

Université de Montréal

**Analyse comparative de l'équivalence  
terminologique en corpus parallèle et en corpus  
comparable :  
application au domaine du changement climatique**

par

Annaïch Le Serrec

Département de linguistique et de traduction

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)  
en traduction  
option terminologie

Avril, 2012

© Annaïch Le Serrec, 2012

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

Analyse comparative de l'équivalence terminologique en corpus parallèle et en  
corpus comparable : application au domaine du changement climatique

présentée par :  
Annaïch Le Serrec

a été évaluée par un jury composé des personnes suivantes :

Patrick Drouin, président-rapporteur  
Marie-Claude L'Homme, directrice de recherche  
Louise Dagenais, membre du jury  
Lynne Bowker, examinatrice externe  
Philippe Langlais, représentant du doyen de la FES

## RÉSUMÉ

Les travaux entrepris dans le cadre de la présente thèse portent sur l'analyse de l'équivalence terminologique en corpus parallèle et en corpus comparable. Plus spécifiquement, nous nous intéressons aux corpus de textes spécialisés appartenant au domaine du changement climatique. Une des originalités de cette étude réside dans l'analyse des équivalents de termes simples. Les bases théoriques sur lesquelles nous nous appuyons sont la terminologie textuelle (Bourigault et Slodzian 1999) et l'approche lexico-sémantique (L'Homme 2005).

Cette étude poursuit deux objectifs. Le premier est d'effectuer une analyse comparative de l'équivalence dans les deux types de corpus afin de vérifier si l'équivalence terminologique observable dans les corpus parallèles se distingue de celle que l'on trouve dans les corpus comparables. Le deuxième consiste à comparer dans le détail les équivalents associés à un même terme anglais, afin de les décrire et de les répertorier pour en dégager une typologie.

L'analyse détaillée des équivalents français de 343 termes anglais est menée à bien grâce à l'exploitation d'outils informatiques (extracteur de termes, aligneur de textes, etc.) et à la mise en place d'une méthodologie rigoureuse divisée en trois parties. La première partie qui est commune aux deux objectifs de la recherche concerne l'élaboration des corpus, la validation des termes anglais et le repérage des équivalents français dans les deux corpus. La deuxième partie décrit les critères sur lesquels nous nous appuyons pour comparer les équivalents des deux types de corpus. La troisième partie met en place la typologie des équivalents associés à un même terme anglais.

Les résultats pour le premier objectif montrent que sur les 343 termes anglais analysés, les termes présentant des équivalents critiquables dans les deux corpus sont relativement peu élevés (12), tandis que le nombre de termes présentant des similitudes d'équivalence entre les corpus est très élevé (272 équivalents identiques et 55 équivalents non critiquables). L'analyse comparative décrite dans ce chapitre confirme notre hypothèse selon laquelle la terminologie employée dans les corpus parallèles ne se démarque pas de celle des corpus comparables.

Les résultats pour le deuxième objectif montrent que de nombreux termes anglais sont rendus par plusieurs équivalents (70 % des termes analysés). Il est aussi constaté que ce ne sont pas les synonymes qui forment le groupe le plus important des équivalents, mais les quasi-synonymes. En outre, les équivalents appartenant à une autre partie du discours constituent une part importante des équivalents. Ainsi, la typologie élaborée dans cette thèse présente des mécanismes de l'équivalence terminologique peu décrits aussi systématiquement dans les travaux antérieurs.

**Mots-clés** : terminologie, équivalence, corpus parallèle, corpus comparable, extraction automatique de termes, changement climatique.

## ABSTRACT

The research undertaken for this thesis concerns the analysis of terminological equivalence in a parallel corpus and a comparable corpus. More specifically, we focus on specialized texts related to the domain of climate change. A unique aspect of this study is based on the analysis of the equivalents of single word terms. The theoretical frameworks on which we rely are the *terminologie textuelle* (Bourigault et Slodzian 1999) and the *lexico-sémantique* approaches (L'Homme 2005).

This study has two objectives. The first is to perform a comparative analysis of terminological equivalents in the two types of corpora in order to verify if the equivalents found in the parallel corpus are different from the ones observed in the comparable corpora. The second is to compare in detail equivalents associated with a same English term, in order to describe them and define a typology.

A detailed analysis of the French equivalents of 343 English terms is carried out with the help of computer tools (term extractor, text aligner, etc.) and the establishment of a rigorous methodology divided into three parts. The first part, common to both objectives of the research concerns the elaboration of the corpus, the validation of the English terms and the identification of the French equivalents in the two corpora. The second part describes the criteria on which we rely to compare the equivalents of the two types of corpora. The third part sets up the typology of equivalents associated with a same English term.

