTHS, pp. 329-339.
- TOURNIER Jean, 1985, *Introduction à la lexicogénétique de l'anglais contemporain*, Genève, Champion-Slatkine, 517 p.
- TOURNIER Jean, 1991a, *Précis de lexicologie anglaise*, Paris, Nathan, 208 p.
- TOURNIER Jean, 1991b, *Structures lexicales de l'anglais*, Paris, Nathan, 190 p.
- TRIMBLE Louis, 1989, *English For Science and Technology*, Cambridge, CUP, 180 p.
- VAL ALVARO José Francisco, 1999, « La composición », in Real Academia española, dir., *Gramática descriptiva de la lengua española*, pp. 4757-4841.
- VERINE Henri dir., 2000, *Les Sciences et leurs langages*, Paris, Editions du CTHS, Congrès national des sociétés historiques et scientifiques, 129^e Amiens 120^e Aix-en-Provence, 362 p.
- *VIRGA, 1997, *Dictionnaire bilingue de la micro-informatique et de l'Internet*, Alleur, Marabout, 566 p.
- VIVES Robert, 1990, « Les Composés nominaux par juxtaposition », *Langue française*, 87, pp. 98-103.
- *VOSS Andreas, 1998, *Dictionnaire de l'informatique et de l'Internet 1999*, Paris, Micro

Application, 731 p.

WALTER Henriette, 1997, « Le Lexique de l'informatique et l'emprise de l'anglais », *La Linguistique*, 33-2, pp. 45-59.

WATBLED Jean-Philippe, 1998, *Les Mots de l'anglais*, Paris, Nathan, coll. 128, 128 p.

WILLIAMS Geoffrey C., 1999, *Les Réseaux collocationnels dans la construction et l'exploitation d'un corpus dans le cadre d'une communauté de discours scientifique*, Thèse de doctorat en anglais, dir. BOUCHER Paul, Université de Nantes, Villeneuve d'Ascq, Septentrion, 330 p.

ZANDVOORT R.W., 1962 (1st éd. 1957), *Handbook of English Grammar*, London, Longmans, 349 p.

SITES INTERNET

Glossaires Internet :

ATI : www.ati.es

Dico du Net : www.outilnet.com/dico

DGLF: www.culture.fr/culture/dglf/cogeter/1-9-00-internet.htm

Glossaire de termes relatifs à Internet (BENZINEB Jean-Karim) :
www.culture.fr/culture/dglf/ressources/lexiques/internet.htm

Le Grand dictionnaire terminologique : www.granddictionnaire.com

McAfee : <http://fr.mcafee.com/virusInfo/default.asp?id=glossary>

Netlingo : <http://www.netlingo.com/inframes.cfm>

Dalloz Internet : www.dicoweb.dalloz.fr

Ressources terminologiques :

CRTT : <http://recherche.univ-lyon2.fr>

Grand dictionnaire terminologique : www.granddictionnaire.com

Meta : www.erudit.org/erudit/meta/

RINT : www.cfwb.be/franca/termin/liste.htm

Dictionnaires :

Cambridge Advanced Learner's Dictionary : www.dictionary.cambridge.org

Les noms composés anglais français et espagnols du domaine d'Internet :

Merriam Webster Online Dictionary : www.m-w.com

DRAE : www.rae.es

Index terminologique

A

- accentuation, 24, 25, 26, 27, 28, 29, 41
- accent, 21, 25, 26, 27, 28, 29, 31, 44, 49, 51, 79, 298, 322
- amalgame, 1, 10, 12, 13, 15, 17, 22, 23, 30, 69, 71, 74, 77, 97, 98, 108, 110, 112, 113, 134, 156, 205, 228, 229, 230
- attesté, 18, 39, 44, 70, 92, 114, 202, 223, 228, 229, 230, 231, 253, 266, 279, 280, 285, 286

B

- brièveté, 88, 117, 191, 287, 288, 289

C

- calque, 45, 46, 70, 106, 118, 120, 121, 122, 125, 126, 132, 135, 137, 138, 139, 140, 141, 147, 148, 152, 160, 167, 168, 174, 180, 200, 204, 228, 229, 230, 231, 232, 288, 289, 293, 299
- calqué, 71, 78, 121, 152, 174, 234
- centricité, 13, 68, 103, 106, 144, 177, 287
- endocentricité, 104, 105, 145, 287, 288
- endocentrique, 13, 23, 28, 37, 38, 39, 67, 68, 91, 93, 103, 104, 105, 106, 133, 134, 137, 139, 145, 169, 171, 174, 177, 178, 229, 232, 234, 287, 288, 289, 358, 360, 361, 364, 365, 367, 368
- exocentricité, 9, 35, 108, 229, 288, 289
- exocentrique, 13, 23, 28, 34, 38, 39, 67, 68, 104, 105, 106, 108, 109, 137, 145, 146, 147, 178, 180, 221, 223, 228, 229, 230, 231, 232, 233, 234, 235, 262, 278, 288, 289, 361, 364, 368, 380
- collocation, 7, 18, 19, 20, 21, 23, 24, 32, 33, 34, 35, 36, 37, 38, 39, 298
- combinatoire intermatricielle, 5, 109, 110, 181
- composition, 3, 4, 7, 8, 9, 10, 13, 14, 18, 22, 23, 32, 39, 44, 46, 50, 51, 55, 56, 69, 72, 83, 86, 89, 97, 99, 101, 105, 109, 110, 111, 112, 113, 114, 115, 117, 122, 138, 140, 145, 151, 152, 165, 168, 169, 175, 182, 183, 285, 297
- conversion, 10, 45, 46, 110
- corpus, 5, 19, 23, 26, 28, 29, 30, 41, 44, 51, 59, 60, 61, 64, 65, 66, 67, 69, 71, 79, 80, 81, 82, 84, 85, 86, 88, 89, 90, 91, 92, 93, 94, 95, 96, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 109, 110, 112, 113, 114, 115, 119, 120, 121, 122, 123, 125, 126, 127, 130, 132, 133, 135, 136, 137, 139, 141, 142, 143, 144, 145, 147, 148, 149, 150, 151, 152, 154, 155, 156, 157, 158, 159, 160, 161, 162, 164, 165, 167, 168, 169, 170, 171, 172, 173, 174, 176, 178, 179, 181, 182, 183, 195, 202, 207, 208, 221, 223, 236, 262, 283, 286, 287, 288, 289, 290, 298, 303, 311, 314, 315, 330, 343, 358

D

- dérivation inverse, 10, 110
- déterminant, 11, 12, 13, 17, 18, 20, 22, 33, 34, 35, 36, 37, 41, 43, 45, 54, 68, 75, 91, 96, 98, 99, 104, 105, 106, 108, 109, 132, 133, 134, 135, 137, 138, 139, 140, 146, 147, 148, 157, 166, 169, 171, 172, 173, 174, 175, 176, 177, 178, 179, 193, 228, 235, 239,

276, 279, 283, 288

- déterminé, 11, 13, 17, 20, 22, 26, 33, 34, 35, 51, 52, 54, 68, 75, 91, 93, 98, 99, 101, 103, 104, 105, 106, 108, 109, 132, 133, 135, 137, 139, 140, 146, 147, 148, 157, 166, 169, 170, 171, 172, 173, 174, 176, 177, 178, 179, 191, 234, 235, 239, 244, 276, 279, 283, 288
- distribution, 14, 18, 24, 31, 41, 108, 224, 242, 331, 336, 343, 347, 358, 361

E

- emprunt, 10, 13, 45, 46, 66, 70, 74, 75, 76, 77, 81, 106, 110, 115, 117, 118, 120, 121, 123, 124, 125, 127, 130, 136, 137, 138, 141, 142, 144, 145, 146, 148, 152, 153, 155, 156, 159, 162, 166, 170, 171, 172, 173, 174, 176, 177, 178, 180, 183, 188, 192, 200, 202, 203, 204, 228, 229, 230, 231, 234, 262, 279, 287, 288, 293, 301, 377, 378, 380
- emprunté, 45, 46, 70, 71, 75, 106, 120, 121, 131, 132, 138, 145, 147, 152, 172, 176, 196, 228, 232, 234
- endogène, 73, 74, 75, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 130, 155, 156, 158, 159, 160, 161, 162, 163, 164, 262, 286, 287, 347, 351, 352, 353, 356, 357
- équivalence type, i, 4, 5, 43, 46, 68, 109, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 228, 229, 231, 232, 234, 256, 262, 272, 276, 277, 279, 289, 290, 311, 312, 381
- équivalent de traduction, 3, 4, 200, 206, 285, 289
- exogène, 73, 74, 75, 118, 119, 120, 123, 124, 126, 127, 129, 130, 147, 155, 157, 158, 159, 160, 161, 164, 180, 286, 287, 350, 352, 355, 357

F

- fréquence, 24, 27, 39, 40, 41, 65, 67, 69, 77, 79, 80, 82, 83, 88, 89, 92, 97, 102, 111, 119, 122, 123, 124, 126, 130, 131, 132, 133, 134, 136, 141, 142, 143, 148, 154, 157, 160, 162, 164, 166, 167, 169, 170, 172, 176, 177, 183, 187, 195, 205, 207, 208, 209, 228, 229, 233, 258, 259, 261, 262, 267, 285, 287, 288, 289, 311, 380, 381
- fréquence « type », 69, 79, 80, 82, 119, 123, 132, 133, 134, 136, 141, 142, 148, 154, 167, 170, 172, 176, 177, 207, 208, 289, 311

G

- Générateur d'équivalents, i, 247, 253, 289
- GE, 247, 250, 253, 254, 256, 257, 258, 259, 261, 262, 263, 266, 270, 271, 273, 275, 276, 279, 280, 284, 289, 290, 371, 376, 380, 381
- graphie, 9, 17, 18, 23, 24, 27, 30, 31, 41, 166
- orthographe, 30, 269

H

- hybride, 73, 74, 75, 119, 121, 124, 125, 126, 155, 156, 158, 159, 161, 171, 174, 175, 176, 209, 228, 262, 287, 347, 351, 352, 353, 356, 357

J

- juxtaposition, 8, 10, 22, 55, 56, 69, 110, 122, 152, 157, 297, 302
- juxtaposé, 13, 15, 22, 50, 51, 56, 81, 123

L

- lexicalisation, 9, 29, 39
- lexicalisé, 7, 9, 10, 15, 16, 18, 29, 42, 301
- lexicalité, 9, 35, 40, 41, 42
- lexicologie, 10, 294, 295, 298, 299, 300, 302
- lexie, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 35, 36, 37, 39, 40, 41, 44, 45, 46, 48, 51, 64, 68, 71, 72, 74, 75, 80, 82, 83, 89, 90, 91, 92, 93, 95, 100, 101, 102, 104, 109, 110, 111, 115, 117, 120, 131, 132, 133, 134, 136, 140, 142, 143, 146, 147, 149, 156, 166, 167, 169, 170, 173, 175, 176, 183, 191, 192, 195, 196, 197, 198, 199, 200, 201, 202, 204, 205, 207, 222, 230, 288, 299

-
- lexie complexe, 9, 14, 17
 - lexie composée, 14, 15, 24, 37
 - lexie dérivée, 14, 15
 - lexie prépositionnelle, 14, 15, 17
 - lexie primaire, 14, 15
 - lexie simple, 14, 15, 35, 192
 - lexique, 1, 2, 3, 5, 9, 10, 11, 12, 14, 15, 17, 21, 22, 23, 39, 46, 57, 59, 60, 62, 64, 75, 81, 83, 89, 91, 110, 111, 115, 117, 118, 120, 121, 122, 127, 133, 140, 144, 151, 156, 165, 168, 181, 186, 192, 196, 198, 202, 204, 206, 228, 236, 253, 256, 285, 287, 288, 290, 296, 297, 298, 300, 303

M

- matrice lexicogénique, 10, 15, 22, 56, 109, 110, 113, 149, 152, 181, 288, 289
- métaphore, 10, 22, 23, 38, 40, 71, 93, 103, 105, 106, 108, 110, 111, 112, 113, 114, 115, 116, 117, 140, 145, 147, 151, 152, 174, 178, 182, 183, 196, 288
- métaphorique, 103, 114, 146, 151, 152, 178, 179, 230
- métonymie, 10, 106, 108, 110, 113, 114, 115, 117, 151, 182, 183, 230, 288
- métonymique, 117, 151, 178
- mot composé, 1, 7, 8, 9, 10, 14, 23, 39, 40, 41, 195, 294, 295, 301
- motivation, 22, 105, 178, 287,
- motivé, 22, 105, 144

N

- nom composé, 3, 4, 5, 7, 9, 10, 11, 13, 16, 19, 21, 22, 23, 26, 38, 44, 46, 48, 54, 59, 65, 66, 67, 68, 69, 70, 71, 79, 104, 106, 107, 108, 114, 118, 120, 121, 128, 146, 147, 150, 151, 154, 165, 179, 180, 182, 186, 187, 192, 193, 195, 198, 201, 204, 229, 235, 236, 240, 256, 270, 285, 286, 287, 288, 289, 290, 291, 296, 298, 301, 321, 330, 334, 338, 343, 347, 353, 358, 361, 365

O

- onomatopée, 10, 105, 110

P

- patron compositionnel, 72, 73, 79, 88, 204, 207, 209, 221, 235, 257, 267, 287, 288, 289
- patron syntaxique, 67, 68, 69, 70, 71, 74, 75, 76, 80, 81, 93, 118, 130, 132, 133, 140, 147, 154, 156, 164, 175, 205, 206, 221, 234, 286, 298, 343, 347
- structure compositionnelle, 69, 79, 80, 119, 154, 163, 206, 234, 261, 262, 276, 286, 312, 370, 381
- structure syntaxique, 16, 70
- préfixation, 10, 15, 110, 113, 115, 117, 151, 183
- pseudo-confixation, 22, 86, 111, 112, 114, 115, 151, 182, 183, 287
- pseudo-confixe, 12, 13, 15, 30, 71, 72, 74, 75, 79, 81, 82, 83, 84, 86, 88, 89, 90, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 111, 112, 113, 121, 122, 123, 125, 126, 127, 130, 131, 133, 134, 135, 136, 138, 139, 140, 141, 142, 143, 145, 156, 157, 160, 162, 165, 166, 167, 168, 169, 170, 172, 173, 174, 175, 176, 177, 178, 182, 204, 234, 287, 288, 311

Q

- quasi-morphème, 11, 12, 13, 15

R

- relation sémantique, 37, 106, 107, 108, 109, 146, 147, 148, 179, 180, 181, 187, 234, 287
- relation N1-N2, 106, 147, 179, 180, 187, 232, 234, 235, 289

S

-
- séquence fortuite, 7, 10, 13, 14, 17, 18, 21, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 38, 39, 285
 - syntagme, 3, 7, 9, 10, 15, 16, 17, 18, 27, 31, 32, 39, 293
 - siglaison, 10, 88, 110, 112, 115, 117, 151, 183
 - sigle, 1, 13, 51, 52, 71, 74, 77, 81, 85, 98, 113, 120, 121, 124, 125, 127, 130, 140, 156, 159, 188, 229, 230, 250, 253, 257, 271, 273, 275, 276, 279, 282, 283, 284, 287, 300, 311, 312, 371, 376, 381
 - suffixation, 10, 15, 110, 115, 117, 151, 183
 - surcomposé, 12, 23, 30, 54, 56, 69, 71, 74, 75, 77, 82, 83, 84, 85, 86, 88, 96, 101, 104, 112, 118, 123, 124, 125, 126, 127, 128, 130, 134, 139, 141, 158, 159, 160, 161, 162, 163, 164, 168, 171, 173, 174, 188, 192, 195, 212, 218, 221, 239, 242, 345, 346, 347, 351, 352, 356, 357
 - système d'équivalences types, 41, 43, 57, 61, 65, 67, 68, 69, 70, 77, 103, 106, 109, 144, 145, 178, 181, 186, 205, 206, 221, 223, 228, 230, 231, 232, 235, 247, 284, 285, 286, 287, 288, 289, 290

T

- terminologie, 3, 9, 12, 13, 15, 22, 28, 38, 59, 62, 67, 72, 79, 82, 83, 84, 85, 96, 117, 122, 123, 124, 125, 130, 133, 145, 149, 154, 157, 160, 187, 200, 236, 240, 287, 288, 293, 294, 297, 298, 299, 301
- traducteur spécialisé, 4, 46, 187, 193, 195, 206, 235, 285, 289
- traduction, 1, 2, 3, 4, 5, 41, 42, 46, 51, 60, 69, 70, 106, 122, 151, 152, 170, 175, 184, 186, 187, 188, 192, 193, 194, 199, 200, 205, 206, 207, 228, 229, 230, 232, 235, 236, 237, 240, 241, 242, 245, 262, 285, 286, 288, 289, 290, 293, 298, 299, 302, 311, 325, 326
- TA, 235, 236, 240, 243, 245, 246, 285, 289
- TAO, 235, 236, 289
- traducteur automatique, 206, 235
- traduction automatique, 186
- troncation, 10, 12, 22, 46, 75, 82, 89, 94, 95, 97, 98, 100, 101, 110, 111, 112, 133, 134, 135, 140, 141, 167, 168, 170, 171, 173, 174
- tronqué, 12, 14, 77, 140, 142, 175
- typologie morphosémantique, 232, 233
- typologie morphosyntaxique, 5, 67, 286, 289
- typologie sémantique, 54, 56, 68

U

- unité lexicale, 4, 9, 10, 12, 14, 16, 17, 23, 24, 38, 40, 42, 74, 75, 104, 120, 145, 156, 193, 285, 290, 296

Annexe 1. Liste des figures

- Figure 1 : Fréquence des pseudo-confixes anglais du corpus
- Figure 2 : Centricité des composés anglais du corpus
- Figure 3 : Fréquence des pseudo-confixes français
- Figure 4 : Centricité des composés français du corpus
- Figure 5 : Fréquence des pseudo-confixes espagnols
- Figure 6: Centricité des composés espagnols du corpus
- Figure 7 : Fréquence des équivalences types dans la traduction des N + N anglais
- Figure 8 : Menu général
- Figure 9 : Fenêtre Présentation
- Figure 10 : Fenêtre Guide de l'utilisateur
- Figure 11 : Fenêtre À Propos
- Figure 12 : Saisie du composé
- Figure 13 : Fenêtre d'affichage des équivalents français
- Figure 14 : Fenêtre d'échec d'affichage des équivalents
- Figure 15 : Fenêtre de vérification orthographique
- Figure 16 : Fenêtre d'échec d'affichage des équivalents après vérification orthographique

- Figure 17 : Fenêtre d'identification de la vedette
- Figure 18 : Boîte de dialogue de saisie du composé
- Figure 19 : Fenêtre d'affichage de la vedette anglaise
- Figure 20 : Boîte de dialogue de saisie du composé
- Figure 21 : Fenêtre d'affichage des équivalents
- Figure 22 : Boîte de dialogue de saisie du sigle
- Figure 23 : Fenêtre d'affichage de la vedette anglaise
- Figure 24 : Boîte de dialogue de saisie du composé
- Figure 25 : Fenêtre d'affichage des équivalents
- Figure 26 : Saisie de la structure compositionnelle
- Figure 27 : Fenêtre d'affichage des équivalences types
- Figure 28 : Fenêtre d'affichage des équivalents français
- Figure 29 : Boîte de dialogue
- Figure 30 : Fenêtre d'affichage des exceptions
- Figure 31 : Menu Informations connexes
- Figure 32 : Boîte de dialogue de saisie du composé
- Figure 33 : Fenêtre d'affichage des termes français implantés
- Figure 34 : Boîte de dialogue
- Figure 35 : Fenêtre d'affichage des variantes anglaises
- Figure 36 : Boîte de dialogue
- Figure 37 : Fenêtre d'affichage des synonymes anglais
- Figure 38 : Boîte de dialogue
- Figure 39 : Fenêtre d'affichage des sigles anglais
- Figure 40 : Boîte de dialogue
- Figure 41 : Fenêtre d'affichage par champ notionnel

Annexe 2. Liste des tableaux

- Tableau 1 : Nombre d'internautes
- Tableau 2 : Accentuation des A + N espagnols
- Tableau 3 : Accentuation des N + N espagnols
- Tableau 4 : Typologie de Tournier
- Tableau 5 : Typologie de G. Gross
- Tableau 6 : Utilisation des patrons syntaxiques des noms composés anglais dans le système d'équivalences types
- Tableau 7 : Structure principale de la typologie des patrons compositionnels anglais
- Tableau 8 : Structure principale de la typologie des patrons compositionnels français
- Tableau 9 : Structure principale de la typologie des types compositionnels espagnols
- Tableau 10 : Structure secondaire de la typologie des patrons syntaxiques
- Tableau 11 : Structures compositionnelles des composés anglais
- Tableau 12 : Structures compositionnelles des surcomposés anglais à trois éléments
- Tableau 13 : Structures compositionnelles des surcomposés anglais à quatre éléments
- Tableau 14 : Structures compositionnelles des surcomposés anglais à cinq et six éléments

- Tableau 15 : Structures compositionnelles des composés et surcomposés anglais
- Tableau 16 : Composés anglais du corpus avec le pseudo-confixe cyber-
- Tableau 17 : Composés anglais du corpus avec le pseudo-confixe hyper
- Tableau 18 : Composés anglais du corpus avec le pseudo-confixe meta-
- Tableau 19 : Composés anglais du corpus avec le pseudo-confixe e-
- Tableau 20 : Composés anglais du corpus avec le pseudo-confixe net
- Tableau 21 : Composés anglais du corpus avec le pseudo-confixe web-
- Tableau 22 : Composés anglais du corpus avec le pseudo-confixe -ware
- Tableau 23 : Synthèse des relations sémantiques des composés anglais du corpus
- Tableau 24 : Combinaisons des matrices lexicogéniques du corpus anglais
- Tableau 25 : Fréquence des combinaisons de matrices dans le corpus anglais
- Tableau 26 : Structures compositionnelles endogènes et exogènes des composés du corpus français
- Tableau 27 : Structures compositionnelles endogènes des composés du corpus français
- Tableau 28 : Structures compositionnelles endogènes et exogènes des surcomposés à trois éléments du corpus français
- Tableau 29 : Structures compositionnelles des surcomposés endogènes à trois éléments du corpus français
- Tableau 30 : Structures compositionnelles endogènes et exogènes des surcomposés à quatre et cinq éléments du corpus français 126
- Tableau 31 : Structures compositionnelles des surcomposés endogènes à quatre et cinq éléments du corpus français
- Tableau 32 : Répartition des noms composés anglais et français selon le nombre de formants
- Tableau 33 : Structures compositionnelles des composés et surcomposés français
- Tableau 34 : Composés français avec le pseudo-confixe cyber-
- Tableau 35 : Composés français avec le pseudo-confixe hyper-
- Tableau 36 : Composés français avec le pseudo-confixe télé-
- Tableau 37 : Composés français avec le pseudo-confixe meta-
- Tableau 38 : Composés français avec le pseudo-confixe -web-
- Tableau 39 : Composés français avec le pseudo-confixe -net
- Tableau 40 : Composés français avec le pseudo-confixe e-
- Tableau 41 : Composés français avec le pseudo-confixe -naute
- Tableau 42 : Composés français avec le pseudo-confixe -ware
- Tableau 43 : Synthèse des relations sémantiques des composés français du corpus

-
- Tableau 44 : Résumés des relations sémantiques fréquentes en français et anglais
 - Tableau 45 : Combinaisons des matrices lexicogéniques du corpus français
 - Tableau 46 : Fréquence des combinaisons de matrices dans le corpus français
 - Tableau 47 : Structures compositionnelles endogènes et exogènes des composés binaires du corpus espagnol
 - Tableau 48 : Structures compositionnelles endogènes des composés binaires du corpus espagnol
 - Tableau 49 : Structures compositionnelles endogènes et exogènes des surcomposés à trois éléments du corpus espagnol
 - Tableau 50 : Structures compositionnelles des surcomposés endogènes à trois éléments du corpus espagnol
 - Tableau 51 : Structures compositionnelles endogènes et exogènes des surcomposés à quatre et cinq éléments du corpus espagnol
 - Tableau 52 : Structures compositionnelles des surcomposés endogènes à quatre et cinq éléments du corpus espagnol
 - Tableau 53 : Les structures compositionnelles des composés et surcomposés espagnols
 - Tableau 54 : Répartition des noms composés anglais et espagnols selon le nombre de formants
 - Tableau 55 : Noms composés espagnols avec le pseudo-confixe ciber-
 - Tableau 56 : Composés espagnols avec le pseudo-confixe hiper-
 - Tableau 57 : Composés espagnols avec le pseudo-confixe tele-
 - Tableau 58 : Composés espagnols avec le pseudo-confixe meta-
 - Tableau 59 : Composés espagnols avec le pseudo-confixe web-
 - Tableau 60 : Composés espagnols avec le pseudo-confixe net-
 - Tableau 61 : Composés espagnols avec le pseudo-confixe e-
 - Tableau 62 : Composés espagnols avec le pseudo-confixe -nauta
 - Tableau 63 : Composés espagnols avec le pseudo-confixe -ware
 - Tableau 64 : Synthèse des relations sémantiques des composés espagnols du corpus
 - Tableau 65 : Relations sémantiques fréquentes
 - Tableau 66 : Combinaisons des matrices du corpus espagnol
 - Tableau 67 : Fréquence des combinaisons de matrices dans le corpus espagnol
 - Tableau 68 : Application des critères définitoires des lexies scientifiques et techniques au domaine d'Internet
 - Tableau 69 : Étiquettes de domaine attribuées par le GDT et Netgloss
 - Tableau 70 : Échantillon trilingue de termes

- Tableau 71 : Structures françaises équivalentes au patron compositionnel anglais N + N
- Tableau 72 : Équivalences du patron N + N
- Tableau 73 : Équivalences du patron A + N
- Tableau 74 : Équivalences du patron PS + N
- Tableau 75 : Équivalences du patron N+ PS
- Tableau 76 : Équivalences du patron A + PS
- Tableau 77 : Équivalences des patrons anglais mineurs
- Tableau 78 : Équivalences du patron N + N + N
- Tableau 79 : Équivalences du patron A + N + N
- Tableau 80 : Équivalences du patron off + N + N
- Tableau 81 : Équivalences du patron N + A + N
- Tableau 82 : Équivalences du patron N + up + N
- Tableau 83 : Équivalences du patron PS + N + N
- Tableau 84 : Équivalences du patron A + PS + N
- Tableau 85 : Équivalences des patrons mineurs (3 éléments)
- Tableau 86 : Équivalences du patron N + N
- Tableau 87 : Équivalences du patron A + N
- Tableau 88 : Équivalences du patron PS + N
- Tableau 89 : Équivalences du patron A + PS
- Tableau 90 : Équivalences du patron N + PS
- Tableau 91 : Équivalences des patrons mineurs
- Tableau 92 : Équivalences du patron N + N + N
- Tableau 93 : Équivalences du patron A + N + N
- Tableau 94 : Équivalences du patron PS + N + N
- Tableau 95 : Équivalences du patron A+ PS + N
- Tableau 96 : Équivalences des patrons mineurs
- Tableau 97 : Test du système d'équivalences sur les composés anglais N + N du corpus
- Tableau 98 : Équivalents français des composés anglais exocentriques
- Tableau 99 : Croisement des typologies morphosémantiques
- Tableau 100 : Test de SYSTRAN
- Tableau 101 : Test de IT Dictionary
- Tableau 102 : Test de Tradinfo
- Tableau 103 : Test de Babylon-Pro : anglais-espagnol.

-
- Tableau 104 : Test de Babylon Pro : anglais-français.
 - Tableau 105 : Les composés binaires anglais
 - Tableau 106 : Les surcomposés anglais à trois éléments
 - Tableau 107: Les surcomposés anglais à quatre éléments
 - Tableau 108 : Les surcomposés à cinq et six éléments
 - Tableau 109 : Les composés binaires français
 - Tableau 110: Les surcomposés français à trois éléments
 - Tableau 111 : Les surcomposés français à quatre et cinq éléments
 - Tableau 112 : Les composés binaires espagnols
 - Tableau 113 : Les surcomposés espagnols à trois éléments
 - Tableau 114 : Les surcomposés espagnols à quatre et cinq éléments
 - Tableau 115 : Les noms composés anglais
 - Tableau 116 : Les noms composés français
 - Tableau 117 : Les noms composés espagnols

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Annexe 3. Application du critère dictionnaire à un échantillon de collocations

Les noms composés anglais français et espagnols du domaine d'Internet :

	Collocations	Tests	Commentaires	Résultats
1	démenti formel	T1= C, B	seulement un exemple dans l'article de la base, pas de définition / exemple: démenti officiel	0
		T2 = 0	autres exemples : déclaration, refus, consentement	
2	ferme résolution	T1= C		0
		T2 = 0	autres exemples : refus, intension, volonté	
3	activité fiévreuse	T1 = C		0
		T2 = 0	autre exemple : excitation	
4	vif intérêt	T1		?
		T2 = 0	exemples cités : émotion, satisfaction, regret, besoin, succès, discussion	
5	soin minutieux	T1 = C		0
		T2 = 0		
6	attention soutenue	T1 = C, B	seulement un exemple dans l'article de la base, pas de définition / exemples : suivie	0
		T2 = 0		
7	agitation fébrile	T1 = C		0
		T2 = 0	autres exemples : mouvement, inquiétude	
8	preuve irréfutable	T1 = C, B	seulement un exemple dans l'article de la base, pas de définition	0
		T2	autre exemple : argument	
9	refus catégorique	T1 = B	seulement un exemple dans l'article de la base, pas de définition / exemple : humiliant	?
		T2 = 0	exemples cités : affirmation, réponse, position, ton	
10	liens indissolubles	T1 = C		0
		T2	autres exemples : attachement, mariage	
11	sanglant outrage	T1	définition : profondément blessant, extrêmement dur et outrageant	?
		T2 = 0		
12	profonde reconnaissance	T1 = C	autres définitions : silence, transformation, influence, différence, joie	?
		T2 = 0		
13	envie folle	T1 = C	autres exemples : idée, tentative, course, passion	0
		T2 = 0		
14	rapidité vertigineuse	T1 = C	autres exemples : hauteur, chute, vitesse	0
		T2 = 0		
15	prix exorbitant	T1 = C, B	seulement un exemple dans l'article de la base, pas de définition / autres ex : normal, élevé	0
		T2 = 0		
16	victoire décisive	T1	exemples cités : bataille, jeu, intervention	?
		T2 = 0		
17	ignorance crasse	T1 = C, B	seulement un exemple dans l'article de la base, pas de	0

Annexe 3. Application du critère dictionnaire à un échantillon de collocations

			définition	
		T2 = 0		
18	laideur repoussante	T1 = C		?
		T2 = N		
19	avarice sordide	T1 = C, B	seulement un exemple dans l'article de la base, pas de définition	0
		T2 = 0		
20	luxe princier	T1 = C		?
		T2 = N		
21	répugnance insurmontable	T1		N
		T2	exemples : angoisse, aversion	

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Annexe 4. Application du triple test de compositionnalité à dix candidats composés

Les noms composés anglais français et espagnols du domaine d'Internet :

	Candidat composé	Critère appliqué	Résultat	Bilan
1	analog modem	T1	0	0
		T2	0	
		T3	0	
2	anonymous surfing	T1	0	0
		T2	0	
		T3	0	
3	bulk email	T1	0	0
		T2	0	
		T3	0	
4	dedicated server	T1	0	0
		T2	0	
		T3	0	
5	download site	T1	0	0
		T2	0	
		T3	0	
6	electronic mail	T1	0	0
		T2	0	
		T3	0	
7	mail server	T1	0	0
		T2	0	
		T3	0	
8	spyware	T1	0	0
		T2	0	
		T3	0	
9	stealth virus	T1	0	0
		T2	0	
		T3	0	
10	user interface	T1	0	0
		T2	0	
		T3	0	

Légende : T1 = Le candidat composé répond à la définition du nom composé T2 = Le candidat composé est inclus dans un ouvrage de référence T3 = Le candidat composé est par un moteur de recherche (> 5 occurrences) Résultat : 0 = Réussite du test, N = Échec du test Bilan : 0 = Le candidat composé est un nom composé, N = Le candidat composé est un autre groupement

Annexe 5. Fiches des revues anglaises, françaises et espagnoles

1. Fiches des revues anglaises

- ***Internet Magazine*** (3 numéros)

Numéros : janvier, juin et novembre 2001.

Type : Revue papier. Existe depuis 2000.

Lectorat : *Internet Magazine* s'adresse aux internautes « sérieux ».

Pourquoi : Il donne des informations sur les nouveautés et les derniers développements. Des experts prodiguent des conseils.

Maison d'édition : Emap automotive.

- ***Network Magazine*** (3 numéros)

Numéros : janvier, février et mars 2001.

Type : Revue électronique (www.networkmagazine.com). Version papier également

disponible. Existe depuis janvier 1999.

Lectorat : Ce magazine est une passerelle entre les entreprises et le NPN (New Public Network).

Pourquoi ? : Il donne des informations pratiques intéressant les spécialistes des technologies de l'information (IT) et met l'accent sur les tutorats, les analyses de produits et il se consacre aussi aux entreprises et NPN.

Maison d'édition : CMP Media.

· ***Wired*** (4 numéros)

Numéros : décembre 2000, février, mai et octobre 2001.

Type : Revue électronique (www.wired.com/wired/current.html). Version papier disponible. Existe depuis mars 1993.

Lectorat : Cette revue s'adresse aux spécialistes des technologies de pointe et aux personnes simplement intéressées par les nouveautés et tournées vers le futur.

Pourquoi : *Wired* couvre tous les acteurs d'Internet – aussi bien les particuliers que les entreprises – et informe ses lecteurs sur les dernières avancées. C'est une revue informatique appréciée (d'après le nombre de parutions), proche du domaine Internet et qui existe depuis de nombreuses années.

Editeur : The Condé Nast Publications Inc.

· ***PC World*** (12 numéros)

Numéros : janvier à décembre 2001. Le nombre de numéros est plus important que pour les autres revues car *PC World* est une revue informatique. Le nombre d'articles relatifs à Internet est inférieur à celui des autres magazines Internet et même à *Wired*, elle aussi une revue informatique.

Type : Revue électronique (www.pcworld.com). Disponible en version papier. Existe depuis 1998.

Lectorat : *PC World* a pour objectif de satisfaire les besoins informationnels des lecteurs qui ont des connaissances techniques ou non. Il est utile aussi bien pour les activités personnelles que professionnelles.

Pourquoi : C'est une aide pour la conception, l'achat, l'intégration et l'utilisation de produits informatiques (software et hardware). Avec 6,9 millions de lecteurs en 2001 – pour la version papier – ce magazine est une des publications informatiques les plus lues. Le site Web a reçu 1,7 millions de visiteurs uniques/mois (Nielsen//NetRatings, Déc. 2002 - Fév. 2003, U.S.).

Editeur : PC World communications, une filiale de IDG.

2. Fiches des revues françaises

· ***Officiel du Net*** (3 numéros)

Numéros : octobre, hors-série d'octobre et novembre 2001.

Type : Revue papier. Existe depuis février 1999.

Lectorat : L'*Officiel du Net* s'adresse à des personnes ayant une base informatique qui recherchent des informations variées.

Pourquoi : Ce magazine regorge de renseignements sur toutes les activités du Net et étudie les logiciels de navigation, les outils de création de site pour dévoiler leurs petits plus. C'est le premier magazine de poche en termes de diffusion et de vente. Pour 2000-2001, sa diffusion totale est 46 701 numéros.¹⁹¹

Editeur : Hagal Aria.

· ***Netsurf*** (3 numéros)

Numéros : janvier, février et mars 2001 (plus de parution à partir de février 2002).

Type : Revue papier. Existe depuis janvier 1996.

Lectorat : *Netsurf* est à la fois généraliste et professionnel. C'est le premier magazine portail destiné à tous les internautes. Il s'adresse à l'ensemble des cadres et des actifs connectés. Il les aide à consommer sur le Net et à intégrer le Net à leur vie.

Pourquoi : C'est un « mode d'emploi » du Net qui s'adapte aux nouvelles réalités. La diffusion de cette revue a atteint 18 178 numéros en 2000-2001 (contre 24 092 en 2000).

Editeur : Pressimage puis Infosources depuis avril 2000.

· ***.Net*** (3 numéros)

Numéros : septembre, octobre et novembre 2001.

Type : Revue papier. Existe depuis 1998.

Lectorat : *.Net* intéresse surtout aux utilisateurs expérimentés ou « futurs initiés ».

Pourquoi ? : Il présente l'actualité des sites et prodigue de nombreux conseils pratiques pour surfer ou construire un site. Les lecteurs y trouvent également une sélection d'adresses Internet, un choix de matériels et de logiciels. Ce magazine a une large diffusion – 42 833 numéros pour l'année 2000-2001.

Editeur : EDICORP Publications.

· ***Web magazine*** (3 numéros)

Numéros : janvier, avril et août 2001.

Type : Revue papier. Existe depuis mai 1999.

Lectorat : *Web magazine* attire un public moins informé et plus large que les revues précédentes.

¹⁹¹ Chiffres certifiés par l'OJD qui valide les ventes et la diffusion : www.ojd.com

Pourquoi : Cette revue constitue une aide à la navigation et donne des astuces pour faciliter le surf. En 2000-2001, c'est le magazine qui a la diffusion la plus importante avec 119 683 numéros.

Editeur : Prisma Presse.

3. Fiches des revues espagnoles

- ***En la Red*** (3 numéros)

Numéros : janvier, février et mars 2002.

Type : Revue papier. Existe depuis 1997.

Lectorat : Ce magazine est susceptible d'intéresser tous les internautes car il aborde toutes les facettes du Web.

Pourquoi : *En la Red* parle des nouveautés du monde Internet parce que, selon la rédaction, la frontière entre Internet et le quotidien de chacun est de plus en plus étroite.

Editeur : Ifactoría del Conocimiento.

- ***Netsurf*** (3 numéros)

Numéros : janvier, février et mars 2001.

Type : Revue papier.

Lectorat : *Netsurf*, version espagnole, est également un magazine portail généraliste et professionnel. Il s'adresse aussi à tous les internautes – peut-être débutants.

Pourquoi : Il présente des sites Web et les nouveaux produits, il explique comment créer et gérer un site. *Netsurf* n'est pas la traduction de la version française.

Editeur : IXO Publishing Ibérica puis Freeway ediciones.

- ***IWorld*** (3 numéros)

Numéros : mars, juin et octobre 2001.

Type : Revue papier. Existe depuis janvier 1997.

Lectorat : *IWorld* s'adresse aux spécialistes intéressés surtout par le e-business.

Pourquoi : Il présente les stratégies d'affaires sur le Net, les nouveaux produits, des sites Internet et des jeux. Il analyse les tendances du Net et les stratégies qui permettront aux entreprises de réaliser leurs projets.

Editeur : IDG communications S.A..

- ***PC World*** (4 numéros)

Numéros : février, juin, septembre et décembre 2001.

Annexe 5. Fiches des revues anglaises, françaises et espagnoles

Type : Revue électronique (www.idg.es/pcworld/). Existe aussi en version papier depuis 1999.

Lectorat : *PC World* a les mêmes objectifs que la version anglaise (cf. p.00).

Pourquoi ? : La version espagnole de *PC World* n'est pas une traduction de la version anglaise. Elle attire 64 537 lecteurs (avril 2001-mars 2002).

Editeur : IDG communications.

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Annexe 6. Rubriques des revues anglaises, françaises et espagnoles

1. Revues anglaises

- Internet Magazine

Rubriques retenues : Browser, Opinion (The Internet Magazine interview, Campaign Update), Features, Site Surveyor (Essentials viewing, Home improvement), Expert Help, Tried and Tested (Lab).

Rubriques rejetées : Browser (Services), Site Surveyor (Site of the month, Opinion, Mailbox), Spotlight on, Tried and Tested (Products, Providers, Hosts).

- Network magazine

Rubriques retenues : le thème du numéro, NPN features, Product focus (software), Viewpoint, News and Analysis, Off the Wires, Global Watch, Tutorials, Best Practices, Network Defense, Business Layer.

Rubriques rejetées : Letters, Product Spotlight, News Products, Wide Angle.

- ***Wired***

Rubriques retenues : New Media (Multimedia, Virtual Communities, Web Development), New Economy, New Technology (Connectivity, Hacking and Warez), New Policy (Censorship, Encryption, Internet Law, Privacy).

Rubriques rejetées : New Media (Film and Special Effect, Gaming, Music and Sound), New Technology (Computers, Genetics, Planes, Trains and Autos, Robots and AI, Space exploration, Y2K).

- ***PC World***

Rubriques retenues : News and Trends, Here's How (Internet Tips, Upgrade Guide), Department (Up Front, Consumer Watch, Web Savvy, Bugs and Fixes, Home Office), New Products (Software, Internet, Navigation).

Rubriques rejetées : Top 100, Here's How (Hardware Tips, Word Processing Tips, Windows Tips), Department (Letters to PC World, New Products, On your side, Home Office).

2. Revues françaises

- **Officiel du Net**

Rubriques retenues : Dernières actualités du Web, Hack News, On a testé...(software), Nettement étrange, Epopée, Dossier, Ce qui fait débat, Il y a un an, Net expliqué, Mise en œuvre, Téléchargement, Petits plus, Trucs et astuces.

Rubriques rejetées : Courrier des lecteurs, Cine club, Arrêt sur image, Espace lecteurs, Sites du mois, Voyage, Tribune libre, High tech, Livres et CD, Charme, Ca nous fait rire.

- ***Net***

Rubriques retenues : Guide pratique (fiches, trucs et astuces), Webmestre (vos sites en questions), Rubriques (forum de discussion, actualités, logiciels), Thématique du mois.

Rubriques rejetées : NetGuide (30 jours d'événements, sélection, musique, pratique, insolite, humour, sport, bookmark), Webmestre (zone php, zone javascript), Rubriques (Net shopping, Matériel, Sur le CD-ROM, Quel fournisseur d'accès, Tapage !).

- ***Web magazine***

Rubriques retenues : Magazine (En direct du Web, Dossier complet, Bons plans, Parlons net, Bonnes idées du mois, Cyberscience), Net business (Stratégies, Echos, Nouveau métier), Cahier pratique (Offres d'accès, Mails, Forum).

Rubriques rejetées : Magazine (Courrier, Culture Web, Actu, shopping, Gros plan sur, Test, CD, DVD, Evénements, Net vu d'ailleurs), Web Guide (Agenda, Test du mois, Sites préférés de..., Charme, Culture, Cuisine, Histoires), Cahier pratique (Jeux).

- Netsurf

Rubriques retenues : Netzine (Economie connectée, En ligne et hors ligne), Nettech (Meilleurs outils, Dernières innovations, Produits indispensables pour naviguer, pour créer sont testés et expliqués pour vous aider à tirer le meilleur du Net), Netzoom (Dossier avec panorama, Conseils), Netbiz (Bien utiliser le Net dans une entreprise).

Rubriques rejetées : Netzoom (Témoignages), @home (Surfer pour se distraire et se détendre).

3. Revues espagnoles

- Netsurf

Rubriques retenues : Actualidad (Otras fronteras, Parlamento digital, e-Economía, Portaleando, Se busca, Casos y cosas, En el laboratorio, En clase, Los números, Ellos y ellas, La entrevista), Informe, En práctica, Net Direct, Crear (Practica HTML, Cómo ?, Análisis, Estudios de caso, Dossier), Netsites (Top hispano, Top inglés, Top francés, Temas, Vuelta a España, A su servicio, Comercio electrónico, ¿ Cómo ?, el ojo de la Red).