The results for the first objective shows that of the 343 English words analyzed, terms with equivalents that can be criticized in both corpora are relatively low in number (12), while the number of terms with similar equivalences between the two corpora is very high (272 identical and 55 equivalents not objectionable). The analysis described in this chapter confirms our hypothesis that terminology used in parallel corpora does not differ from that used in comparable corpora.

The results of the second objective show that many English terms are rendered by several equivalents (70% of analyzed terms). It is also noted that synonyms are not the largest group of equivalents but near-synonyms. Also, equivalents from another part of

speech constitute an important part of the equivalents analyzed. Thus, the typology developed in this thesis presents terminological equivalent mechanisms rarely described as systematically in previous work.

**Keywords:** terminology, equivalence, aligned corpora, comparable corpora, automatic term extraction, climate change.

# TABLE DES MATIÈRES

RÉSUMÉ .....	iii
ABSTRACT.....	v
TABLE DES MATIÈRES .....	vii
LISTE DES TABLEAUX.....	xiv
LISTE DES FIGURES .....	xviii
LISTE DES SYMBOLES, ABRÉVIATIONS ET CONVENTIONS TYPOGRAPHIQUES.....	xxi
REMERCIEMENTS .....	xxiv
1 INTRODUCTION .....	1
1.1 Problématique .....	1
1.2 Objectifs et hypothèse .....	3
1.3 Cadres théoriques .....	4
1.4 Plan de la thèse.....	6
CHAPITRES COMMUNS AUX DEUX VOLETS DE LA RECHERCHE .....	9
2 FONDEMENTS THÉORIQUES DE L'ÉQUIVALENCE.....	10
2.1 L'équivalence en traductologie .....	10
2.1.1 Typologies de l'équivalence en traductologie.....	11
2.1.2 Récapitulation.....	16
2.2 L'équivalence en lexicographie .....	16
2.2.1 Équivalence exacte .....	21
2.2.2 Équivalence partielle .....	22
2.2.2.1 Équivalence par supériorité ou par infériorité .....	22
2.2.2.2 Équivalence par intersection.....	24
2.2.2.3 Équivalence et disparité d'usage .....	24
2.2.3 Non-équivalence.....	24
2.2.4 Autres typologies de l'équivalence en lexicographie .....	25
2.2.4.1 Équivalence sémantique et équivalence pragmatique.....	25
2.2.4.2 Équivalence traductionnelle et équivalence explicative .....	26
2.2.5 Récapitulation.....	26
2.3 L'équivalence en terminologie.....	28

2.3.1	L'équivalence d'un point de vue conceptuel.....	31
2.3.1.1	Équivalence exacte.....	31
2.3.1.2	Équivalence partielle.....	32
2.3.1.2.1	Équivalence par supériorité ou par infériorité.....	32
2.3.1.2.2	Équivalence par intersection.....	34
2.3.1.2.3	Équivalence et disparité d'usage.....	35
2.3.1.3	Non-équivalence.....	35
2.3.2	Équivalence terminologique telle qu'observée en corpus.....	35
2.3.3	Équivalence terminologique d'un point de vue structurel.....	38
2.3.4	Traitement des termes polysémiques et des homonymes.....	40
2.3.5	Récapitulation.....	41
2.4	Conclusion du chapitre.....	41
3	Textes électroniques et corpus.....	43
3.1	Textes électroniques.....	43
3.2	Corpus.....	44
3.2.1	Corpus parallèles.....	45
3.2.2	Corpus comparables.....	47
3.3	Analyse des critiques touchant les corpus parallèles.....	48
3.3.1	Analyse du groupe A.....	49
3.3.2	Analyse du groupe B.....	51
3.3.3	Récapitulatif des analyses des groupes A et B.....	55
3.4	Conclusion de l'analyse critique.....	55
3.5	Conclusion du chapitre et contribution à la recherche.....	59
4	TECHNIQUES D'EXTRACTION LEXICALE ET TERMINOLOGIQUE.....	60
4.1	Extraction de termes en corpus monolingue.....	60
4.1.1	Approches linguistiques.....	61
4.1.2	Approches statistiques.....	63
4.2	Extraction des équivalents en corpus parallèles.....	63
4.2.1	Alignement de textes parallèles au niveau de la phrase.....	64
4.2.2	Extraction des équivalents en corpus parallèle.....	65
4.2.2.1	Extraction lexicale en corpus parallèle.....	67