Rubriques rejetées : Actualidad (Ocio, Las letras, Primer tiempo).

- En la Red

Rubriques retenues : Actual.com, Informe, Entrevista, Ciberespacio, Noticias, Web entrevista, e-scaparte, Download.

Rubriques rejetées : Primer Vistazo a..., ¡Mira, sin cables !, Cine, Tecnología en casa, Juegos, Laboratorio (hardware), Kit de construcción, Música online, Toolbox.

- Iworld

Rubriques retenues : Artículos, Nuevos productos (software), Actualidad, Noticias, Empresas, Comercio electrónico, Arte digital, Mundo Internet 2001, Entrevista, Informe, De viaje por la Red, Tecnología, Práctico, Software público, Soluciones y trucos, Primer contacto, Mujeres, Educación, Opinión, Seguridad, Economía digital.

Rubriques rejetées : Cartas de los lectores, Nuevos productos (hardware), Juegos en la Red, Libros, Guía del ocio.

- PC World

Les noms composés anglais français et espagnols du domaine d'Internet :

Rubriques retenues : Correo electrónico, Actualidad, Reportaje, Nuevos productos (software, internet), De viaje por la Red, Comparativa, Seguridad, Práctico (Turcos Internet), Soluciones y trucos, Entrevista, Dossier (Internet), Noticias (Internet).

Rubriques rejetées : Editorial, Cartas, Nuevas tecnologías, Nuevos productos (Hardware, Periféricos, Mundo de Bolsillo), Sucedío hace 10 años, Bazar, Primeros contactos, Comparativa, Test, Tutor, Práctico (PC, Windows, Linux), Mundo Linux, Rincón del programador, Ocio, Tecnología y sociedad, Juegos, Quinta dimensión.

Annexe 7. Les noms composés du corpus

1. Les noms composés anglais

access control
ad blocker
ad-blocking program
address bar
address book
ad-free zone
ADSL modem
ADSL provider
alias page
always-on access
always-on broadband connection

always-on connection
always-on routed Internet
analog modem
anonymous FTP
anonymous surfing
antivirus provider
Application Service Provider
archive manager
archive moderator
Asymmetric Digital Subscriber Line
asynchronous transfer
attack program
audience measurement
backdoor
bandwidth
banner ad
banner advertising
black hat
blacklist
Blind Carbon Copy
blocking access
bookmark
bookmark manager
Boolean operator
Boolean search
boot virus
bottleneck
broadband Internet
browser cache
browser sniffer
browsing habit
buffer overflow attack
bug hunter
bulk email

Bulletin Board System
business to consumer
Business to Employee
business-to-business
cable Internet access
cable modem
Cascading Style Sheet
central server
Challenge Handshake Authentication Protocol
chat group
chat host
chat server
chat session
chat software
chat window
chatbox
chat-mate
chatroom
click and mortar
click-and-mortar business
client interface
client-server architecture
collocated server package
commercial site
comparison site
connection kit
connection speed
connection time
contact group
contact list
cookie builder
cookie file
cookie ID
cookie management

Les noms composés anglais français et espagnols du domaine d'Internet :

cookie manager
cyber activity
cyber album
cyber directory
cyber market
cyber monitoring
cyber referendum
cyber shop
cyber visitor
cyberattack
cyberbrand
cybercafé
cyberconference
cyber-conflict
cybercop
cybercrime
cybercriminal
cybercriminality
cyber-debate
cyber-delinquency
cyber-delinquent
cybereditor
cyberethics
cyberflirt
cyberfrontier
cyber-grocer
cyberhacking
cyberlanguage
cyberlaw
cyberlove
cybermerchant
cyberprotection
cyberpunk
cyber-recruitment

cybersex
cybersoldier
cyberspace
cybersquatter
cybersquatting
cyberstress
cybersupermarket
cyberterrorism
cyberterrorist
cyberunivers
cyberworld
data packet
dedicated server
Denial of Service
Denial of Service attack
dial-up connection
dial-up modem
digital signature
discussion area
discussion board
discussion forum
discussion group
distribution list
DNS server
domain name
Domain Name System
DoS attack
dot com
dot-com industry
dotcom startup
dot-org
download
download manager
download site

download speed
download time
downloading
Dynamic HyperText Markup Language
dynamic IP address
ebiz
ebook
e-business
e-business integrator
e-card
e-commerce package
e-commerce site
e-commerce
e-customer
e-employee
e-job
e-learning
electronic hoax
electronic mail account
electronic message
electronic messaging
electronic messaging client
electronic newsletter
electronic signature
electroniccommerce
e-mail account
e-mail address
email alert
e-mail attachment
e-mail client
e-mail encryptor
e-mail filtering
e-mail header
email list

email manager
e-mail newsletter
email provider
e-mail server
email software
email virus
email worm
email
e-market place
emoticon
encrypted email
encrypted virus
encryption key
encryption program
encryption system
e-purchase
e-tourism
e-trade
e-trader
executable attack
eXtensible Markup Language
external modem
e-zine site
file sharing
file swapping
File Transfer Protocol
file-sharing program
file-swapping program
filtering package
filtering rule
firewall
flamewar
forwarding address
fraggle attack

Les noms composés anglais français et espagnols du domaine d'Internet :

frame relay
frame-to-frame animation
freeware
Frequently Asked Questions
FTP access
FTP client
FTP file
FTP server
FTP space
general portal
groupware
guestbook
hacked page
hacked site
hacker attack
hacker community
hacking attack
hacktivism
hacktivist
hardware firewall
high-speed access
high-speed connection
high-speed Internet
high-speed Web access
history folder
hit counter
home page
hosted data
hosting company
hosting package
hosting URL
hotline
HTML editor
HTML integrator

HTML page
HTML site
HTTP server
human-assisted search engine
hyperlink
hyperlinked video
HyperText Markup Language
HyperText Transfer Protocol
inbox
incoming mail server
indexed page
indexing engine
individual ADSL
information superhighway
instant message
instant messaging
instant messenger
instant translator
intelligent agent
Internet
Internet access
Internet access authorization
Internet access package
Internet Access Provider
Internet cache
Internet café
Internet community
Internet connection
Internet connection software
Internet content provider
Internet cookie
Internet gateway
Internet name
Internet portal

Internet Protocol
Internet server
Internet Service Provider
Internet tool
Internet traffic load
Internet user
intrusion detection system
invisible Web
IP address
IP network
IP router
IRC channel
IRC client
IRC server
Java applet
junk email
keylogger
leased line connection
live update
Local Area Network
local ISP
log file
log sniffer
low-speed connection
mail bombing
mail checker
mail forwarding
mail server
mailbox
mailing list
mailing-list management software
mailto
malicious code
Managed Service Provider

massive attack
message board
message filter
meta portal
meta search engine
meta tag
mirror site
mobile Internet
modem
MP3 file
MP3 site
multiaccess portal
multicriteria search
multipartite virus
musical website
Napsterized file
navigation bar
Net access
Net congestion message
Net connection
Net connection speed
Net economy
Net marketer
Net marketing
Net user
netizen
network administrator
network connection
network firewall
network of networks
newsgroup
newsgroup server
newsgroup software
newsgroupie

Les noms composés anglais français et espagnols du domaine d'Internet :

newsletter
official page
official site
offline browser
offline browsing
online privacy
online security
Online Service Provider
online shop
online shopping
open source app
openware
outcoming mail server
P2P network
page view
parody site
password
password cracker
peer-to-peer
peer-to-peer network
permanent cookie
personal home page
personal site
ping sweep
Point to Point Protocol
polymorphic virus
POP account
POP server
pop up
POP3 email box
POP3 mail box
pop-down ad
pop-under ad
pop-up ad

pop-up message
port blocking
port scanning
portal of portals
portal site
Post Office Protocol
price comparison site
private key
private virtual network
program virus
proxy firewall
proxy server
PSTN connection
public IP address
public key
read notification
receipt notification
remodelled site
remote access
remote computer
remote server
rollover
rotation banner
satellite modem
script blocking
script file
script kiddy
script virus
scrolling banner
search engine
search engine spider
secure certificate
secure server
Secure Socket Layer

security hole
server attack
server connection
server package
server-to-server architecture
shadow password
shared hosting
shared hosting package
shared IP address
shared mail server
shared server
shared server package
shareware
shopbot
Short Message Service
showroom
Simple Mail Transfer Protocol
simultaneous connection
site creation tool
site directory
site owner
site structure
Small Office Home, Office
SMS messaging
smurf attack
snail mail
software firewall
source code
space provider
spam protection
specific portal
spyware
startup
static IP address

static Web page
stealth virus
storage site
streaming video
style sheet
surfing habit
techno-nerd
teleworking
temporary cookie
thematic directory
thematic portal
thumbnail
transfer speed
Transmission Control , Protocol/Internet , Protocol
Trojan Horse
Trojan Horse detector
Trojan virus
underground site
Uniform Resource Locator
unique visitor
uplink
upload
URL history
Usenet group
Usenet posting
user account
user authentication
user interface
user interface expert
username
video chat
video chatting
videostream
viral attack

Les noms composés anglais français et espagnols du domaine d'Internet :

viral code
viral infection
viral worm
virtual hosting
virus alert
virus attack
virus definition
virus update
virus writer
virus-signature file
visible Web
Vmail
voice chat
voice email
voice IP
WAP locking
WAP page
WAP portal
Web accelerator
Web access
Web address
Web agency
Web attack
Web browser
web bug
Web cobrowsing
Web collaborating
Web community
Web design
Web designer
Web developer
Web directory
Web hosting
Web hosting company

Web manager
Web marketer
Web page
Web radio
web remix
web remixer
Web server
Web software
Web space
Web surfer
Web surfing habit
Web user interface
webcam
webcenter
weblog
weblogger
weblogging
webmail
webmaster
website
white hacker
white hat
Wireless Application Protocol
wireless connection
wireless Internet connection
wireless Web
World Wide Web
worm alert
worm attack

2. Les noms composés français

- accès à distance
- accès distant
- accès FTP
- accès haut-débit
- accès Internet
- accès Internet par câble
- accès permanent
- accès Web
- accès web haut-débit
- accusé de lecture
- accusé de réception
- achat en ligne
- administrateur de site
- administrateur réseau
- administrateur Web
- adresse de courrier électronique
- adresse d'hébergement
- adresse électronique
- adresse e-mail
- adresse Internet
- adresse IP
- adresse IP dynamique
- adresse IP fixe
- adresse IP partagée
- adresse IP publique
- adresse URL
- adresse Web
- ADSL individuel
- adware
- agent intelligent
- alerte de virus
- alerte e-mail
- animation frame par frame

annuaire de sites
annuaire thématique
annuaire Web
antihacker
antispam
applet Java
apprenti pirate
apprentissage en ligne
araignée de moteur de recherche
architecture client-serveur
architecture serveur-serveur
article Usenet
aspirateur de site
aspirateur Internet
aspirateur Web
assistant de téléchargement
Asymmetric Digital Subscriber Line
attaque buffer overflow
attaque de hacker
attaque de piratage
attaque de serveur
attaque de ver
attaque de virus
attaque Denial of Service
attaque DoS
attaque fraggle
attaque massive
attaque smurf
attaque virale
attaque Web
autorisation d'accès à Internet
autorité de certification
autoroute de l'information
babillard électronique

Les noms composés anglais français et espagnols du domaine d'Internet :

backdoor
balise meta
bande passante
bandeau publicitaire
bannière défilante
bannière publicitaire
bannière rotative
barre d'adresses
barre de navigation
black hat
blacklist
blocage de port
bloqueur de popup
boîte aux lettres
boîte aux lettres électronique
boîte aux lettres POP3
boîte de réception
bookmark
buffer overflow
business to business
business to consumer
Business to Employee
cache du navigateur
cache Internet
café Internet
canal de chat
canal IRC
canular électronique
carnet d'adresses
CD de connexion
certificat de sécurité
chasseur de bug
chat room
chat vocal

chatbox
cheval de Troie
chiffrement d'e-mail
citoyen du Net
clé de chiffrement
clé privée
clé publique
click and mortar
client de chat
client de courrier
client de courrier électronique
client de messagerie électronique
client e-mail
client FTP
client IRC
code malveillant
code source
code viral
commerce électronique
communauté du Web
communauté hacker
communauté Internet
comparateur de prix
compte de courrier
compte de courrier électronique
compte de messagerie
compte e-mail
compte pop
compte utilisateur
compteur d'accès
compteur de visite
concepteur Web
connexion à bande large permanente
connexion à faible débit

connexion dédiée
connexion en dialup
connexion haut débit
connexion Internet
connexion Internet sans fil
connexion par ligne commutée
connexion permanente
connexion réseau
connexion RTC
connexion sans fil
connexion serveur
connexion simultanée
contact list
contrôle d'accès
cookie Internet
cookie permanent
cookie temporaire
copie carbone invisible
courrier électronique
courrier électronique chiffré
courrier escargot
courrier non-sollicité
courrier postal
cracker de mot de passe
cyber amour
cyber civisme
cyber univers
cyberacheteur
cyberactivité
cyberalbum
cyberannuaire
cyberattaque
cyberboutique
cybercafé

cybercaméra
cyber-commerçant
cyberconférence
cyberconflit
cyberconsommateur
cybercrime
cybercriminalité
cybercriminel
cyberdébat
cyber-délinquence
cyberdélinquent
cyberdroit
cyberéditeur
cyberépicier
cyberespace
cyberflic
cyber-flirt
cyberfrontière
cyberguerre
cyberlangue
cybermarché
cybermarque
cybermonde
cyber-piraterie
cyber-plébiscite
cyberprotection
cyber-punk
cyberpunkitude
cyber-recrutement
cybersexe
cybersoldat
cybersquatter
cybersquatting
cyberstress

Les noms composés anglais français et espagnols du domaine d'Internet :

cyber-supermarché
cybersurveillance
cyberterrorisme
cyberterroriste
cybervisiteur
définition de virus
Déni de Service
design Web
designer Web
détecteur de chevaux de Troie
développeur Web
dialogue en direct
diffuseur de virus
donnée hébergée
dot org
dotcom
download
e-achat
e-biz
ebook
e-business
e-card
échange de fichier(s)
e-commerce
e-courrier
éditeur HTML
e-économie
e-employé
e-learning
e-mail
e-mail chiffré
e-mail crypté
e-mail non sollicité
e-métier

émoticône
en-tête de courrier
entreprenaute
entreprise point com
envoi groupé
e-picerie
e-place de marché
ergonome web
espace de discussion
espace ftp
espace Web
espiongiciel
e-tourisme
e-xagone
eXtensible Markup Language
faille de sécurité
fenêtre de chat
fenêtre pop under
fenêtre pop up
fenêtre publicitaire
feuille de style
feuille de style en cascade
fichier de signature de virus
fichier espion
fichier FTP
fichier journal
fichier log
fichier MP3
fichier Napster
File Transfer Protocol
filtrage d'e-mails
filtre anti-pub
filtre de messages
firewall

firewall logiciel
Foire Aux Questions
forum de discussion
fouineur de log
fournisseur ADSL
fournisseur d'accès
fournisseur d'accès Internet
fournisseur d'antivirus
fournisseur d'applications hébergées
fournisseur de contenu
fournisseur de contenu Internet
fournisseur de courrier
fournisseur de courrier électronique
fournisseur de services en ligne
fournisseur de services gérés
fournisseur de services Internet
fournisseur d'e-mail
fournisseur d'espace
fournisseur d'hébergement
fournisseur d'hébergement web
frame relay
freeware
Frequently Asked Questions
FSI local
FTP anonyme
générateur de cookies
générateur de virus
gestion de cookies
gestion de mailing list
gestionnaire d'archives
gestionnaire de cookies
gestionnaire de courrier électronique
gestionnaire de favoris
gestionnaire de téléchargement

gestionnaire Web
goulot d'étranglement
groupe de chat
groupe de contact
groupe de discussion
groupe de News
groupe Usenet
groupware
habitude de navigation
hacker blanc
hacker noir
hacktivisme
hacktiviste
haut débit
hébergement de site
hébergement partagé
hébergement virtuel
hébergeur de chat
home page
hotline
hotliner
hyperlien
HyperText Markup Language
inbox
infection virale
intégrateur e-business
intégrateur HTLM
interface client
interface graphique web
interface utilisateur
interface Web
internaute
Internet
Internet bande large

Les noms composés anglais français et espagnols du domaine d'Internet :

Internet haut débit
Internet mobile
Internet nomade
intimité en ligne
ISP local
jeune poussé
journal de connexion
key logger
kit de connexion
langage DHTML
Langage HTML
langage XML
lien consulté
lien hypertexte
Ligne d'abonnement numérique asymétrique
liste de contact
liste de diffusion
liste de distribution
liste noire
livre d'or
log de connexion
logiciel anti popup
logiciel anti-pub
logiciel d'accès Internet
logiciel de chat
logiciel de chiffrement
logiciel de connexion Internet
logiciel de courrier
logiciel de cryptage de courrier
logiciel de filtrage
logiciel de navigation
logiciel de news
logiciel de serveur
logiciel d'hébergement

logiciel espion
logiciel gratuit
logiciel libre
logiciel partagé
logiciel spyware
logiciel Web
magasin en ligne
mail bombing
mail vocal
mailing list
mailto
message board
message de congestion
message électronique
message instantané
message Internet
message non-sollicité
message pop up
message vocal
messagerie électronique
messagerie instantanée
messagerie SMS
mesure d'audience
meta tag
metabalise
métamoteur
metaportal
mise à jour anti-virus
mise à jour automatique
modem
modem ADSL
modem analogique
modem câble
modem commuté

modem externe
modem RTC
modem satellite
mot de passe
moteur à assistance humaine
moteur de recherche
moteur d'indexation
navigation anonyme
navigation hors-ligne
navigation off-line
Net-économie
nétiquette
netmarketeur
netmarketing
netsurfer
netsurfeur
newsgroup
newsletter
newsletter électronique
newsletter e-mail
nom de domaine
nom d'utilisateur
non-connecté
notificateur de mail
open source
openware
opérateur booléen
ordinateur distant
ossature de site
outil de création de site
outil de recherche
outil Internet
pack ADSL
package d'hébergement partagé

page alias
page d'accueil
page d'accueil perso
page d'accueil personnelle
page de démarrage
page HTML
page indexée
page officielle
page perso
page personnelle
page piratée
page visitée
page vue
page WAP
page Web
page Web statique
paquet d'information
pare-feu
pare-feu proxy
pare-feu réseau
partage de fichier(s)
passerelle Internet
peer-to-peer
pièce jointe
ping sweep
pirate informatique
pixel espion
point com
pop under
pop up
portail de portails
portail généraliste
portail Internet
portail multi-accès

Les noms composés anglais français et espagnols du domaine d'Internet :

portail spécifique
portail thématique
portail WAP
porte de derrière
porte de service
porte dérobée
programme d'accès Internet
programme d'attaque
programme de partage de fichier
programme d'échange de fichier
programme d'e-commerce
programme espion
programme open source
propriétaire de site
protection anti-spam
Protocole d'applications sans fil
Protocole de Transfert de Fichier
protocole http
protocole Internet
protocole pop
protocole PPP
protocole SMTP
protocole TCP/IP
publicité non-sollicitée
rapatriement du courrier
recherche multicritère
règle de filtrage
renifleur de navigateur
requête booléene
réseau des réseaux
réseau IP
réseau P2P
réseau peer-to-peer
réseau privé virtuel

responsable de la veille
robot logiciel
rollover
routeur IP
salon de chat
salon de discussion
scan de port
script kiddy
sécurité en ligne
serveur central
serveur de chat
serveur de courrier
serveur de courrier entrant
serveur de courrier sortant
serveur de News
serveur de newsgroup
serveur dédié
serveur distant
serveur DNS
serveur FTP
serveur HTTP
serveur Internet
serveur IRC
serveur mail
serveur mandataire
serveur partagé
serveur partagé de courrier
serveur POP
serveur proxy
serveur sécurisé
serveur Web
session de chat
shareware
shopbot

shopping en ligne
Short Message Service
showroom
signature de virus
signature électronique
signature numérique
site à assistance humaine
site d'archivage
site de comparaison
site de comparaison de prix
site de parodie
site de téléchargement
site d'e-commerce
site e-zine
site filleul
site hacké
site HTML
site Internet
site marchand
site mercantile
site mirror
site MP3
site officiel
site parodique
site parrain
site perso
site personnel
site piraté
site portail
site remodelé
site underground
site visité
site Web
site Web musical

site webzine
snail mail
spyware
start-up
start-upeur
surf anonyme
système de chiffrement
système de détection d'intrusion
Système de Nom de Domaine
techno-nerd
téléchargement
téléformation
télémaintenance
télétravail
télétraveilleur
temps de connexion
temps de téléchargement
thumbnail
Toile d'araignée mondiale
traducteur instantané
transfert asynchrone
Transmission Control Protocol/Internet Protocol
upload
user authentication
ver de courrier
vidéo hyperliée
video streaming
videochat
videostream
virus crypté
virus de boot
virus de courrier électronique
virus de programme
virus de script

Les noms composés anglais français et espagnols du domaine d'Internet :

virus e-mail
virus furtif
virus multipartite
virus polymorphique
virus troyen
visiteur unique
vitesse de connexion
vitesse de connexion Internet
vitesse de téléchargement
vitesse de transfert
voice IP
voix sur IP
waplockage
wappanaute
Web agency
web bug
Web collaborating
web designer
Web invisible
Web mobile
web radio
web sans fil
Web visible
web-bar
webcam
webcenter
weblog
weblogger
weblogging
webmail
webmarketeur
webmaster
webmestre
Web-remixeuse

webzine
white hat
Wireless Application Protocol
World Wide Web

3. Les noms composés espagnols

acceso a alta velocidad
acceso a Internet
acceso a Internet a alta velocidad
acceso a Internet por cable
acceso a la Red
acceso FTP
acceso permanente
acceso por cable a Internet
acceso remoto
acceso Web
actualización automática
actualización de virus
acuse de lectura
acuse de recibido
administrador de red
administrador Web
ADSL individual
ADSL multipuesto
adware
agenda de direcciones
agente inteligente
agujero de seguridad
alerta de correo
alerta de virus
alojamiento de sitio web
ancho de banda

animación frame a frame
antihacker
antipiratería
antispam
antivirus
aplicación gratuita
aplicación libre de todo coste
aplicación Web
applet Java
archivo adjunto
archivo de registro
archivo log
arquitectura cliente-servidor
artículo de Usenet
ataque de denegación de servicio
ataque de gusano
ataque de virus
ataque DOS
ataque fraggle
ataque hacker
ataque masivo
ataque smurf
ataque vírico
ataque web
autenticación de usuario
autopista de la información
autorización de acceso a Internet
banda ancha
banner de publicidad
banner publicitario
banner rotatorio
barra de direcciones
barra de navegación
barrido de puerto

barrido ping
bloqueador de pop-up
bloqueo de puerto
boletín de noticias
boletín de noticias electrónico
bookmark
buffer overflow
bulo electrónico
Business to Consumer
Business to Employee
búsqueda booleana
buzón de correo
buzón de correo POP3
buzón de mail
buzón electrónico
caballo de Troya
cable módem
caché de Internet
caché del browser
caché del navegador
café Internet
cámara Web
canal de chat
canal IRC
certificado de seguridad
charla de voz
chatbox
chequeador de email
ciber contratación
ciber supermercado
ciberactividad
ciberadicto
ciberálbum
ciberamor

Les noms composés anglais français et espagnols du domaine d'Internet :

ciberataque
cibercafé
cibercentro
cibercomerciante
cibercomprador
ciberconferencia
ciberconflicto
cibercrimen
cibercriminal
cibercriminalidad
ciberdebate
ciberdelincuencia
ciberdelincuente
ciberdelito
ciberderecho
ciberdirectorio
cibereditor
ciberespacio
ciberfrontera
ciberguerra
ciberlenguaje
ciberley
cibermayor
cibermercado
cibermundo
cibernauta
ciberocupa
ciberocupación
ciberpiratería
ciber-plebescito
ciberprotección
ciberpunk
cibersexo
cibersoldado

cibertendero
ciberterrorismo
ciberterrorista
cibertienda
ciberuniverso
cibervisitante
ciudadano de la red
clave de acceso
clave de cifrado
clave privada
clave pública
click and mortar
cliente de correo
cliente de mail
cliente de mensajería electrónica
cliente FTP
cliente IRC
código fuente
código malicioso
comercio electrónico
comparador de precios
compartimiento de ficheros
comunidad hacker
comunidad internauta
comunidad Internet
conexión a Internet
conexión ADSL multipuesto
conexión de alta velocidad
conexión de baja velocidad
conexión de línea dedicada
conexión de red
conexión dedicada
conexión inalámbrica
conexión inalámbrica a Internet

conexión permanente
conexión permanente de banda ancha
conexión por dial-up
conexión por línea comutada
conexión RTC
conexión servidor
conexión simultánea
contador de accesos
contador de visitas
contraseña en la sombra
control de acceso
cookie de Internet
cookie permanente
cookie temporal
copia oculta
correo basura
correo cifrado
correo electrónico
correo electrónico cifrado
correo electrónico comercial
correo electrónico junk
correo electrónico no deseado
correo indeseado
correo postal
correo vocal
correo web
cortafuegos
cortafuegos proxy
cuello de botella
cuenta de correo
cuenta de correo electrónico
cuenta de usuario
cuenta e-mail
cuenta POP

cyber-punk
definición de virus
denegación de servicio
Denial of Service
desarrollador Web
desbordamiento de bufer
detector de troyanos
dirección de e-mail
dirección electrónica
dirección Internet
dirección IP
dirección IP dinámica
dirección IP fija
dirección IP pública
dirección URL
dirección visitada
dirección Web
directorio de sitios
directorio temático
directorio Web
diseñador Web
diseño Web
documento anejo
Domain Name System
dot org
download
e-biz
ebook
e-business
e-card
e-commerce
e-compra
editor HTML
e-empleado

Les noms composés anglais français et espagnols du domaine d'Internet :

e-empleo
e-learning
e-mail
email comercial
e-market place
e-marketing
emoticono
empresa a empresa
empresa de alojamiento web
empresa de hosting
empresa de housing
empresa.com
encriptador de mensajes
enlace de hipertexto
espacio ftp
espacio Web
estructura de sitio
etiqueta Meta
e-turismo
explorador Web
fichero de firmas de virus
fichero FTP
fichero log
fichero MP3
fichero Napster
fichero script
File Transfer Protocol
filtrado de correo
filtro de mensajes
firewall
firewall hardware
firewall personal
firewall software
firma de virus

firma digital
firma electrónica
foro de debate
foro de discusión
forum de discusión
frame relay
freeware
Frequently Asked Questions
FTP anónimo
gateway
generador de virus
gestión de cookies
gestor de archivos
gestor de bookmarks
gestor de cookies
gestor de correo
gestor de descarga
gestor de mail
gestor Web
groupware
grupo de charla
grupo de contacto
grupo de discusión
grupo de noticias
grupo Usenet
gusano de correo
gusano vírico
hábitos de navegación
hacker blanco
hacker negro
hacktivismo
hacktivista
herramienta de creación de sitios
herramienta de Internet

hipervínculo
hoja de estilo
hoja de estilo en cascada
home page
home perso
hosting compartido
hotline
inbox
infección de virus
información alojada
integrador e-business
integrador HTML
intercambio de ficheros
interfaz cliente
interfaz de usuario
interfaz gráfica web
interfaz Web
internauta
Internet
Internet de alta velocidad
Internet fija
Internet móvil
Internet Protocol
intimidad on line
ISP local
jubilnauta
junk e-mail
kit de conexión
lector de news
Lenguaje de Marcado de Hipertexto
lenguaje DHTML
lenguaje HTML
lenguaje XML
libreta de direcciones

libro de visita
Línea de Subscripción Asimétrica Digital
lista de contacto
lista de correo
lista de discusión
lista de distribución
lista negra
mailbombing
mailto
marketing Web
medida de audiencia
mensaje cifrado
mensaje corto
mensaje de correo
mensaje de correo electrónico
mensaje de voz
mensaje instantáneo
mensaje Internet
mensaje pop-up
mensaje SMS
mensaje vocal
mensajería electrónica
mensajería instantánea
mensajería SMS
message board
meta tag
metabuscador
metaportal
módem
módem ADSL
módem analógico
módem de cable
módem dial-up
módem externo

Les noms composés anglais français et espagnols du domaine d'Internet :

módem RDSI
módem satelite
motor de búsquedas
motor de indexación
navegación anónima
navegación off-line
navegador de Internet
navegador fuera de línea
navegador offline
navegador web
negocio a negocio
negocio electrónico
netcam
Net-economía
netiqueta
newsletter
newsletter electrónica
no conectado
nombre de dominio
nombre de usuario
notificador de correo
openware
operador booleano
ordenador remoto
página alias
página de inicio
página de inicio personal
página hackeada
página HTML
página indexada
página inicial
página oficial
página perso
página personal

página visitada
página vista
página WAP
página Web
página Web estática
paquete de información
pasarela de Internet
password cracker
peer-to-peer
pirata informático
pop up
popunder
portal de portales
portal específico
portal generalista
portal Internet
portal multiacceso
portal temático
portal WAP
Preguntas más frecuentes
programa anti pop-up
programa de ataque
programa de chat
programa de cifrado
programa de compartido de fichero
programa de conexión a Internet
programa de correo
programa de e-commerce
programa de intercambio de ficheros
programa gratuito
programa multi-banner
programa spyware
protección anti-spam
Protocolo de Aplicación de Telefonía Inalámbrica

Protocolo de Transferencia de Ficheros
protocolo http
protocolo Internet
protocolo POP
protocolo punto a punto
protocolo SMTP
proveedor de acceso
proveedor de acceso a Internet
proveedor de alojamiento
proveedor de alojamiento web
proveedor de antivirus
proveedor de ASDL
proveedor de contenido
proveedor de correo
proveedor de espacio
proveedor de hosting
proveedor de Internet
proveedor de servicios
proveedor de servicios administrados
proveedor de servicios de aplicaciones
proveedor de servicios de internet
proveedor de servicios en línea
puerta trasera
puntocom
radio Web
recién conectado
red de área local
red de redes
red IP
red P2P
red peer- to-peer
red privada virtual
regla de filtrado
reventador de contraseña

robot de compras
rollover
router IP
sala de charla
sala de chat
salón de chat
script kiddie
seguridad en línea
servidor central
servidor compartido
servidor de chat
servidor de correo
servidor de correo entrante
servidor de correo saliente
servidor de DNS
servidor de dominio
servidor de Internet
servidor de mail
servidor de News
servidor de noticias
servidor dedicado
servidor FTP
servidor HTTP
servidor IRC
servidor POP
servidor proxy
servidor remoto
servidor seguro
servidor Web
sesión de charla
shadow password
shareware
shopbot
sistema de cifrado

Les noms composés anglais français et espagnols du domaine d'Internet :

sistema de detección de intrusión
sistema de detección de intrusos
sistema de nombre de dominio
sitio comercial
sitio de archivo
sitio de descarga
sitio de e-commerce
sitio de Internet
sitio hackeado
sitio HTML
sitio mirror
sitio MP3
sitio oficial
sitio personal
sitio portal
sitio underground
sitio Web
Small Office Home Office
snail mail
software compartido
software de acceso a Internet
software de filtrado
software de navegación
software de servidor
software espía
software libre
spyware
tablón de anuncios
tag meta
telaraña electrónica
telecentro
teletrabajador
teletrabajo
thumbnail

tiempo de conexión
tiempo de descarga
tienda on-line
traductor instantáneo
transferencia asincrónica
Transmission Control Protocol/Internet Protocol
velocidad de conexión
velocidad de conexión a Internet
velocidad de descarga
velocidad de transferencia
venta electrónica
ventana de chat
ventana emergente
ventana flotante
video chat
video hipervinculado
vídeo streaming
virus cifrado
virus de arranque
virus de correo electrónico
virus de programa
virus de script
virus multipartito
virus polimórfico
virus polimorfo
virus stealth
virus troyano
visitante único
voice over IP
voz sobre IP
web bug
Web invisible
web musical
Web personal

Web radio
Web visible
webcam
weblog
weblogger
webmaster
website
Wireless Application Protocol
World Wide Web

Annexe 8. Typologies morphosyntaxiques des noms composés du corpus

1. Les noms composés anglais

Tableau 105 : Les composés binaires anglais

¹⁹² Le chiffre figurant à la fin de chaque catégorie figurent deux chiffres (entre parenthèses) indique le nombre de composés créés par le patron syntaxique observé.

Les noms composés anglais français et espagnols du domaine d'Internet :

Cat.	Structures	Composés ¹⁹²
1	N + N	access control, ad-blocker, address bar, address book, ADSL modem, ADSL provider, alias page, antivirus provider, archive manager, archive moderator, attack program, audience measurement, bandwidth, banner ad, banner advertising, blocking access, bookmark, boot virus, bottleneck, browser cache, browser sniffer, browsing habit, bug hunter, cable modem, chat group, chat host, chat server, chat session, chat window, chatbox, chate-mate, chatroom, client interface, comparison site, connection kit, connection speed, connection time, contact group, contact list, cookie builder, cookie file, cookie ID, cookie management, cookie manager, data packet, discussion area, discussion board, discussion forum, discussion group, distribution list, DNS server, domain name, DoS attack, dot com, dot-org, emoticon, encryption key, encryption program, encryption system, file-sharing, file-swapping, filtering package, filtering rule, firewall, flamewar, forwarding address, fraggle attack, frame relay, FTP access, FTP client, FTP file, FTP server, FTP space, guestbook, hacker attack, hacker community, hacking attack, hacktivism, hacktivist, history folder, hit counter, home page, hosting company, hosting package, hosting URL, HTML editor, HTML integrator; HTML page, HTML site, HTTP server, indexing engine, Internet access, Internet cache, Internet café, Internet community, Internet connection, Internet cookie, Internet name, Internet portal, Internet Protocol, Internet server, Internet tool, Internet user, IP address, IP network, IP router, IRC channel, IRC client, IRC server, Java applet, keylogger, log file, log sniffer, mail bombing, mail checker, mail forwarding, mail server, mailbox, mailing list, message board, message filter, mirror site, modem, MP3 file, MP3 site, navigation bar, netizen, network administrator, network connection, newsgroup, newsgroupie, newsletter, P2P network, page view, parody site, password, ping sweep, POP account, POP server, port blocking, port scanning, portal site, program virus, proxy server, PSTN connection, read notification, receipt notification, rotation banner, satellite modem, script blocking, script file, script kiddie, script virus, scrolling banner, search engine, security hole, server attack, server connection, server package, shopbot, showroom, site directory, site owner, site structure, SMS messaging, smurf attack, snail mail, source code, space provider, spam protection, storage site, streaming video, style sheet, surfing habit, techno-nerd, thumbnail, transfer

¹⁹² Le chiffre figurant à la fin de chaque catégorie figurent deux chiffres (entre parenthèses) indique le nombre de composés créés par le patron syntaxique observé.

Annexe 8. Typologies morphosyntaxiques des noms composés du corpus

		speed, URL history, Usenet group, Usenet posting, user account, user authentication, user interface, username, video chat, video chatting, videotstream, virus alert, virus attack, virus definition, virus writer, Vmail, voice chat, voice IP, WAP locking, WAP page, WAP portal, wireless connection, worm alert, worm attack (201)
2	A + N	analog modem, anonymous FTP, anonymous surfing, asynchronous transfer, backdoor, black hat, blacklist, Boolean operator, Boolean search, central server, commercial site, dedicated server, digital signature, electronic commerce, electronic hoax, electronic message, electronic messaging, electronic signature, encrypted virus, executable attack, external modem, general portal, hacked page, hacked site, hosted data, hotline, indexed page, individual ADSL, instant message, instant messaging, instant messenger, instant translator, intelligent agent, Internet, local ISP, malicious code, massive attack, mobile Internet, multiaccess portal, multicriteria search, multipartite virus, official page, official site, permanent cookie, personal site, polymorphic virus, private key, public key, remodelled site, remote access, remote computer, remote server, secure certificate, secure server, shared hosting, shared server, simultaneous connection, specific portal, stealth virus, temporary cookie, thematic directory, thematic portal, Trojan Horse, Trojan virus, unique visitor, viral attack, viral code, viral infection, viral worm, virtual hosting, white hacker, white hat (72)
3	PS + N	cyber delinquent, cyber directory, cyber-grocer, cyber market, cyber monitoring, cyber shop, cyber visitor, cyber-activity, cyber-album, cyberattack, cyberbrand, cyber-café, cyberconference, cyber-conflict, cybergop, cybercrime, cybercriminal, cybercriminality, cyber-debate, cyber-delinquency, cybereditor, cyberethics, cyberflirt, cyberfrontier, cyberhacking, cyberlanguage, cyberlaw, cyberlove, cybermerchant, cyberprotection, cyberpunk, cyber-recruitment, cyber-referendum, cybersex, cybersoldier, cyberspace, cybersquatter, cybersquatting, cyberstress, cyberterrorist, cyberterrorism, cyberunivers, cyberworld, ebook, e-business, e-card, e-commerce, e-customer, e-employee, e-job, e-learning, email, e-purchase, e-tourism, e-trade, e-trader, hyperlink, meta portal, meta tag, Net access, Net connection, Net economy, Net marketing, Net user, teleworking, Web accelerator, Web access, Web address, Web agency, Web attack, Web browser, Web bug, Web cobrowsing, Web collaborating, Web community, Web design, Web designer, Web developer, Web directory, Web hosting, Web manager, Web page, Web radio, Web remix, Web remixer, Web server, Web space, Web surfer, webcam,

Les noms composés anglais français et espagnols du domaine d'Internet :

		webcenter, weblog, weblogger, weblogging, webmail, webmaster, Website (96)
4	N + PS	groupware, shareware, spyware, wireless Web (4)
5	A + PS	freeware, openware, invisible web, visible web (4)
6	in + N	inbox (1)
7	up + N	uplink (1)
8	N + up	pop up, startup (2)
9	N+ to	mailto (1)
10	N + over	rollover (1)
11	down + V	download, downloading (2)
12	up + V	upload (1)
13	N+to+N	business to business, business-to-consumer, business to employee, peer-to-peer (4)
14	N+and+N	click and mortar (1)
15	N+of+N	Denial of Service, network of networks, portal of portals (3)
16	Flud	Napsterized file, e-biz, Net marketer, Web marketer (4)

Tableau 106 : Les surcomposés anglais à trois éléments

Cat	Structures	Surcomposés
1	N + N + N	ad-blocking program, Application Service Provider, bookmark manager, Bulletin Board System, cable Internet access, Cascading Style Sheet, client-server architecture, Domain Name System, dot-com industry, File Transfer Protocol, file-sharing program, file-swapping program, Internet access authorization, Internet Access Provider, Internet access package, Internet content provider, Internet gateway, Internet Service Provider, Internet traffic load, intrusion detection system, network firewall, newsgroup server, password cracker, POP3 mail box, Post Office Protocol, price comparison site, proxy firewall, search engine spider, shadow password, site creation tool, user interface expert, virus-signature file, wireless Internet connection, Wireless Application Protocol (34)
2	A + N + N	blind carbon copy, broadband Internet, collocated server package, dynamic IP address, electronic messaging client, electronic newsletter, electronic mail account, high-speed access, high-speed connection, high-speed Internet, leased line connection, Local Area Network, low-speed connection, Managed Service Provider, open source app, personal home page, public IP address, shared hosting package, shared IP address, shared mail server, shared server package, Short Message Service, static IP address, Trojan horse detector, Secure Socket Layer, Uniform Resource Locator (26)
3	A + A + N	private virtual network (1)
4	on + N + N	online privacy, online security, online shop, online shopping (4)
5	off + N + N	offline browser, offline browsing (2)
6	under + N + N	underground site (1)
7	N + A + N	ad-free zone, World Wide Web, chat software (3)
8	N + up + N	dial-up connection, dial-up modem, live update, pop-up message, pop-up ad, virus update (6)
9	N + down + N	pop-down ad (1)
10	N + under + N	pop-under ad (1)
11	PS + A + N	cybersupermarket, hyperlinked video, Web software (3)
12	PS + N + N	e-business integrator, e-commerce package, e-commerce site, e-mail account, e-mail address, email alert, e-mail attachment, e-mail client, email encryptor, e-mail filtering, e-mail header, email list, email manager, email provider, email server, email virus, email worm, e-market place, e-zine site meta search engine, Net congestion message, Net connection speed, web user interface, Web hosting company, Web surfing habit (25)
13	A + PS + N	bulk email, encrypted email, junk email, musical website, static

Les noms composés anglais français et espagnols du domaine d'Internet :

		Web page (5)
14	Adv + V + N	download manager, download speed, download site, download time (4)
15	Adv + A + N	Frequently Asked Questions (1)
16	N+ PS + N	voice email (1)
17	Adv + on + N	always-on access, always-on connection (2)
18	N + of + N + N	Denial of Service attack (1)
29	N + to + N + N	frame-to-frame animation, peer-to-peer network, Point to Point protocol, server-to-server architecture (4)
20	N + and + N + N	click-and-mortar business (1)

Tableau 107: Les surcomposés anglais à quatre éléments

Cat.	Structures	Surcomposés
1	N + over + N + N	buffer overflow attack (1)
2	N + N + N + up	dotcom start up (1)
3	on + N + N + N	Online Service Provider (1)
4	in + N + N + N	incoming mail server (1)
5	out + N + N + N	outgoing mail server (1)
6	N + N + A + N	Internet connection software, newsgroup software (2)
7	A + A + N + N	Asymmetric Digital Subscriber Line, human-assisted search engine (2)
8	Adv + on + A + N	always-on routed Internet (1)
9	N + A + A + N	information superhighway (1)
10	A + N + PS + N	high-speed Web access (1)
11	PS + N + N + N	e-mail newsletter, HyperText Transfer Protocol (2)
12	N + PS + N + N	POP3 email box (1)
13	PS + N + A + N	email software (1)
14	A + N + N + N	Hardware firewall, Simple Mail Transfer Protocol, Small Office Home Office, software firewall (4)
15	A + N + up + N	eXtensible Markup Language (1)

Tableau 108 : Les surcomposés à cinq et six éléments

Cat.	Structures	Surcomposés
Cinq éléments		
1	Adv + on + A + N + N	always-on broadband connection (2)
2	N + N + N + N + N	Transmission Control Protocol/Internet Protocol (1)
3	PS + N + N + up + N	HyperText Markup Language (1)
4	N + N + V + N + N	Challenge Handshake Authentication Protocol (1)
5	N + N + N + A + N	mailing-list management software, newsgroup software (2)
Six éléments		
1	A+ PS + N + N + up + N	Dynamic HyperText Markup Language (1)

2. Les noms composés français

Tableau 109 : Les composés binaires français

¹⁹³ Les noms composés soulignés sont des composés endogènes hybrides. A la fin de chaque catégorie figurent deux chiffres (entre parenthèses). Le premier indique le nombre de composés créés par le patron syntaxique observé et le deuxième, souligné, correspond au nombre de composés endogènes hybrides.