4.2.2.2	Extraction de termes en corpus parallèle .....	69
4.3	Extraction des équivalents en corpus comparable.....	70
4.3.1	Type d'unité identifiée.....	73
4.3.2	Type et taille des corpus .....	74
4.3.3	Traitement du corpus .....	76
4.3.4	Liste de mots source à traduire.....	77
4.3.5	Construction des vecteurs de contexte.....	78
4.3.6	Pondération des cooccurrents.....	79
4.3.7	Dictionnaire pivot.....	80
4.3.8	Traduction des vecteurs de contexte.....	80
4.3.9	Comparaison des vecteurs.....	81
4.3.10	Stratégies pour améliorer l'approche par traduction directe .....	82
4.3.11	Validation.....	86
4.3.12	Résultats .....	87
4.3.13	Récapitulatif .....	88
4.4	Conclusion et contribution à la recherche.....	93
5	MÉTHODOLOGIE COMMUNE AUX DEUX VOLETS DE LA RECHERCHE .....	95
5.1	Constitution des corpus.....	95
5.1.1	Critères de sélection des textes.....	96
5.1.2	Prétraitement des textes .....	100
5.1.3	Noms des fichiers .....	103
5.1.4	Taille des corpus.....	104
5.1.5	Alignement du corpus parallèle au niveau des phrases .....	104
5.1.5.1	Description de l'aligneur d'Alinea .....	105
5.1.5.2	Résultats de l'alignement du corpus parallèle au niveau des phrases.....	106
5.2	Extraction des candidats termes .....	107
5.2.1	Fonctionnement de TermoStat .....	108
5.2.2	Résultats de l'extraction des CT anglais.....	109
5.3	Nettoyage de la liste d'extraction et validation des CT anglais .....	111
5.3.1	Première étape : nettoyage de la liste de candidats termes .....	111
5.3.2	Deuxième étape : validation des 527 CT .....	112

5.3.2.1	Critères d'identification des termes .....	112
5.3.2.2	Étapes de la validation.....	114
5.3.3	Résultats de la validation des termes anglais .....	116
5.4	Collecte des cooccurrents des termes anglais .....	117
5.5	Repérage et validation des équivalents dans le corpus parallèle.....	121
5.5.1	Repérage des équivalents dans le corpus parallèle .....	121
5.5.2	Validation des équivalents dans le corpus parallèle .....	124
5.6	Repérage et validation des équivalents dans le corpus comparable .....	127
5.6.1	Méthode de repérage et de validation des équivalents en corpus comparable	129
5.6.2	Exemples supplémentaires de recherches avancées des équivalents.....	133
5.7	Conclusion du chapitre .....	138
6.	RÉSULTATS DE LA MÉTHODOLOGIE COMMUNE AUX DEUX VOILETS DE LA RECHERCHE	142
6.1	Description détaillée de la liste de termes anglais.....	142
6.1.1	La répartition .....	142
6.1.2	Le score .....	148
6.1.3	La fréquence.....	150
6.2	Résultats de la validation des équivalents dans le corpus parallèle .....	152
6.2.1	Données globales .....	152
6.2.2	Nombre d'équivalents par termes anglais .....	152
6.2.3	Équivalents complexes et équivalents simples.....	154
6.2.4	Distribution des équivalents par partie du discours .....	155
6.3	Résultats de la validation des équivalents du corpus comparable.....	156
6.3.1	Données globales .....	156
6.3.2	Nombre d'équivalents par termes anglais .....	156
6.3.3	Équivalents complexes et équivalents simples.....	158
6.3.4	Distribution des équivalents par partie du discours .....	159
6.4	Conclusion du chapitre .....	159
PREMIER VOILET : ANALYSE COMPARATIVE DE L'ÉQUIVALENCE TERMINOLOGIQUE EN CORPUS PARALLÈLE ET EN CORPUS COMPARABLE .....		162
7	MÉTHODOLOGIE ET RÉSULTATS DE L'ANALYSE COMPARATIVE DE L'ÉQUIVALENCE EN CORPUS PARALLÈLE ET EN CORPUS COMPARABLE .....	163