Cat.	Structures	Composés ¹⁹³
1	Composés endogènes purs et composés endogènes hybrides	
1.1	N + de + N	accusé de lecture, accusé de réception, administrateur de site, adresse d'hébergement, alerte de virus, annuaire de sites, aspirateur de site, <u>attaque de hacker</u> , attaque de piratage, attaque de serveur, attaque de ver, attaque de virus, autorité de certification, autoroute de l'information, barre d'adresses, barre de navigation, blocage de port, <u>bloqueur de [popup]</u> , boîte de réception, cache du navigateur, <u>canal de chat</u> , carnet d'adresses, <u>CD de connexion</u> , certificat de sécurité, <u>chasseur de bug</u> , cheval de Troie, <u>chiffrement d'[e-mail]</u> , clé de chiffrement, <u>client de chat</u> , client de courrier, comparateur de prix, compte de courrier, compte de messagerie, compteur d'accès, compteur de visite, contrôle d'accès, définition de virus, Déni de Service, diffuseur de virus, échange de fichiers, en-tête de courrier, espace de discussion, faille de sécurité, <u>fenêtre de chat</u> , feuille de style, <u>filtrage d'[e-mails]</u> , filtre de messages, forum de discussion, <u>fouineur de log</u> , <u>fournisseur d'[e-mail]</u> , fournisseur d'accès, fournisseur de contenu, fournisseur de courrier, fournisseur d'espace, fournisseur d'hébergement, <u>générateur de cookies</u> , générateur de virus, <u>gestion de [mailing list]</u> , <u>gestion de cookies</u> , gestionnaire d'archives, <u>gestionnaire de cookies</u> , gestionnaire de favoris, goulot d'étranglement, <u>groupe de chat</u> , groupe de contact, groupe de discussion, <u>groupe de News</u> , habitude de navigation, hébergement de site, <u>hébergeur de chat</u> , journal de connexion, <u>kit de connexion</u> , liste de contact, liste de diffusion, liste de distribution, livre d'or, <u>log de connexion</u> , <u>logiciel de chat</u> , logiciel de chiffrement, logiciel de courrier, logiciel de filtrage, logiciel de navigation, <u>logiciel de news</u> , logiciel de serveur, logiciel d'hébergement, message de congestion, mesure d'audience, mot de passe, moteur de recherche, moteur d'indexation, nom de domaine, nom d'utilisateur, notificateur de mail, ossature de site, outil de recherche, page d'accueil, page de démarrage, paquet d'information, partage de fichier(s), portail de portails, porte de derrière, porte de service, <u>programme d'[e-commerce]</u> , programme d'attaque, propriétaire de site, rapatriement de courrier, règle de filtrage, renifleur de navigateur, réseau des réseaux, <u>responsable de l'avenir</u> , <u>salon de chat</u> , salon de

¹⁹³

Les noms composés soulignés sont des composés endogènes hybrides. A la fin de chaque catégorie figurent deux chiffres (entre parenthèses). Le premier indique le nombre de composés créés par le patron syntaxique observé et le deuxième, souligné, correspond au nombre de composés endogènes hybrides.

Annexe 8. Typologies morphosyntaxiques des noms composés du corpus

		discussion, <u>scan de port</u> , <u>serveur de [newsgroup]</u> , <u>serveur de chat</u> , serveur de courrier, <u>serveur de News</u> , <u>session de chat</u> , signature de virus, <u>site d[e-commerce]</u> , site d'archivage, site de comparaison, site de parodie, système de chiffrement, temps de connexion, ver de courrier, <u>virus de boot</u> , virus de programme, <u>virus de script</u> , vitesse de connexion, vitesse de transfert (131 / 32)
1.2	N + à + N	accès à distance, boîte aux lettres, Foire Aux Questions (3)
1.3	N + de + PS	<u>citoyen du Net</u> , <u>communauté du Web</u> (2 / 2)
1.4	N + A	accès distant, accès permanent, adresse électronique, <u>ADSL individuel</u> , agent intelligent, annuaire thématique, attaque massive, attaque virale, babillard électronique, bande passante, bandeau publicitaire, bannière défilante, bannière publicitaire, bannière rotative, canular électronique, <u>chat vocal</u> , clé privée, clé publique, code malveillant, code viral, commerce électronique, connexion dédiée, connexion permanente, connexion simultanée, <u>cookie permanent</u> , <u>cookie temporaire</u> , courrier électronique, courrier postal, donnée hébergée, <u>[e-mail] chiffré</u> , <u>[e-mail] crypté</u> , envoi groupé, fenêtre publicitaire, FSI local, <u>FTP anonyme</u> , <u>hacker blanc</u> , <u>hacker noir</u> , hébergement partagé, hébergement virtuel, infection virale, <u>Internet mobile</u> , <u>Internet nomade</u> , <u>ISP local</u> , lien consulté, liste noire, logiciel gratuit, logiciel libre, logiciel partagé, <u>mail vocal</u> , message électronique, message instantané, message vocal, messagerie électronique, messagerie instantanée, <u>modem analogique</u> , <u>modem commuté</u> , <u>modem externe</u> , navigation anonyme, <u>[newsletter] électronique</u> , opérateur booléen, ordinateur distant, page indexée, page officielle, page perso, page personnelle, page piratée, page visitée, page vue, pièce jointe, pirate informatique, portail généraliste, portail multi-accès, portail spécifique, portail thématique, porte dérobée, recherche multicritère, requête booléenne, réseau local, serveur central, serveur dédié, serveur distant, serveur partagé, serveur sécurisé, signature électronique, signature numérique, site mercantile, site officiel, site parodique, site perso, site personnel, site piraté, site remodelé, site visité, surf anonyme, traducteur instantané, transfert asynchrone, virus crypté, virus furtif, virus multipartite, virus polymorphique, virus troyen, visiteur unique (102 / 17)
1.5	A + N	haut débit, jeune pousse (2)
		<u>protection anti-spam</u> , <u>accès FTP</u> , <u>accès Internet</u> , administrateur réseau, <u>adresse [e-mail]</u> , <u>adresse Internet</u> , <u>adresse IP</u> , <u>adresse URL</u> , <u>alerte [e-mail]</u> , <u>applet Java</u> , apprenti pirate, <u>article Usenet</u> , <u>aspirateur Internet</u> , <u>attaque</u>

Les noms composés anglais français et espagnols du domaine d'Internet :

1.6	N + N	[Denial of Service], attaque DoS, attaque fraggle, attaque smurf, cache Internet, café Internet, canal IRC, client [e-mail], client FTP, client IRC, code source, communauté hacker, communauté Internet, compte [e-mail], compte pop, compte utilisateur, connexion Internet, connexion réseau, connexion RTC, connexion serveur, cookie Internet, courrier escargot, éditeur HTML, émoticône, espace FTP, fenêtre pop-under, fenêtre pop-up, fichier espion, fichier FTP, fichier journal, fichier log, fichier MP3, fichier Napster, filtre anti-pub, firewall logiciel, fournisseur ADSL, fournisseur d'antivirus, groupe Usenet, hacktivisme, hacktiviste, intégrateur [e-business], intégrateur HTML, interface client, interface utilisateur, langage DHTML, langage HTML, langage XML, logiciel anti [popup], logiciel anti-pub, logiciel spyware, logiciel espion, message pop-up, message Internet, messagerie SMS, modem ADSL, modem câble, modem RTC, modem satellite, navigation off-line, [newsletter] [e-mail], outil Internet, pack ADSL, page alias, page HTML, page WAP, portail WAP, passerelle Internet, pixel espion, point com, portail Internet, programme [open source], programme espion, protocole http, protocole Internet, protocole pop, protocole PPP, protocole SMTP, protocole TCP/IP, réseau peer-to-peer, réseau IP, réseau P2P, robot logiciel, routeur IP, serveur DNS, serveur FTP, serveur HTTP, serveur Internet, serveur IRC, serveur mail, serveur mandataire, serveur POP, serveur proxy, site e-zine, site underground, site webzine, site filleul, site HTML, site Internet, site marchand, site mirror, site MP3, site parrain, site portail, virus e-mail (117 / 93)
1.7	PS + N	cyber amour, cyber univers, cyberacheteur, cyberactivité, cyberalbum, cyberannuaire, cyberattaque, cyberboutique, cybercafé, cybercaméra, cybergévison, cyber-commerçant, cyberconférence, cyberconflit, cyberconsommateur, cybercrime, cybercriminalité, cybercriminel, cyberdébat, cyber-délinquance, cyberdélinquant, cyberdroit, cyberéditeur, cyberépicier, cyberespace, cyberflic, cyber-flirt, cyberfrontière, cyberguerre, cyberlangue, cybermarché, cybermarque, cybermonde, cyber-piraterie, cyber-plébiscite, cyberprotection, cyber-punk, cyber-recrutement, cybersexe, cybersoldat, cyberstress, cyber-supermarché, cybersurveillance, cyberterrorisme, cyberterroriste, cybervisiteur, e-achat, e-commerce, e-courrier, e-économie, e-employé, e-métier, e-tourisme, hyperlien, metabalise, métamoteur, metaportal, Net économie, néétique, netsurfeur, téléchargement, téléformation, télémaintenance,

		télétravail, télétraveilleur, <u>web-bar</u> , <u>Web-remixeur</u> (67 / 62)
1.8	N + PS	accès Web, administrateur Web, adresse Web, annuaire Web, aspirateur Web, attaque Web, balise meta, concepteur Web, design Web, designer Web, développeur Web, entreprenaute, ergonome Web, espace Web, espiongiciel, gestionnaire Web, interface Web, internaute, logiciel Web, page Web, serveur web, site Web (22 / 3)
1.9	PS + A	Web mobile, Web invisible, Web visible (3 / 3)
1.10	V + N	pare-feu (1)
1.11	non + N	non-connecté (1)
1.12	Flud	e-picerie, e-xagone, hotliner, netmarketeur, site hacké, start-upeur, waplockage, wappanaute, webmestre, webmarketeur (10 / 6)
2	Composés exogènes	
2.1	Emp2	adware, backdoor, black hat, blacklist, bookmark, buffer overflow, Business to Business, Business to Consumer, Business to Employee, chat room, chatbox, click and mortar, contact list, cybersquatter, cybersquatting, dot org, dotcom, download, e-biz, ebook, e-business, e-card, e-learning, e-mail, firewall, frame relay, freeware, groupware, home page, hotline, inbox, Internet, key/logger, mail bombing, mailing list, mailto, message board, meta tag, modem, netmarketing, netsurfer, newsgroup, newsletter, open source, openware, Peer to Peer, ping sweep, pop under, pop-up, rollover, script kiddie, shareware, shopbot, showroom, snail mail, spyware, start-up, techno-nerd, thumbnail, upload, user authentication, video streaming, videochat, videotream, voice IP, web agency, web bug, web collaborating, web designer, web radio, webcam, webcenter, weblog, weblogger, weblogging, webmail, webmaster, webzine, white hat (79)

Tableau 110: Les surcomposés français à trois éléments

Les noms composés anglais français et espagnols du domaine d'Internet :

Cat.	Structures	Surcomposés
1		Surcomposés endogènes purs et surcomposés endogènes hybrides
1.1	N+de+N+A	adresse de courrier électronique, client de courrier électronique, client de messagerie électronique, compte de courrier électronique, fournisseur d'applications hébergées, fournisseur de courrier électronique, fournisseur de services gérés, gestionnaire de courrier électronique, <u>package d'hébergement partagé</u> , page d'accueil perso, page d'accueil personnelle, serveur de courrier entrant, serveur de courrier sortant, Toile d'araignée mondiale, virus de courrier électronique (15 / 1)
1.2	N+à+N+A	boîte aux lettres électronique, mise à jour automatique, moteur à assistance humaine, site à assistance humaine, connexion à faible débit (5)
1.3	PS+N+de+N	<u>e-place de marché</u> (1 / 1)
1.4	N+de+PS+N	assistant de téléchargement, gestionnaire de téléchargement, site de téléchargement, temps de téléchargement, vitesse de téléchargement (5)
1.5	N+de+N+PS	fournisseur d'hébergement Web (1 / 1)
1.6	N+de+N+N	fournisseur d'accès Internet, <u>fournisseur de contenu Internet</u> , <u>fournisseur de services Internet</u> , <u>logiciel d'accès Internet</u> , <u>programme d'accès Internet</u> , <u>logiciel de connexion Internet</u> , <u>vitesse de connexion Internet</u> (7 / 7)
1.7	N+à+N+N	boîte aux lettres POP3, mise à jour antivirus (2 / 1)
1.8	N+N+par+N	<u>accès Internet par câble</u> , <u>animation frame par frame</u> (2 / 2)
1.9	N+A+de+N	serveur partagé de courrier (1)
1.10	N+de+N+en+N	feuille de style en cascade, fournisseur de services en ligne (2)
1.11	N+de+N+à+N	<u>autorisation d'accès à Internet</u> (1)
1.12	N+de+N+de+N	logiciel de cryptage de courrier, programme de partage de fichier, outil de création de site, programme d'échange de fichier, Protocole de Transfert de Fichier, Système de Nom de Domaine, site de comparaison de prix, détecteur de chevaux de Troie, système de détection d'intrusion, <u>cracker de mot de passe</u> , araignée de moteur de recherche, fichier de signature de virus (12 / 1)
1.13	N+N+A	copie carbone invisible, <u>adresse IP dynamique</u> , <u>adresse IP fixe</u> , <u>adresse IP publique</u> , <u>adresse IP partagée</u> , <u>Internet bande large</u> (6 / 5)
1.14	N+A+N	accès haut débit, connexion haut débit, <u>Internet haut débit</u> (3 / 1)

1.15	PS+N+N	<u>cyberpunkitude</u> (1 / 1)
1.16	N+PS+N	<u>lien hypertexte</u> (1 / 1)
1.17	N+PS+A	<u>vidéo hyperliée</u> , <u>page Web statique</u> , <u>site Web musical</u> (3 / 3)
1.18	N+A+PS	<u>interface graphique web</u> (1)
1.19	PS+sans+N	<u>Web sans fil</u> (1)
1.20	N sur N	<u>voix sur IP</u> (1)
1.21	V+N+N	<u>pare-feu réseau</u> , <u>pare-feu proxy</u> (2 / 1)
1.22	N+sans+N	<u>connexion sans fil</u> (1)
1.23	N+hors+N	<u>navigation hors-ligne</u> (1)
1.24	N+non+A	courrier non-sollicité, message non-sollicité, publicité non-sollicitée, <u>e-mail non sollicité</u> (4 / 1)
1.25	N + en + N	apprentissage en ligne, dialogue en direct, achat en ligne, intimité en ligne, magasin en ligne, sécurité en ligne, shopping en ligne, <u>connexion en dialup</u> (8 / 1)
1.26	N+N+N	architecture client/serveur, architecture serveur-serveur, <u>attaque buffer overflow</u> , <u>entreprise point com</u> (4 / 2)
1.27	N+A+A	courrier électronique chiffré, réseau privé virtuel (2)
2	Surcomposés exogènes	
	Emp3	Frequently Asked Questions, Short Message Service, World Wide Web, File Transfer Protocol, Wireless Application Protocol (5)

Tableau 111 : Les surcomposés français à quatre et cinq éléments

Cat.	Structures	Surcomposés
1	Surcomposés endogènes purs et surcomposés endogènes hybrides	
1.1	N + de + N + A + A	ligne d'abonnement numérique asymétrique (1)
1.2	N + à + N + A + A	connexion à bande large permanente (1)
1.3	N + de + N + sans +N	protocole d'applications sans fil (1)
1.4	N + PS + A + N	<u>accès web haut débit</u> (1 / 1)
1.5	N + N + sans + N	<u>connexion Internet sans fil</u> (1 / 1)
1.6	N + par + N + A	connexion par ligne commutée (1)
2	Surcomposés exogènes	
2.1	Emp4	Asymmetric Digital Subscriber Line, HyperText Transfer Protocol, eXtensible Markup Language (3)
2.2	Emp5	HyperText Markup Language, Transmission Control Protocol/Internet Protocol (2)

3. Les noms composés espagnols

Les noms composés anglais français et espagnols du domaine d'Internet :

Tableau 112 : Les composés binaires espagnols

Cat	Structures	Composés
1		Composés endogènes purs et composés endogènes hybrides
1.1	N + A	acceso permanente, acceso remoto, actualización automática, <u>ADSL individual</u> , <u>ADSL multipuesto</u> , agente inteligente, aplicación gratuita, archivo adjunto, ataque masivo, ataque vírico, banda ancha, <u>banner publicitario</u> , <u>banner rotatorio</u> , bulo electrónico, búsqueda booleana, buzón electrónico, clave privada, clave pública, código malicioso, comercio electrónico, conexión dedicada, conexión inalámbrica, conexión permanente, conexión simultánea, <u>cookie permanente</u> , <u>cookie temporal</u> , copia oculta, correo cifrado, correo electrónico, correo postal, correo vocal, dirección electrónica, dirección visitada, directorio temático, documento anexo, <u>e-mail comercial</u> , firma digital, firma electrónica, <u>FTP anónimo</u> , <u>ISP local</u> , gusano vírico, <u>hacker blanco</u> , <u>hacker negro</u> , <u>home perso</u> , <u>site anejo</u> , <u>hosting compartido</u> , información alojada, <u>Internet fija</u> , <u>Internet móvil</u> , lista negra, mensaje cifrado, mensaje corto, mensaje instantáneo, mensaje vocal, mensajería electrónica, mensajería instantánea, <u>módem analógico</u> , <u>módem externo</u> , navegación anónima, negocio electrónico, operador booleano, ordenador remoto, página indexada, página inicial, página oficial, página perso, página personal, página visitada, página vista, pirata informático, portal específico, portal generalista, portal multiacceso, portal temático, programa gratuito, puerta trasera, correo indeseado, servidor central, servidor compartido, servidor dedicado, servidor remoto, servidor seguro, <u>sitio underground</u> , sitio comercial, sitio oficial, sitio personal, <u>software compartido</u> , <u>software libre</u> , traductor instantáneo, transferencia asincrónica, venta electrónica, ventana emergente, ventana flotante, virus cifrado, virus multipartito, virus polimórfico, virus polimorfo, <u>virus stealth</u> , virus troyano, visitante único (100 / 22)
1.2	A + N	recién conectado (1)
1.3	N + N	acceso FTP, <u>applet Java</u> , <u>archivo log</u> , <u>ataque DoS</u> , <u>ataque fraggle</u> , <u>ataque hacker</u> , <u>ataque smurf</u> , <u>barrido ping</u> , <u>café Internet</u> , <u>canal IRC</u> , <u>cliente FTP</u> , <u>cliente IRC</u> , código fuente, comunidad hacker, comunidad Internet, conexión RTC, conexión servidor, correo basura, <u>cuenta e-mail</u> , <u>cuenta POP</u> , dirección Internet, dirección IP, dirección URL, <u>editor HTML</u> , emotícono, <u>empresa.com</u> , espacio ftp, <u>fichero FTP</u> , <u>fichero log</u> , <u>fichero MP3</u> , <u>fichero Napster</u> , <u>fichero script</u> , <u>firewall software</u> , <u>firewall hardware</u> , grupo Usenet, <u>hacktismo</u> , <u>hacktivista</u> , <u>integrador e-business</u> , <u>integrador HTML</u> , interfaz cliente, <u>intimidad on line</u> , <u>lenguaje DHTML</u> , <u>lenguaje HTML</u> ,

		<u>lenguaje XML</u> , <u>mensaje pop-up</u> , <u>mensaje Internet</u> , <u>mensaje SMS</u> , <u>mensajería SMS</u> , <u>módem dial-up</u> , <u>módem ADSL</u> , <u>módem RDSL</u> , <u>módem satélite</u> , <u>navegación off-line</u> , <u>navegador offline</u> , <u>página alias</u> , <u>página HTML</u> , <u>página WAP</u> , <u>portal Internet</u> , <u>portal WAP</u> , <u>programa anti pop-up</u> , <u>programa spyware</u> , <u>programa multi-banner</u> , <u>protección anti-spam</u> , <u>protocolo http</u> , <u>protocolo Internet</u> , <u>protocolo POP</u> , <u>protocolo SMTP</u> , <u>puntocom</u> , <u>red peer-to-peer</u> , <u>red IP</u> , <u>red P2P</u> , <u>servidor FTP</u> , <u>servidor HTTP</u> , <u>servidor IRC</u> , <u>servidor POP</u> , <u>servidor proxy</u> , <u>site FTP</u> , <u>site WAP</u> , <u>sitio HTML</u> , <u>sitio mirror</u> , <u>sitio MP3</u> , <u>sitio portal</u> , <u>software espia</u> , <u>tag HTML</u> , <u>tienda on-line</u> (85 / 78)
1.4	PS + N	<u>ciber contratación</u> , <u>ciber supermercado</u> , <u>ciberactividad</u> , <u>ciberadicto</u> , <u>ciberálbum</u> , <u>ciberamor</u> , <u>ciberataque</u> , <u>cibercafé</u> , <u>cibercentro</u> , <u>cibercomerciante</u> , <u>cibercomprador</u> , <u>ciberconferencia</u> , <u>ciberconflicto</u> , <u>cibercrimen</u> , <u>cibercriminal</u> , <u>cibercriminalidad</u> , <u>ciberdebate</u> , <u>ciberdelincuencia</u> , <u>ciberdelincuente</u> , <u>ciberdelito</u> , <u>ciberderecho</u> , <u>ciberdirectorio</u> , <u>cibereditor</u> , <u>ciberespacio</u> , <u>ciberfrontera</u> , <u>ciberguerra</u> , <u>ciberlenguaje</u> , <u>ciberley</u> , <u>cibermayor</u> , <u>cibermercado</u> , <u>cibermundo</u> , <u>cibernauta</u> , <u>ciberocupa</u> , <u>ciberocupación</u> , <u>ciberpiratería</u> , <u>ciber-plebiscito</u> , <u>ciberporteción</u> , <u>cibersexo</u> , <u>cibersoldado</u> , <u>cibertenedor</u> , <u>ciberterrorismo</u> , <u>ciberterrorista</u> , <u>cibertienda</u> , <u>ciberuniverso</u> , <u>cibervisitante</u> , <u>e-compra</u> , <u>e-empleado</u> , <u>e-empleo</u> , <u>e-turismo</u> , <u>hipervínculo</u> , <u>metabuscador</u> , <u>metaportal</u> , <u>Net-economía</u> , <u>netiqueta</u> , <u>teletrabajador</u> , <u>teletrabajo</u> (56 /54)
1.5	PS + A	<u>web musical</u> , <u>web personal</u> , <u>web invisible</u> , <u>web visible</u> (4 / 4)
1.6	N + PS	<u>acceso web</u> , <u>administrador web</u> , <u>aplicación web</u> , <u>ataque web</u> , <u>cámara web</u> , <u>correo web</u> , <u>desarrollador web</u> , <u>dirección web</u> , <u>directorio web</u> , <u>diseñador web</u> , <u>diseño web</u> , <u>espacio web</u> , <u>etiqueta Meta</u> , <u>explorador web</u> , <u>gestor web</u> , <u>interfaz web</u> , <u>internauta</u> , <u>jubilnauta</u> , <u>marketing web</u> , <u>navegador web</u> , <u>página web</u> , <u>radio web</u> , <u>servidor web</u> , <u>sitio web</u> , <u>tag meta</u> (25 /25)
1.7	no + N	no conectado (1)
1.8	V + N	cortafuegos (1)
		actualización de virus, acuse de lectura, acuse de recibido, administrador de red, agenda de direcciones, agujero de seguridad, alerta de correo, alerta de virus, ancho de banda, archivo de registro, <u>artículo de Usenet</u> , ataque de gusano, ataque de virus, autenticación de usuario, autopista de la información, <u>banner de publicidad</u> , barra de direcciones, barra de navegación, barrido de puerto, <u>bloqueador de pop-up</u> , bloqueo de puerto, boletín de noticias, buzón de correo, <u>buzón de mail</u> , caballo de Troya, <u>caché de Internet</u> , <u>caché del browser</u> , caché del navegador, <u>canal de chat</u> ,

Annexe 8. Typologies morphosyntaxiques des noms composés du corpus

1.9	N + de + N	<p>certificado de seguridad, charla de voz, ciudadano de la red, clave de acceso, clave de cifrado, cliente de correo, <u>cliente de mail</u>, comparador de precios, compartimiento de ficheros, conexión de red, contador de accesos, contador de visitas, control de acceso, <u>cookie de Internet</u>, cuello de botella, cuenta de correo, cuenta de usuario, definición de virus, denegación de servicio, <u>desbordamiento de bufer</u>, detector de troyanos, <u>dirección de email</u>, directorio de sitios, <u>empresa de hosting</u>, <u>empresa de housing</u>, encriptador de mensajes, estructura de sitio, filtrado de correo, filtro de mensajes, firma de virus, foro de debate, foro de discusión, forum de discusión, generador de virus, <u>gestión de cookies</u>, gestor de archivos, <u>gestor de bookmarks</u>, <u>gestor de cookies</u>, gestor de correo, gestor de descarga, <u>gestor de mail</u>, grupo de charla, grupo de contacto, grupo de discusión, grupo de noticias, gusano de correo, hábito de navegación, <u>herramienta de Internet</u>, hoja de estilo, infección de virus, intercambio de ficheros, interfaz de usuario, <u>kit de conexión</u>, <u>lector de news</u>, libreta de direcciones, libro de visita, lista de contacto, lista de correo, lista de discusión, lista de distribución, medida de audiencia, mensaje de correo, mensaje de voz, <u>módem de cable</u>, motor de búsqueda, motor de indexación, <u>navegador de Internet</u>, nombre de dominio, nombre de usuario, notificador de correo, página de inicio, paquete de información, <u>pasarela de Internet</u>, portal de portales, <u>programa de e-commerce</u>, programa de ataque, <u>programa de chat</u>, programa de cifrado, programa de correo, proveedor de acceso, <u>proveedor de ADSL</u>, proveedor de alojamiento, proveedor de antivirus, proveedor de contenido, proveedor de correo, proveedor de espacio, <u>proveedor de hosting</u>, <u>proveedor de Internet</u>, proveedor de servicios, red de redes, regla de filtrado, reventador de contraseña, robot de compras, sala de charla, <u>salón de chat</u>, <u>servidor de chat</u>, servidor de correo, <u>servidor de DNS</u>, servidor de dominio, <u>servidor de Internet</u>, <u>servidor de mail</u>, <u>servidor de News</u>, servidor de noticias, sesión de charla, sistema de cifrado, <u>sitio de e-commerce</u>, sitio de archivo, sitio de descarga, <u>sitio de Internet</u>, <u>software de conexión</u>, <u>software de filtrado</u>, <u>software de servidor</u>, tablón de anuncios, tiempo de conexión, tiempo de descarga, velocidad de conexión, velocidad de descarga, velocidad de transferencia, <u>ventana de chat</u>, virus de arranque, virus de programa, <u>virus de script</u> (152 / 41)</p>
1.10	N + a + N	acceso a la Red, <u>acceso a Internet</u> , <u>conexión a Internet</u> , empresa a empresa, negocio a negocio (5 / 2)

Les noms composés anglais français et espagnols du domaine d'Internet :

1.11	N + en + N	contraseña en la sombra (1)
1.12	Fdiv	<u>chequeador de e-mail</u> , <u>página hackeada</u> , <u>sitio hackeado</u> (3 / 3)
2	Composés exogènes	
2.1	Emp2	adware, bookmark, buffer overflow, Business to Consumer, Business-to-Employee, cable modem, chatbox, click and mortar, cyber-punk, Denial of Service, dot org, download, e-biz, ebook, e-business, e-card,e-commerce, e-learning, e-mail, e-marketing, firewall, frame relay, freeware, gateway, groupware, home page, hotline, inbox, Internet, mailbombing, mailto, message board, meta tag, modem, netcam, newsletter, openware, password, peer-to-peer, popunder, pop-up, rollover, router IP, script kiddie, shareware, shopbot, snail mail, spyware, thumbnail, video chat, video streaming, web bug, web radio, webcam, weblog,weblogger, webmaster, website (58)

Tableau 113 : Les surcomposés espagnols à trois éléments

Cat	Structures	Surcomposés
1	Surcomposés endogènes purs et surcomposés endogènes hybrides	
1.1	N + N + A	conexión ADSL multipuesto, dirección IP pública, dirección IP dinámica, dirección IP fija, firewall personal, newsletter electrónica, telaraña electrónica (7 /6)
1.2	N + A + A	correo electrónico comercial, correo electrónico cifrado, correo electrónico junk, red privada virtual (4 / 1)
1.3	N + A + PS	interfaz gráfica web (1 /1)
1.4	N + N + PS	comunidad internauta (1 / 1)
1.5	N +PS + A	página web estática, video hipervínculado (2 /2)
1.6	V + N + N	cortafuegos proxy (1 / 1)
1.7	N + sobre + N	voz sobre IP (1 / 1)
1.8	N + en + N	seguridad en línea (1)
1.9	N + más + A	preguntas más frecuentes (1)
1.10	N + N + N	arquitectura cliente-servidor (1)
1.11	N + de + N + A	cliente de mensajería electrónica, conexión de línea dedicada, cuenta de correo electrónico, mensaje de correo electrónico, página de inicio personal, proveedor de servicios administrados, red de área local, servidor de correo entrante, servidor de correo saliente, virus de correo electrónico (10)
1.12	N + a + A + N	acceso a alta velocidad (1)
1.13	N + de + A + N	conexión de alta velocidad, conexión de baja velocidad, Internet de alta velocidad (3 / 1)
1.14	N + de + PS + N	enlace de hipertexto (1 / 1)
1.15	N + de + N + PS	empresa de alojamiento web, proveedor de alojamiento web, alojamiento de sitio web (3 / 3)
1.16	N + A + a + N	conexión inalámbrica a Internet (1 / 1)
1.17	N + N + a + N	animación frame a frame, protocolo punto a punto (2 / 1)
1.18	N + de + N + N	buzón de correo POP3 (1 / 1)
1.19	N + de + N + a + N	software de acceso a Internet, proveedor de acceso a Internet, autorización de acceso a Internet (3 / 1)
1.20	N + de + N + de+ N	ataque de denegación de servicio, fichero de firmas de virus, herramienta de creación de sitios, programa de conexión a Internet, programa de compartido de ficheros, programa de intercambio de ficheros, Protocolo de Transferencia de Ficheros, proveedor de servicios de aplicaciones, proveedor de servicios de Internet, sistema de detección de intrusos, sistema de detección de

		intrusiones, sistema de nombre de dominio, <u>velocidad de conexión a Internet (13 / 3)</u>
1.21	N + fuera + de + N	navegación fuera de línea (1)
2	Surcomposés exogènes	
2.1	Emp3	Domain Name System, e-market place, File Transfer Protocol, Frequently Asked Question, junk e-mail, password cracker, shadow password, voice over IP, World Wide Web, Wireless Application Protocol (10)

Tableau 114 : Les surcomposés espagnols à quatre et cinq éléments

Cat.	Structures	Surcomposés
1	Surcomposés endogènes purs et surcomposés endogènes hybrides	
1.1	N + A + no + N	correo electrónico no deseado (1)
1.2	N + por + N + A	conexión por linea conmutada (1)
1.3	N + a + N + a + A + N	<u>acceso a Internet a alta velocidad (1 / 1)</u>
1.4	N + A + de + N + A	conexión permanente de banda ancha (1)
1.5	N + de + N + de + PS + N	<u>Lenguaje de Marcado de Hipertexto (1 / 1)</u>
1.6	N + de + N + A + A	Línea de Subscripción Asimétrica Digital (1)
1.7	N + de + N + de + N + A	Protocolo de Aplicación de Telefonía Inalámbrica (1)
1.8	N + de + N + en + N	proveedor de servicios en línea, hoja de estilo en cascada (2)
1.9	N + a + N + por + N	<u>acceso a Internet por cable (1 / 1)</u>
1.10	N + por + N + a + N	<u>acceso por cable a Internet (1 / 1)</u>
1.11	N + A + de + todo + N	aplicación libre de todo coste (1)
2	Surcomposés exogènes	
2.1	Emp4	Small Office Home Office (1)
2.2	Emp5	Transmission Control Protocol / Internet Protocol (1)

Annexe 9. Typologies de la centricité des noms composés du corpus

1. Les noms composés anglais

Tableau 115 : Les noms composés anglais

Les noms composés anglais français et espagnols du domaine d'Internet :

Centricité	Composés anglais
1- Composés endocentriques N2N1 est un N1	access control, ad blocker, ad-blocking program, address bar, address book, ad-free zone, ADSL modem, ADSL provider, alias page, always-on access, always-on broadband connection, always-on connection, always-on routed Internet, analog modem, anonymous FTP, anonymous surfing, antivirus provider, Application Service Provider, archive manager, archive moderator, Asymmetric Digital Subscriber Line, asynchronous transfer, attack program, audience measurement, banner ad, banner advertising, blacklist, Blind Carbon Copy, blocking access, bookmark, bookmark manager, Boolean operator, Boolean search, boot virus, broadband Internet, browser cache, browser sniffer, browsing habit, buffer overflow attack, bug hunter, bulk email, Bulletin Board System, cable Internet access, cable modem, Cascading Style Sheet, central server, Challenge Handshake Authentication Protocol, chat group, chat host, chat server, chat session, chat software, chat window, chatbox, chat-mate, chatroom, click-and-mortar business, client interface, client-server architecture, collocated server package, commercial site, comparison site, connection kit, connection speed, connection time, contact group, contact list, cookie builder, cookie file, cookie management, cookie manager, cyber activity, cyber album, cyber directory, cyber market, cyber monitoring, cyber referendum, cyber shop, cyber visitor, cyberattack, cyberbrand, cybercafé, cyberconference, cyber-conflict, cybercop, cybercrime, cybercriminal, cybercriminality, cyber-debate, cyber-delinquency, cyber-delinquent, cybereditor, cyberethics, cyberflirt, cyberfrontier, cyber-grocer, cyberhacking, cyberlanguage, cyberlaw, cyberlove, cybermerchant, cyberprotection, cyber-recruitment, cybersex, cybersoldier, cyberspace, cyberstress, cybersupermarket, cyberterrorism, cyberterrorist, cyberunivers, cyberworld, data packet, dedicated server, Denial of Service attack, dial-up connection, dial-up modem, digital signature, discussion area, discussion board, discussion forum, discussion group, distribution list, DNS server, domain name, Domain Name System, DoS attack, dot-com industry, dotcom startup, download, download manager, download site, download speed, download time, downloading, Dynamic HyperText Markup Language, dynamic IP address, ebiz, ebook, e-business, e-business integrator, e-card, e-commerce package, e-commerce site, e-commerce, e-customer, e-employee, e-job, e-learning, electronic hoax, electronic mail account, electronic message, electronic messaging, electronic messaging client, electronic newsletter, electronic signature,

electronic-commerce, e-mail account, e-mail address, email alert, e-mail attachment, e-mail client, e-mail encryptor, e-mail filtering, e-mail header, email list, email manager, e-mail newsletter, email provider, e-mail server, email software, email virus, email worm, email, e-market place, emoticon, encrypted email, encrypted virus, encryption key, encryption program, encryption system, e-purchase, e-tourism, e-trade, e-trader, executable attack, eXtensible Markup Language, external modem, e-zine site, file sharing, file swapping, File Transfer Protocol, file-sharing program, file-swapping program, filtering package, filtering rule, flamewar, forwarding address, fraggle attack, frame-to-frame animation, freeware, Frequently Asked Questions, FTP access, FTP client, FTP file, FTP server, FTP space, general portal, groupware, guestbook, hacked page, hacked site, hacker attack, hacker community, hacking attack, hardware firewall, high-speed access, high-speed connection, high-speed Internet, high-speed Web access, history folder, hit counter, HMTL site, home page, hosted data, hosting company, hosting package, hosting URL, hotline, HTML editor, HTML integrator, HTML page, HTTP server, human-assisted search engine, hyperlinked video, HyperText Markup Language, HyperText Transfer Protocol, inbox, incoming mail server, indexed page, indexing engine, individual ADSL, information superhighway, instant message, instant messaging, instant messenger, instant translator, Internet, Internet access, Internet access authorization, Internet access package, Internet Access provider, Internet cache, Internet café, Internet community, Internet connection, Internet connection software, Internet content provider, Internet gateway, Internet name, Internet portal, Internet Protocol, Internet server, Internet Service Provider, Internet tool, Internet traffic load, Internet user, intrusion detection system, invisible Web, IP address, IP network, IP router, IRC channel, IRC client, IRC server, Java applet, junk email, keylogger, leased line connection, live update, Local Area Network, local ISP, log file, log sniffer, low-speed connection, mail checker, mail forwarding, mail server, mail bombing, mailbox, mailing list, mailing-list management software, malicious code, managed Service Provider, massive attack, message board, message filter, meta portal, meta search engine, meta tag, mirror site, mobile Internet, MP3 file, MP3 site, multiaccess portal, multicriteria search, multipartite virus, musical website, Napsterized file, navigation bar, Net access, Net congestion message, Net connection, Net connection speed, Net economy, Net marketer, Net marketing, Net user, netizen, network administrator,

	network connection, network firewall, newsgroup, newsgroup server, newsgroup software, newsgroupie, newsletter, official page, official site, offline browser, offline browsing, online privacy, online security, Online Service Provider, online shop, online shopping, open source app, openware, outcoming mail server, P2P network, page view, parody site, password, password cracker, peer-to-peer network, personal home page, personal site, ping sweep, Point to Point Protocol, polymorphic virus, POP account, POP server, POP3 email box, POP3 mail box, pop-down ad, pop-under ad, pop-up ad, pop-up message, port blocking, port scanning, portal site, Post Office Protocol, price comparison site, private key, private virtual network, program virus, proxy firewall, proxy server, PSTN connection, public IP address, public key, read notification, receipt notification, remodelled site, remote access, remote computer, remote server, rotation banner, satellite modem, script file, script virus, script-blocking, scrolling banner, search engine, secure certificate, secure server, security hole, server attack, server connection, server package, server-to-server architecture, shadow password, shared hosting, shared hosting package, shared IP address, shared mail server, shared server, shared server package, shareware, Short Message Service, showroom, Simple Mail Transfer Protocol, simultaneous connection, site creation tool, site directory, site owner, site structure, SMS messaging, smurf attack, snail mail, software firewall, source code, space provider, spam protection, specific portal, spyware, static IP address, static Web page, stealth virus, storage site, streaming video, style sheet, surfing habit, techno-nerd, teleworking, thematic directory, thematic portal, transfer speed, Trojan Horse detector, Trojan virus, underground site, Uniform Resource Locator, unique visitor, upload, URL history, Usenet group, Usenet posting, user account, user authentication, user interface, user interface expert, username, video chat, video-chatting, videotostream, viral attack, viral code, viral infection, viral worm, virtual hosting, virus alert, virus attack, virus definition, virus update, virus writer, virus-signature file, visible Web, Vmail, voice chat, voice email, WAP locking, WAP page, WAP portal, Web accelerator, Web access, Web address, Web agency, Web attack, Web browser, web bug, Web cobrowsing, Web collaborating, Web community, Web design, Web designer, Web developer, Web directory, Web hosting, Web hosting company, Web manager, Web marketer, Web page, Web radio, web remix, web remixer, Web server, Web software, Web space, Web surfer, Web surfing habit, Web user interface, webcam, webcenter, weblog,
--	--

	weblogger, weblogging, webmail, webmaster, website, white hacker, Wireless Application Protocol, wireless connection, wireless Internet connection, wireless Web, World Wide Web, worm alert, worm attack
2- Composés endocentriques N1N2 est un N1	business to consumer, Business to Employee, business-to-business, Denial of Service, network of networks, peer-to-peer, portal of portals, voice IP
3- Composés endocentriques N2N1 est un N1 et un N2	hacktivism, hacktivist, modem, Small Office Home Office, Transmission Control Protocol/Internet Protocol
4- Composés exocentriques *N1N2 est un N1 et un N2	backdoor, bandwidth, black hat, bottleneck, click and mortar, cookie ID, cyberpunk, cybersquatter, cybersquatting, dotcom, dot-org, firewall, frame relay, hyperlink, intelligent agent, Internet cookie, mailto, permanent cookie, pop up, rollover, script kiddie, search engine spider, Secure Socket Layer, shopbot, startup, temporary cookie, thumbnail, Trojan Horse, uplink, white hat

2. Les noms composés français

Tableau 116 : Les noms composés français

Les noms composés anglais français et espagnols du domaine d'Internet :

Centricité	Composés français
1- Composés endocentriques N1N2 est un N1	accès à distance, accès distant, accès FTP, accès haut-débit, accès Internet, accès Internet par câble, accès permanent, accès Web, accès web haut-débit, accusé de lecture, accusé de réception, achat en ligne, administrateur de site, administrateur réseau, administrateur Web, adresse de courrier électronique, adresse d'hébergement , adresse électronique, adresse e-mail, adresse Internet, adresse IP, adresse IP dynamique, adresse IP fixe, adresse IP partagée, adresse IP publique, adresse URL, adresse Web, ADSL individuel, alerte de virus, alerte e-mail, animation frame par frame, annuaire de sites, annuaire thématique, annuaire Web, applet Java, apprentissage en ligne, architecture client-serveur, architecture serveur-serveur, article Usenet, aspirateur de site, aspirateur Internet, aspirateur Web, assistant de téléchargement, Asymmetric Digital Subscriber Line, attaque buffer overflow, attaque de hacker, attaque de piratage, attaque de serveur, attaque de ver, attaque de virus, attaque Denial of Service, attaque DoS, attaque fraggle, attaque massive, attaque smurf, attaque virale, attaque Web, autorisation d'accès à Internet, autorité de certification, autoroute de l'information, babillard électronique, balise méta, bandeau publicitaire, bannière défilante, bannière publicitaire, bannière rotative, barre d'adresses, barre de navigation, blacklist, blocage de port, bloqueur de popup, boîte aux lettres, boîte aux lettres électronique, boîte aux lettres POP3, boîte de réception, BtoE, business to business, business to consumer, Business to Employee, cache du navigateur, cache Internet, café Internet, canal de chat, canal IRC, canular électronique, carnet d'adresses, CD de connexion, certificat de sécurité, chasseur de bug, chat vocal, chiffrement d'e-mail, citoyen du Net, clé de chiffrement, clé privée, clé publique, client de chat, client de courrier, client de courrier électronique, client de messagerie électronique, client e-mail, client FTP, client IRC, code malveillant, code source, code viral, commerce électronique, communauté du Web, communauté hacker, communauté Internet, comparateur de prix, compte de courrier, compte de courrier électronique, compte de messagerie, compte e-mail, compte pop, compte utilisateur, compteur d'accès, compteur de visite, concepteur Web, connexion à bande large permanente, connexion à faible débit, connexion dédiée, connexion en dialup, connexion haut débit, connexion Internet, connexion Internet sans fil, connexion par ligne commutée, connexion permanente, connexion réseau, connexion RTC, connexion sans fil, connexion serveur, connexion simultanée, contrôle d'accès, copie carbone invisible,

courrier électronique, courrier électronique chiffré, courrier escargot, courrier non-sollicité, courrier postal, cracker de mot de passe, définition de virus, Dénie de Service, design Web, designer Web, détecteur de chevaux de Troie, développeur Web, dialogue en direct, diffuseur de virus, donnée hébergée, download, échange de fichier(s), éditeur HTML, e-mail chiffré, e-mail crypté e-mail non sollicité, en-tête de courrier, entreprise point com, envoi groupé, e-place de marché, ergonome web, espace de discussion, espace ftp, espace Web, faille de sécurité, fenêtre de chat, fenêtre pop under, fenêtre pop up, fenêtre publicitaire, feuille de style, feuille de style en cascade, fichier de signature de virus, fichier espion, fichier FTP, fichier journal, fichier log, fichier MP3, fichier Napster, filtrage d'e-mails, filtre anti-pub, filtre de messages, firewall logiciel, Foire Aux Questions, forum de discussion, fouineur de log, fournisseur ADSL, fournisseur d'accès, fournisseur d'accès Internet, fournisseur d'antivirus, fournisseur d'applications hébergées, fournisseur de contenu, fournisseur de contenu Internet, fournisseur de courrier, fournisseur de courrier électronique, fournisseur de services en ligne, fournisseur de services gérés, fournisseur de services Internet, fournisseur d'e-mail, fournisseur d'espace, fournisseur d'hébergement, fournisseur d'hébergement web, FSI local, FTP anonyme, générateur de cookies, générateur de virus, gestion de cookies, gestion de mailing list, gestionnaire d'archives, gestionnaire de cookies, gestionnaire de courrier électronique, gestionnaire de favoris, gestionnaire de téléchargement, gestionnaire Web, goulot d'étranglement, groupe de chat, groupe de contact, groupe de discussion, groupe de News, groupe Usenet, habitude de navigation, hacker blanc, hacker noir, hébergement de site, hébergement partagé, hébergement virtuel, hébergeur de chat, infection virale, intégrateur e-business, intégrateur HTLM, interface client, interface graphique web, interface utilisateur, interface Web, Internet bande large, Internet haut débit, Internet mobile, Internet nomade, intimité en ligne, ISP local, journal de connexion, kit de connexion, langage DHTML, Langage HTML, langage XML, lien consulté, Ligne d'abonnement numérique asymétrique, liste de contact, liste de diffusion, liste de distribution, liste noire, livre d'or, log de connexion, logiciel anti popup, logiciel anti-pub, logiciel d'accès Internet, logiciel de chat, logiciel de chiffrement, logiciel de connexion Internet, logiciel de courrier, logiciel de cryptage de courrier, logiciel de filtrage, logiciel de navigation, logiciel de news, logiciel de serveur, logiciel d'hébergement, logiciel espion, logiciel gratuit, logiciel libre, logiciel partagé, logiciel spyware, logiciel Web,

	magasin en ligne, mail vocal, message de congestion, message électronique, message instantané, message Internet, message non-sollicité, message pop up, message vocal, messagerie électronique, messagerie instantanée, messagerie SMS, mesure d'audience, mise à jour anti-virus, mise à jour automatique, modem ADSL, modem analogique, modem câble, modem commuté, modem externe, modem RTC, modem satellite, mot de passe, moteur à assistance humaine, moteur de recherche, moteur d'indexation, navigation anonyme, navigation hors-ligne, navigation off-line, newsletter électronique, newsletter e-mail, nom de domaine, nom d'utilisateur, non-connecté, notificateur de mail, opérateur booléen, ordinateur distant, ossature de site, outil de création de site, outil de recherche, outil Internet, pack ADSL, package d'hébergement partagé, page alias, page d'accueil, page d'accueil perso, page d'accueil personnelle, page de démarrage, page HTML, page indexée, page officielle, page perso, page personnelle, page piratée, page visitée, page vue, page WAP, page Web, page Web statique, paquet d'information, pare-feu proxy, pare-feu réseau, partage de fichier(s), passerelle Internet, peer-to-peer, pièce jointe, pirate informatique, pixel espion, portail de portails, portail généraliste, portail Internet, portail multi-accès, portail spécifique, portail thématique, portail WAP, programme d'accès Internet, programme d'attaque, programme de partage de fichier, programme d'échange de fichier, programme d'e-commerce, programme espion, programme open source, propriétaire de site, protection anti-spam, Protocole d'applications sans fil, Protocole de Transfert de Fichier, protocole http, protocole Internet, protocole pop, protocole PPP, protocole SMTP, protocole TCP/IP, publicité non-sollicitée, rapatriement de courrier, recherche multicritère, règle de filtrage, renifleur de navigateur, requête booléene, réseau des réseaux, réseau IP, réseau P2P, réseau peer-to-peer, réseau privé virtuel, responsable de la veille, robot logiciel, routeur IP, salon de chat, salon de discussion, scan de port, sécurité en ligne, serveur central, serveur de chat, serveur de courrier, serveur de courrier entrant, serveur de courrier sortant, serveur de News, serveur de newsgroup, serveur dédié, serveur distant, serveur DNS, serveur FTP, serveur http, serveur Internet, serveur IRC, serveur mail, serveur mandataire, serveur partagé, serveur partagé de courrier, serveur POP, serveur proxy, serveur sécurisé, serveur Web, session de chat, shopping en ligne, signature de virus, signature électronique, signature numérique, site à assistance humaine, site d'archivage, site de
--	--

	comparaison, site de comparaison de prix, site de parodie, site de téléchargement, site d'e-commerce, site e-zine, site filleul, site hacké, site HTML, site Internet, site marchand, site mercantile, site mirror, site MP3, site officiel, site parodique, site parrain, site perso, site personnel, site piraté, site portail, site remodelé, site underground, site visité, site Web, site Web musical, site webzine, snail mail, surf anonyme, système de chiffrement, système de détection d'intrusion, Système de Nom de Domaine, temps de connexion, temps de téléchargement, Toile d'araignée mondiale, traducteur instantané, transfert asynchrone, ver de courrier, vidéo hyperliée, virus crypté, virus de boot, virus de courrier électronique, virus de programme, virus de script, virus e-mail, virus furtif, virus multipartite, virus polymorphique, virus troyen, visiteur unique, vitesse de connexion, vitesse de connexion Internet, vitesse de téléchargement, vitesse de transfert, voice IP, voix sur IP
2- Composés endocentriques N2N1 est un N1	adware, bookmark, buffer overflow, chatbox, chat room, contact list, cyber amour, cyber civisme, cyber univers, cyberacheteur, cyberactivité, cyberalbum, cyberannuaire, cyberattaque, cyberboutique, cybercafé, cybercaméra, cyber-commerçant, cyberconférence, cyberconflit, cyberconsommateur, cybercrime, cybercriminalité, cybercriminel, cyberdébat, cyber-délinquance, cyberdélinquent, cyberdroit, cyberéditeur, cyberépicier, cyberespace, cyberflic, cyber-flirt, cyberfrontière, cyberguerre, cyberlangue, cybermarché, cybermarque, cybermonde, cyber-piraterie, cyber-plébiscite, cyberprotection, cyber-recrutement, cybersexe, cybersoldat, cyberstress, cyber-supermarché, cybersurveillance, cyberterrorisme, cyberterroriste, cybervisiteur, e-achat, e-biz, ebook, e-business, e-card, e-commerce, e-courrier, e-économie, e-employé, e-learning, e-mail, e-métier, émoticonne, e-picnic, espiogiciel, e-tourisme, e-xagone, eXtensible Markup Language, File Transfer protocol, freeware, Frequently Asked Questions, groupware, haut débit, home page, hotline, hotliner, HyperText Markup Language, internaute, Internet, key logger, mail bombing, mailing list, message board, metabalise, métamoteur, metaportail, meta tag, newsgroup, newsletter, Net-économie, néritquette, netmarketeur, netmarketing, netsurfer, netsurfeur, openware, ping sweep, shareware, Short Message Service, spyware, techno-nerd, téléchargement, téléformation, télémaintenance, télétravail, télétravailleur, Transmission Control Protocol/Internet Protocol, user authentication, videochat, videotstream, video streaming, wappanaute, waplockage, Web agency, web bug, webcam, webcenter, Web collaborating, web designer, Web invisible, weblog, weblogger,

Les noms composés anglais français et espagnols du domaine d'Internet :

	weblogging, webmail, webmarketeur, webmaster, webmestre, Web mobile, web radio, Web-remixeur, web sans fil, Web visible, webzine, white hat, Wireless Application Protocol, World Wide Web
3- Composés endocentriques N1N2 est un N1 et un N2	apprenti pirate, entreprenaute, hacktivisme, hacktiviste, modem
4- Composés exocentriques *N1N2 est un N1 et un N2	agent intelligent, backdoor, araignée de moteur de recherche, bande passante, black hat, cheval de Troie, click and mortar, cookie Internet, cookie permanent, cookie temporaire, cyber-punk, cyberpunkitude, cybersquatter, cybersquatting, dot org, dotcom, firewall, frame relay, hyperlien, inbox, jeune pousse, lien hypertexte, mailto, open source, pare-feu, point com, pop under, pop up, porte de derrière, porte de service, porte dérobée, rollover, script kiddie, shopbot, showroom, start-up, start-upeur, thumbnail, upload

3. Les noms composés espagnols

Tableau 117 : Les noms composés espagnols

Centricité	Composés
1- Composés endocentriques N1N2 est un N1	acceso a alta velocidad, acceso a Internet, acceso a Internet a alta velocidad, acceso a Internet por cable, acceso a la Red, acceso FTP, acceso permanente, acceso por cable a Internet, acceso remoto, acceso Web, actualización automática, actualización de virus, acuse de lectura, acuse de recibido, administrador de red, administrador Web, ADSL individual, ADSL multipuesto, agenda de direcciones, agujero de seguridad, alerta de correo, alerta de virus, alojamiento de sitio web, animación frame a frame, aplicación gratuita, aplicación libre de todo coste, aplicación Web, applet Java, archivo adjunto, archivo de registro, archivo log, arquitectura cliente-servidor, artículo de Usenet, ataque de denegación de servicio, ataque de gusano, ataque de virus, ataque DOS, ataque fraggle, ataque hacker, ataque masivo, ataque smurf, ataque vírico, ataque web, autenticación de usuario, autopista de la información, autorización de acceso a Internet, banner de publicidad, banner publicitario, banner rotatorio, barra de direcciones, barra de navegación, barrido de puerto, barrido ping, bloqueador de pop-up, bloqueo de puerto, boletín de noticias, boletín de noticias electrónico, bulo electrónico, búsqueda booleana, Business to Consumer, Business to Employee, buzón de correo, buzón de correo POP3, buzón de mail, buzón electrónico, caché de Internet, caché del browser, caché del navegador, café Internet, cámara Web, canal de chat, canal IRC, certificado de seguridad, charla de voz, chequeador de email, ciudadano de la red, clave de acceso, clave de cifrado, clave privada, clave pública, cliente de correo, cliente de mail, cliente de mensajería electrónica, cliente FTP, cliente IRC, código fuente, código malicioso, comercio electrónico, comparador de precios, compartimiento de ficheros, comunidad hacker, comunidad internauta, comunidad Internet, conexión a Internet, conexión ADSL multipuesto, conexión de alta velocidad, conexión de baja velocidad, conexión de línea dedicada, conexión de red, conexión dedicada, conexión inalámbrica, conexión inalámbrica a Internet, conexión permanente, conexión permanente de banda ancha, conexión por dial-up, conexión por línea conmutada, conexión RTC, conexión servidor, conexión simultánea, contador de accesos, contador de visitas, contraseña en la sombra, control de acceso, copia oculta, correo basura, correo cifrado, correo electrónico, correo electrónico cifrado, correo electrónico comercial, correo electrónico junk, correo electrónico no deseado, correo indeseado, correo postal, correo vocal, correo web,

	cortafuegos proxy, cuenta de correo, cuenta de correo electrónico, cuenta de usuario, cuenta e-mail, cuenta POP, definición de virus, denegación de servicio, Denial of Service, desarrollador Web, desbordamiento de bufer, detector de troyanos, dirección de e-mail, dirección electrónica, dirección Internet, dirección IP, dirección IP dinámica, dirección IP fija, dirección IP pública, dirección URL, dirección visitada, dirección Web, directorio de sitios, directorio temático, directorio Web, diseñador Web, diseño Web, documento anexo, editor HTML, email comercial, e-market place, empresa a empresa, empresa de alojamiento web, empresa de hosting, empresa de housing, empresa.com, encriptador de mensajes, espacio ftp, espacio Web, estructura de sitio, etiqueta Meta, explorador Web, fichero de firmas de virus, fichero FTP, fichero log, fichero MP3, fichero Napster, fichero script, filtrado de correo, filtro de mensajes, firewall hardware, firewall personal, firewall software, firma de virus, firma digital, firma electrónica, foro de debate, foro de discusión, forum de discusión, FTP anónimo, generador de virus, gestión de cookies, gestor de archivos, gestor de bookmarks, gestor de cookies, gestor de correo, gestor de descarga, gestor de mail, gestor Web, grupo de charla, grupo de contacto, grupo de discusión, grupo de noticias, grupo Usenet, gusano de correo, gusano vírico, hábitos de navegación, hacker blanco, hacker negro, herramienta de creación de sitios, herramienta de Internet, hoja de estilo, hoja de estilo en cascada, home perso, hosting compartido, infección de virus, información alojada, integrador e-business, integrador HTML, intercambio de ficheros, interfaz cliente, interfaz de usuario, interfaz gráfica web, interfaz Web, Internet de alta velocidad, Internet fija, Internet móvil, intimidad on line, ISP local, kit de conexión, lector de news, Lenguaje de Marcado de hipertexto, lenguaje DHTML, lenguaje HTML, lenguaje XML, libreta de direcciones, libro de visita, Línea de Subscripción Asimétrica Digital, lista de contacto, lista de correo, lista de discusión, lista de distribución, lista negra, marketing Web, medida de audiencia, mensaje cifrado, mensaje corto, mensaje de correo, mensaje de correo electrónico, mensaje de voz, mensaje instantáneo, mensaje Internet, mensaje pop-up, mensaje SMS, mensaje vocal, mensajería electrónica, mensajería instantánea, mensajería SMS, módem ADSL, módem analógico, módem de cable, módem dial-up, módem externo, módem RDSL, módem satélite, motor de búsqueda, motor de indexación, navegación anónima, navegación off-line, navegador de Internet,
--	--

navegador fuera de línea, navegador offline, navegador web, negocio a negocio, negocio electrónico, newsletter electrónica, nombre de dominio, nombre de usuario, notificador de correo, operador booleano, ordenador remoto, página alias, página de inicio, página de inicio personal, página hackeada, página HTML, página indexada, página inicial, página oficial, página perso, página personal, página visitada, página vista, página WAP, página Web, página Web estática, paquete de información, pasarela de Internet, peer-to-peer, pirata informático, portal de portales, portal específico, portal generalista, portal Internet, portal multiacceso, portal temático, portal WAP, Preguntas más frecuentes, programa anti pop-up, programa de ataque, programa de chat, programa de cifrado, programa de compartido de fichero, programa de conexión a Internet, programa de correo, programa de e-commerce, programa de intercambio de ficheros, programa gratuito, programa multi-banner, programa spyware, protección anti-spam, Protocolo de Aplicación de Telefonía Inalámbrica, Protocolo de Transferencia de Ficheros, protocolo http, protocolo Internet, protocolo POP, protocolo punto a punto, protocolo SMTP, proveedor de acceso, proveedor de acceso a Internet, proveedor de alojamiento, proveedor de alojamiento web, proveedor de antivirus, proveedor de ASDL, proveedor de contenido, proveedor de correo, proveedor de espacio, proveedor de hosting, proveedor de Internet, proveedor de servicios, proveedor de servicios administrados, proveedor de servicios de aplicaciones, proveedor de servicios de internet, proveedor de servicios en línea, radio Web, red de área local, red de redes, red IP, red P2P, red peer- to-peer, red privada virtual, regla de filtrado, reventador de contraseña, robot de compras, router IP, sala de charla, sala de chat, salón de chat, seguridad en línea, servidor central, servidor compartido, servidor de chat, servidor de correo, servidor de correo entrante, servidor de correo saliente, servidor de DNS, servidor de dominio, servidor de Internet, servidor de mail, servidor de News, servidor de noticias, servidor dedicado, servidor FTP, servidor http, servidor IRC, servidor POP, servidor proxy, servidor remoto, servidor seguro, servidor Web, sesión de charla, sistema de cifrado, sistema de detección de intrusión, sistema de detección de intrusos, sistema de nombre de dominio, sitio comercial, sitio de archivo, sitio de descarga, sitio de e-commerce, sitio de Internet, sitio hackeado, sitio HTML, sitio mirror, sitio MP3, sitio oficial, sitio personal, sitio portal, sitio underground, sitio

Les noms composés anglais français et espagnols du domaine d'Internet :

	Web, software compartido, software de acceso a Internet, software de filtrado, software de navegación, software de servidor, software espía, software libre, tablón de anuncios, tag meta, telaraña electrónica, tiempo de conexión, tiempo de descarga, tienda on-line, traductor instantáneo, transferencia asincrónica, velocidad de conexión, velocidad de conexión a Internet, velocidad de descarga, velocidad de transferencia, venta electrónica, ventana de chat, ventana emergente, ventana flotante, video hipervinculado, virus cifrado, virus de arranque, virus de correo electrónico, virus de programa, virus de script, virus multipartito, virus polimórfico, virus polimorfo, virus stealth, virus troyano, visitante único, voice over IP, voz sobre IP, Web invisible, web musical, Web personal, Web visible,
2- Composés endocentriques N2N1 est un N1	adware, bookmark, buffer overflow, cable módem, chatbox, ciber contratación, ciber supermercado, ciberactividad, ciberadicto, ciberálbum, ciberamor, ciberataque, cibercafé, cibercentro, cibercomerciante, cibercomprador, ciberconferencia, ciberconflicto, cibercrimen, cibercriminal, cibercriminalidad, ciberdebate, ciberdelincuencia, ciberdelinciente, ciberdelito, ciberderecho, ciberdirectorio, cibereditor, ciberespacio, ciberfrontera, ciberguerra, ciberlenguaje, ciberley, cibermayor, cibermercado, cibermundo, cibernauta, ciberocupa, ciberocupación, ciberpiratería, ciber-plebescito, ciberprotección, cibersexo, cibersoldado, cibertenedor, ciberterrorismo, ciberterrorista, cibertienda, ciberuniverso, cibervisitante, Domain Name System, e-biz, ebook, e-business, e-card, e-commerce, e-compra, e-empleado, e-empleo, e-learning, e-mail, e-marketing, emotícono, e-turismo, File Transfer Protocol, freeware, Frequently Asked Questions, groupware, home page, hotline, inbox, internauta, Internet, Internet Protocol, junk e-mail, mailbombing, message board, metabuscador, metaportal, meta tag, netcam, Net-economía, netiqueta, newsletter, no conectado, openware, password cracker, recién conectado, shadow password, shareware, snail mail, spyware, telecentro, teletrabajador, teletrabajo, video chat, vídeo streaming, web bug, weblog, weblogger, Web radio, webcam, webmaster, website, Wireless Application Protocol, World Wide Web
3- Composés endocentriques N1N2 est un N1 et un N2	hacktivismo, hacktivista, jubilnauta, módem, Small Office Home Office, Transmission Control Protocol/Internet Protocol
4- Composés	ancho de banda, agente inteligente, banda ancha, caballo de

exocentriques *N1N2 est un N1 et un N2	Troya, ciberpunk, click and mortar, cookie de Internet, cookie permanente, cookie temporal, cortafuegos, cuello de botella, cyber-punk, dot org, download, enlace de hipertexto, firewall, frame relay, gateway, hipervínculo, mailto, pop up, popunder, puerta trasera, puntocom, rollover, script kiddie, shopbot, thumbnail
---	--

Les noms composés anglais français et espagnols du domaine d'Internet :

Annexe 10. Types de fenêtres

Les quatre types d'agencements envisageables pour créer les fenêtres « Présentation » et « Mode d'emploi » sont :

- Construction mono-fenêtre. Toutes les informations sont contenues dans une fenêtre.
- Construction multi-fenêtres. La quantité d'information délivrée nécessite la création de plusieurs fenêtres reliées entre elles. La navigation d'une fenêtre à une autre se fera en cliquant sur des boutons « page précédente » et « page suivante ».
- Construction info bulle. Les informations apparaîtront lorsque l'utilisateur placera le curseur au-dessus d'un mot différencié des autres par sa couleur, par exemple. Cette construction n'est possible que si le texte est de taille modeste.
- Construction liens. Les informations seront contenues dans plusieurs pages. Pour naviguer d'une page à une autre l'utilisateur cliquera sur le mot ou la suite de mots soulignés.

Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Annexe 11. Autres procédures de réalisation du générateur d'équivalents envisagées

1. Une table unique

1.1. Travail préliminaire

- Calculer le nombre de structures utilisées en langue cible pour rendre une structure anglaise.
- Construire une table unique de taille importante comprenant toutes les vedettes anglaises, les champs des équivalents français, de leur genre, de leur nombre et les termes implantés s'il en existe.
- Attribuer une structure compositionnelle par champ équivalent. Il sera nécessaire de laisser des champs vides lorsque les équivalents créés seront insatisfaisants.
- Définir un ordre définitif de saisie des éléments dans la table.

- Créer des requêtes (requête paramétrée et requête Sélection) pour l'extraction des données dans la table et leur affichage.
- Créer des formulaires d'entrée de données pour la saisie des enregistrements dans la table, de boîte de dialogue pour exécuter des actions et de menu général pour ouvrir d'autres formulaires.
- Saisir chaque vedette anglaise avec ses équivalents français et les informations connexes (genre, nombre, termes implantés).

Structure de la table :

- Enregistrement 1
 - Champ n°1 : Vedette anglaise
 - Champ n°2 : Equivalent 1
 - Champ n°3 : Genre 1
 - Champ n°4 : Nombre 1
 - Champ n°5 : Equivalent 2
 - Champ n°6 : Genre 2
 - Champ n°7 : Nombre 2
 - (jusqu'à équivalent *n*)
 - Champ n°5 : Terme implanté 1
 - Champ n°6 : Genre terme implanté 1
 - Champ n°7 : Nombre terme implanté 1
 - Champ n°8 : Terme implanté 2
 - Champ n°9 : Genre terme implanté 2
 - Champ n°10 : Nombre terme implanté 2
- Enregistrement *n*

Notes :

- Tous les champs se répètent pour chaque enregistrement, c'est-à-dire pour chaque composé anglais (vedette), soit 553 fois.
- On pourrait envisager d'ajouter des champs pour les variantes, synonymes et sigles liés à la vedette anglaise.

1.2. Fonction principale du générateur : Recherche des équivalents

- **Lancer l'application**¹⁹⁴

- Ouvrir la fenêtre *Menu général* du GE
- Cliquer sur le bouton *Sommaire*
- Cliquer sur le bouton *Équivalents français*
 - = Ouverture de la fenêtre *Saisie du composé anglais*
- **Saisir le composé**
- Cliquer sur le bouton *OK*
 - Une requête paramétrée isole la vedette dans la table unique
 - Une seconde requête extrait les équivalents selon l'ordre de saisie des champs équivalents de la table unique
- Ouverture de la fenêtre *Affichage des équivalents français*
- Cliquer sur le bouton *Termes implantés* (option facultative)
- Ou fermer la fenêtre (= retour au *Sommaire*)
- Ou quitter l'application

1.3. Avantages

- La structure de base du générateur est simple.
- Aucune programmation.

1.4. Inconvénients

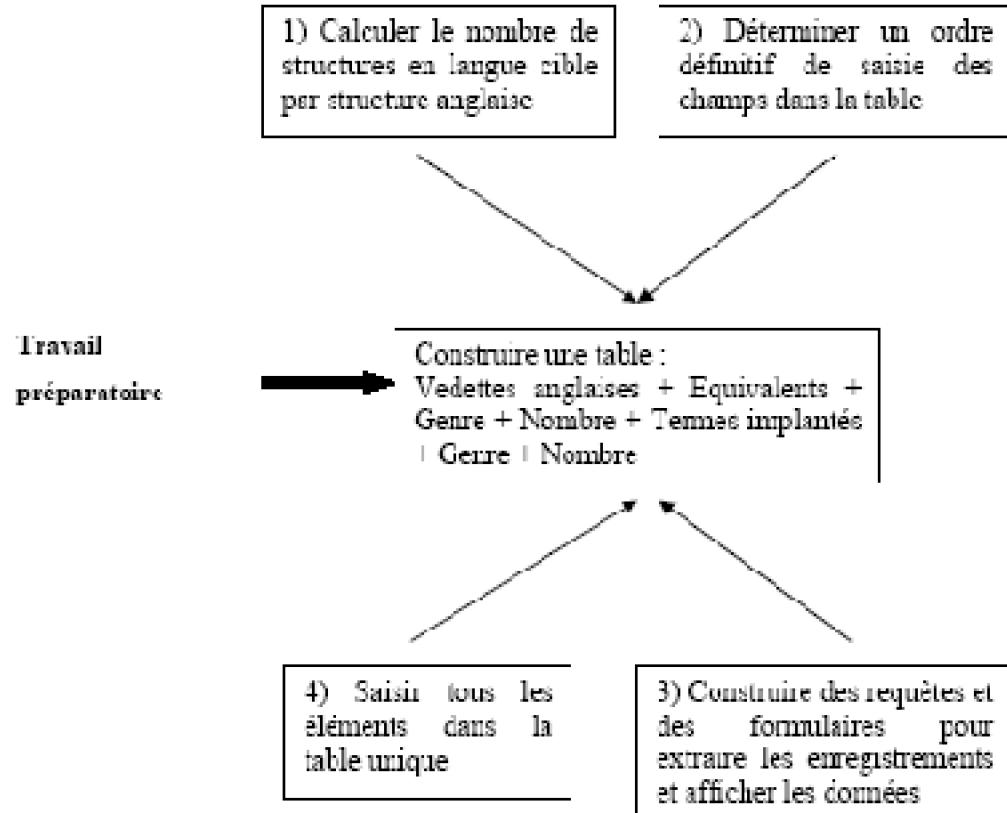
- La table créée est très chargée. Comme elle contient de nombreux champs, le risque d'erreur n'est pas négligeable lors de la phase de saisie qui sera fastidieuse et très lourde.
- Le nombre d'éléments à entrer dans la table est considérable : 553 vedettes anglaises, les équivalents français correspondant à chaque vedette selon leur ordre défini – soit $553 \times n^{195}$ –, les termes déjà utilisés en langue cible s'ils existent. Pour chaque équivalent et terme implanté il faut également faire figurer le genre et le nombre.
- Les champs laissés vides pour les cas dans lesquels les équivalents ne seront pas acceptables provoqueront des problèmes dans l'ordre d'affichage des résultats.

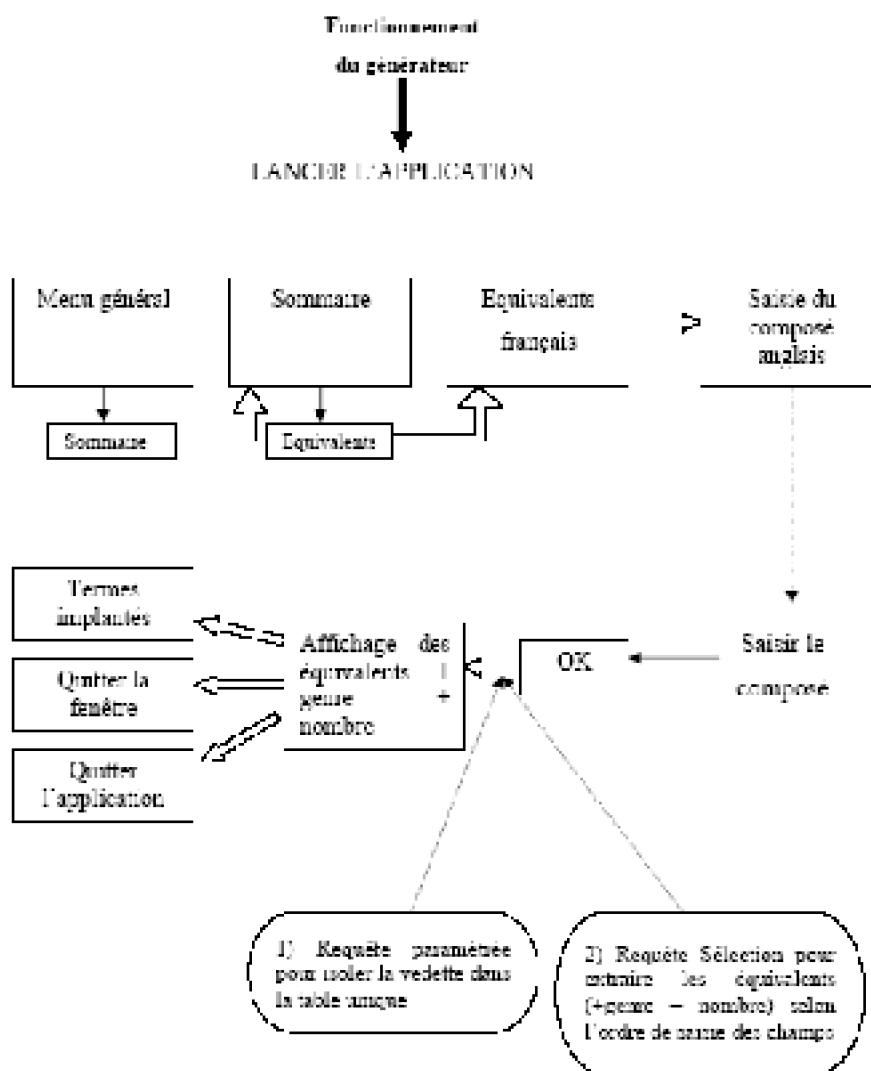
1.5. Schémas

¹⁹⁴ Les schémas sont des opérations visibles pour l'utilisateur. Caractères normaux = opérations sous-jacentes.

¹⁹⁵ Le nombre d'équivalents sera très certainement fixé à cinq.

Les noms composés anglais français et espagnols du domaine d'Internet :





2. Une table unique basée sur le type compositionnel des composés sources et cibles

2.1. Travail préliminaire

- Calculer le nombre de structures utilisées en langue cible pour rendre une structure anglaise.
- Construire une unique table comprenant les champs suivants : composé anglais, structure du composé anglais, équivalents en langue cible, structure des équivalents,

genre, nombre des équivalents et les termes implantés s'il en existe.

- Créer des requêtes (requêtes paramétrées et requêtes Sélection) pour l'extraction des données dans la table et leur affichage.
- Créer des formulaires d'entrée de données pour la saisie des enregistrements dans la table, de boîte de dialogue pour exécuter des actions et de menu général pour ouvrir d'autres formularies.
- Saisir toutes les données.

Structure de la table :

- Enregistrement n°1
- Champ n°1 : Vedette anglaise
- Champ n°2 : Structure vedette anglaise
- Champ n°3 : Equivalent 1
- Champ n°4 : Structure équivalent 1
- Champ n°5 : Genre 1
- Champ n°6 : Nombre 1
- Champ n°7 : Equivalent 2
- Champ n°8 : Structure équivalent 2
- Champ n°9 : Genre 2
- Champ n°10 : Nombre 2
- (jusqu'à l'équivalent *n*)
- Champ : Terme implanté 1
- Champ : Genre terme implanté 1
- Champ : Nombre terme implanté 1
- Champ : Terme implanté 2
- Champ : Genre terme implanté 2
- Champ : Nombre terme implanté 2

Notes :

- Tous ces champs se répètent pour chaque enregistrement, soit 553 fois.
- Le nombre d'équivalents sera très certainement fixé à cinq.
- Il est envisageable d'ajouter des champs pour les variantes, les synonymes et les sigles liés à la vedette anglaise.

2.2. Recherche des équivalents : fonction principale du générateur

- **Lancer l'application**¹⁹⁶
 - Ouvrir la fenêtre *Menu général* du GE
 - Cliquer sur le bouton *Sommaire*
 - Cliquer sur le bouton *Equivalents français*

= Ouverture de la fenêtre Saisie du composé anglais
- **Saisir le composé**
 - Cliquer sur le bouton *OK*
 - Une requête paramétrée isole la vedette anglaise dans la table unique
 - Une seconde requête Sélection extrait la structure de la vedette saisie
 - Boucle de décision : si la structure vedette est 'XY' alors exécution de l'action n°1 ou 2 ou 3 ou *n*
 - Une requête Sélection récupère les équivalents selon l'action exécutée
- Ouverture de la fenêtre Affichage des équivalents français.
 - Cliquer sur le bouton *Termes implantés* (option facultative)
 - Ou fermer la fenêtre (= retour au *Sommaire*)
 - Ou quitter l'application

2.3. Avantages

- Ce système est transposable tant que le nombre de requêtes reste raisonnable.
- Plus de champs vides, la structure de la table est moins rigide et la saisie est moins contraignante que dans la procédure n°1.

2.4. Inconvénients

- Comme pour la procédure n°1, la table est chargée. Le risque d'erreur lors de la phase de saisie est également non négligeable ici.
- Les boucles de décision sont très lourdes.
- La multiplication des boucles de décision rend incertain l'ordre d'affichage des équivalents.
- Les données ne peuvent pas être importées (donc phase de saisie lourde).
- Les requêtes ne sont pas modifiables. Il est impossible d'actualiser la base de données pour entrer de nouveaux enregistrements car cela provoquerait la

¹⁹⁶ Caractéristiques opérationnelles possibles pour utiliser le générateur d'équivalents suivantes

protégé en vertu de la loi du droit d'auteur.

modification des actions. Le temps d'affichage sera long.

2.5. Explication des opérations sous-jacentes

- Address book
 - Requête paramétrée pour isoler la vedette anglaise dans la table unique. *Address book* = enregistrement n°4
 - Requête pour extraire la structure de la vedette saisie. La structure morphologique de *address book* est N + N
 - Boucle de décision

Si la structure de la vedette saisie est N + N alors l'action exécutée sera la n°1 = équivalent 1 (N + de + N)

- Equivalent 2 (N + N)
- Equivalent 3 (emprunt)
- Equivalent 4 (N)
- Equivalent 5 (N + A)
- Implanté 1
- Implanté 2

Suite à l'exécution de l'action n°1, l'affichage des équivalents français respecte l'ordre suivant : 1) carnet d'adresses (équiv 1)

- 2) carnet-adresses (équiv. 2)
- 3) address book (équiv. 3)
- 4) carnet (équiv 4)
- 5) carnet relationnel (équiv 5)

Implanté 1

Implanté 2

- black hat
 - Requête paramétrée pour isoler la vedette anglaise dans la table unique. *Black hat* = enregistrement n°15
 - Requête pour extraire la structure de la vedette saisie. La structure morphologique de *black hat* est A + N
 - Commande pour lancer une boucle de décision
 - Si la structure de la vedette saisie est A +N alors l'action exécutée sera la n°3 =

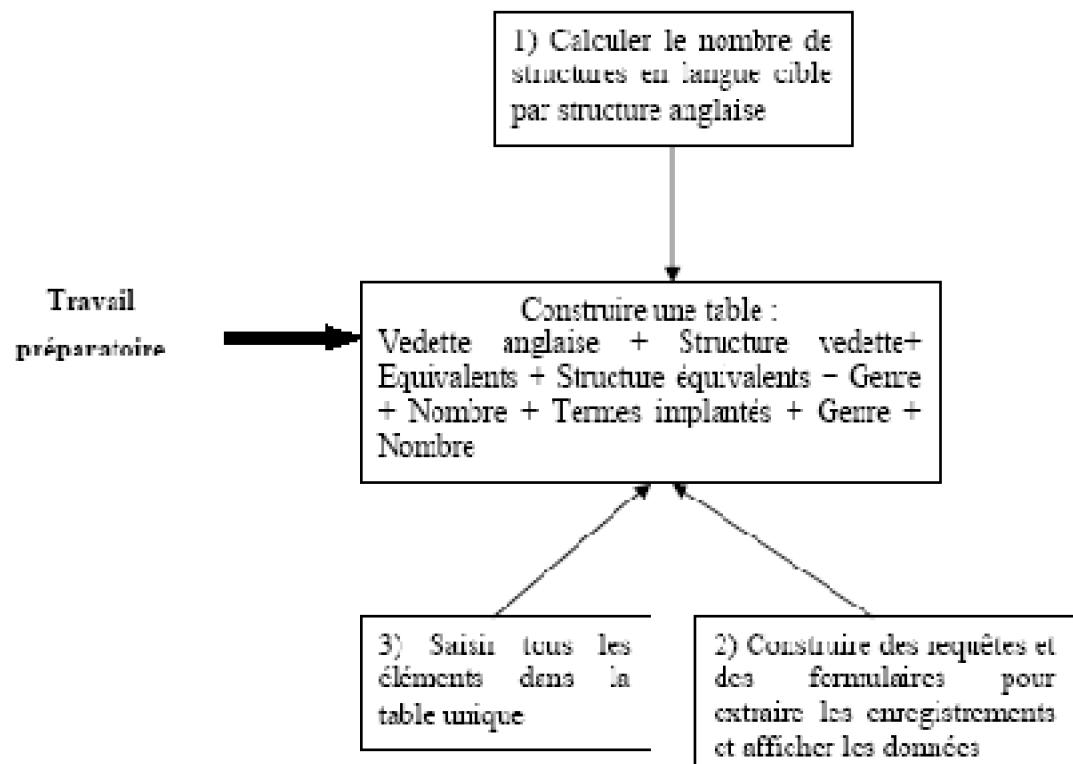
équivalent 5 (N + A)

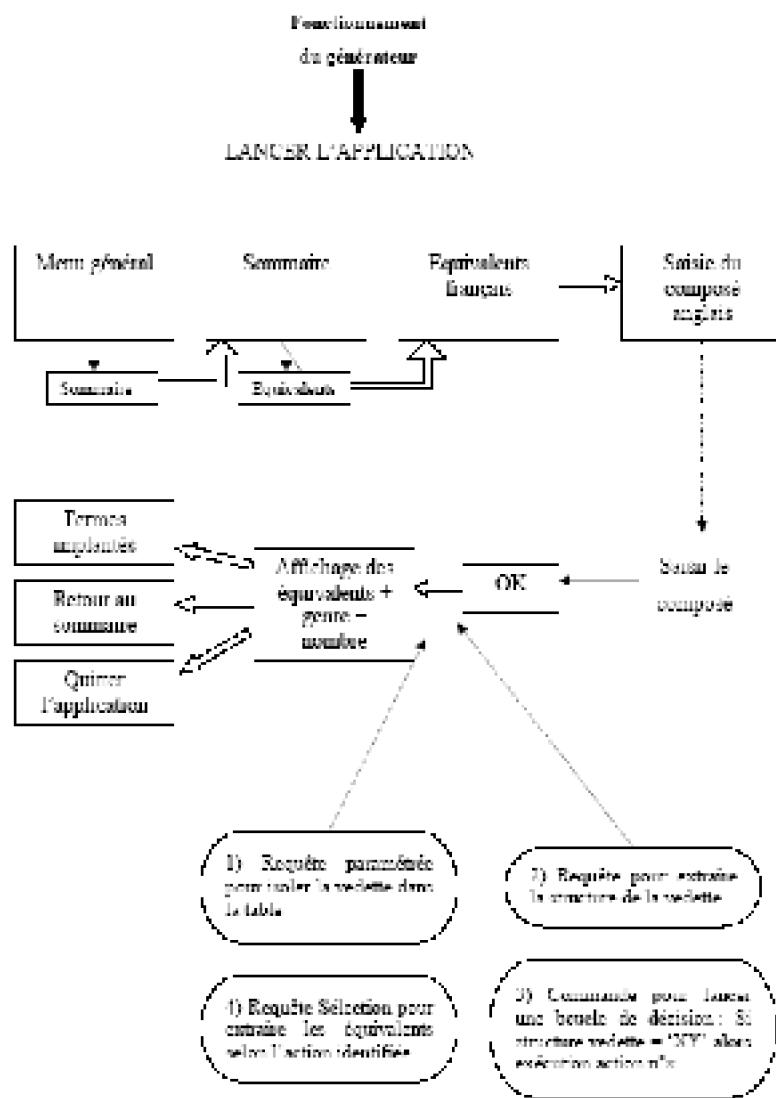
- * Equivalent 3 (emprunt)
- * Equivalent 2 (N + N)
- * Equivalent 6 (N + à + N)
- * Equivalent 4 (N)
- * Implanté 1
- * Implanté 2

Suite à l'exécution de l'action n°3, l'affichage des équivalents français respecte l'ordre suivant : 1) chapeau noir (équiv 5)

- 2) black hat (équiv 3)
- 3) chapeau sorcier (équiv 2)
- 4) chapeau à sorcier (équiv 6)
- 5) Vide (équiv 4)
- Implanté 1
- Implanté 2

2.6. Schémas





Les noms composés anglais français et espagnols du domaine d'Internet :

protégé en vertu de la loi du droit d'auteur.

Annexe 12. La base de données du générateur d'équivalents

1. Les tables

Voici la liste des dix-neuf tables de la base de données du GE :

- ‘Équivalents français’. Elles sont au nombre de cinq, soit une table pour chacun des équivalents potentiels. Les champs qui composent chacune des cinq tables sont : numéro automatique, vedette anglaise, équivalent français, fréquence (de la structure de l’équivalent français), genre, nombre.
- ‘Structures compositionnelles’. Comme les tables ‘Équivalents français’, elles seront au nombre de cinq, une pour chaque équivalent potentiel. Ces cinq tables donneront accès aux statistiques permettant la formation d’un composé français lorsque la vedette anglaise est absente de la base de données. Les champs seront : numéro automatique, structure anglaise, structure française, fréquence (de la structure française).
- ‘Exceptions’. Cette table listera tous les composés anglais qui sont des exceptions au GE, c'est-à-dire ceux auxquels nous ne pouvons pas appliquer les statistiques (les

composés exocentriques comme EN *firewall*, *black hat*). Les champs seront : numéro automatique, vedette anglaise, équivalent français 1, genre 1, nombre 1, équivalent français 2, genre 2, nombre 2, emprunt.

- ‘Termes implantés’. Deux tables de ce type seront créées. Elles recenseront les termes français différents des équivalents créés grâce aux statistiques. Proposer les termes français employés est une nécessité puisque les termes fournis par le GE leur font parfois concurrence. Les champs seront : numéro automatique, terme implanté, genre et nombre.
- ‘Champs notionnels’. Cette table recensera les composés anglais qui ont un élément en commun (tous les composés contenant EN *mail* par exemple). Elle sera composée des champs : numéro automatique, élément commun et composés anglais.
- ‘Fréquence’. Cette table recensera les fréquences des structures des équivalents français. Utilisée dans une requête, elle permettra le classement par ordre décroissant des équivalents avant leur affichage dans le formulaire. Elle contiendra les champs : numéro automatique et fréquence.
- ‘Table générale’. Cette table recense les informations relatives à la vedette anglaise (existence de synonymes, de variantes, de sigles). Elle est composée des champs : numéro automatique, vedette anglaise, synonyme 1, synonyme 2, synonyme 3, variante 1, variante 2, variante 3, sigle (s'il existe plusieurs sigles tous seront inscrits dans la même case).
- ‘Vedette anglaise’. Cette mini-table contenant la liste des vedettes anglaises sera importée dans toutes les tables comprenant un champ ‘vedette anglaise’ dans le but de limiter les erreurs de saisies. Les champs seront : numéro automatique et vedette anglaise.

2. Les requêtes

2.1. Requêtes Sélection

La base de données du GE contiendra neuf requêtes Sélection. Elles permettront d'extraire les données des tables lorsque l'utilisateur du GE souhaitera connaître les termes contenant tel ou tel élément (*mail*), les équivalents d'une exception (*black hat*), le sigle correspondant à une vedette anglaise (*Denial of Service*), le synonyme d'une vedette anglaise (*message board*), les variantes d'une vedette anglaise (*email*), trouver une vedette anglaise d'après une variante (*email*), un synonyme (*bulletin board*) ou un sigle (*BBS*).

2.2. Requêtes Union

Annexe 12. La base de données du générateur d'équivalents

Les trois requêtes Union regrouperont les champs de deux à cinq tables dans un champ de résultats. Elles seront utilisées pour assembler les données d'au moins deux tables. Dans le GE, elles serviront à extraire les données relatives aux équivalents en langue cible, aux équivalences types et aux termes implantés, c'est-à-dire lorsque l'utilisateur interrogera le GE pour : connaître les équivalents en langue cible d'une vedette anglaise ; accéder aux structures compositionnelles de la langue cible permettant de rendre un type morphologique anglais ; disposer de la liste des termes de la langue cible implantés correspondant à une vedette anglaise.