7.1	Méthodologie de l'analyse comparative des équivalents du corpus parallèle aux équivalents du corpus comparable .....	163
7.1.1	Recensement des termes anglais ayant des équivalents identiques dans les deux corpus .....	164
7.1.2	Recensement et analyse des termes anglais ayant des différences d'équivalence dans les deux corpus.....	165
7.1.2.1	Collecte d'informations sur les équivalents différents.....	166
7.1.2.2	Analyse des équivalents différents .....	168
7.1.2.3	Notation des termes anglais ayant des équivalents différents .....	172
7.1.2.4	Classification des termes anglais en fonction du lieu (corpus) où se produit la différence d'équivalence.....	172
7.1.2.4.1	Différence d'équivalence unilatérale.....	173
7.1.2.4.2	Différence d'équivalence bilatérale.....	174
7.2	Résultats de l'analyse comparative des équivalents du corpus parallèle aux équivalents du corpus comparable .....	175
7.2.1	Analyse comparative quantitative globale .....	175
7.2.2	Résultats de l'analyse des termes anglais avec des équivalents identiques dans les deux corpus.....	180
7.2.3	Termes anglais avec des différences d'équivalence dans les deux corpus .....	184
7.2.3.1	Résultats de l'analyse des termes anglais avec des équivalents différents non critiquables dans le CP ou le CC ( $\pm$ ) .....	187
7.2.3.2	Analyse des termes anglais avec équivalent différent critiquable dans le CP ( $\neq$ CP) .....	189
7.2.3.3	Résultats de l'analyse des termes anglais avec équivalent différent critiquable dans le CC ( $\neq$ CC).....	191
7.3	Conclusion du chapitre .....	196
	DEUXIÈME VOLET : ANALYSE COMPARATIVE DES ÉQUIVALENTS FRANÇAIS ASSOCIÉS À UN MÊME TERME ANGLAIS .....	199
8	MÉTHODOLOGIE ET RÉSULTATS DE L'ANALYSE COMPARATIVE DES ÉQUIVALENTS FRANÇAIS ASSOCIÉS À UN MÊME TERME ANGLAIS .....	200

8.1	Méthodologie de l'analyse comparative des équivalents associés à un même terme anglais .....	200
8.1.1	Équivalent de la même partie du discours que l'ÉqT .....	203
8.1.1.1	Équivalents synonymes .....	203
8.1.1.2	Équivalent quasi-synonymes par rapport à l'ÉqT.....	205
8.1.1.3	Équivalent hyperonyme par rapport à l'ÉqT .....	206
8.1.1.4	Équivalent hyponyme par rapport à l'ÉqT .....	208
8.1.1.5	Équivalent holonyme par rapport à l'ÉqT .....	209
8.1.1.6	Équivalent méronyme par rapport à l'ÉqT .....	209
8.1.1.7	Équivalent dont le système conceptuel présente une divergence par rapport à celui de l'autre langue.....	210
8.1.1.8	Équivalent présentant un autre point de vue par rapport à l'ÉqT .....	212
8.1.2	Équivalent transcatégoriel par rapport à l'ÉqT .....	214
8.1.2.1	Équivalent transcatégoriel par rapport à l'ÉqT et morphologiquement relié à celui-ci.....	214
8.1.2.2	Équivalent transcatégoriel par rapport à l'ÉqT, mais non morphologiquement relié à l'ÉqT .....	217
8.2	Résultats de l'analyse comparative des équivalents associés à un même terme anglais .....	218
8.3	Autres observations sur l'équivalence .....	222
8.3.1	ÉqT français dont la partie du discours est différente de celle du terme anglais . .....	222
8.3.2	Équivalents des termes anglais adverbes .....	223
8.3.3	Équivalents des éléments et composés chimiques.....	224
8.4	Conclusion du chapitre .....	225
9	CONCLUSION GÉNÉRALE.....	229
	BIBLIOGRAPHIE.....	234
	ANNEXES.....	I
	Annexe A : Bibliographie des corpus .....	I
	Annexe B : Liste des 343 termes anglais validés.....	VIII
	Annexe C : Liste des ressources terminologiques et autres dictionnaires.....	XVII

Annexe D : Liste des équivalents complexes du corpus parallèle et du corpus comparable .....	XX
Annexe E : Liste des 272 termes anglais ayant des équivalents identiques dans les deux corpus .....	XXIII
Annexe F : Analyse des 67 termes anglais avec des différences d'équivalence ....	XXXVII

## LISTE DES TABLEAUX

Tableau 2.1. Liste partielle des unités lexicales proposées par Liberman (2004). .....	18
Tableau 2.2. Exemple d'équivalence partielle ( <i>singe</i> ).....	22
Tableau 2.3. Exemple d'équivalence partielle ( <i>amour</i> ).....	23
Tableau 2.4. Exemple d'équivalent traductionnel et d'équivalent explicatif.....	26
Tableau 2.5. Exemples de transformations du patron français NOM + prép/dét + NOM vers l'anglais .....	39
Tableau 2.6. Différents degrés de cognation .....	40
Tableau 2.7. Représentation de la forme graphique <i>flood</i> et de ses équivalents .....	41
Tableau 3.1. Catégories offrant un jugement de valeur des groupes A et B .....	55
Tableau 4.1. <i>Is projected to increase</i> et quelques-uns de ses équivalents possibles .....	67
Tableau 4.2. Quelques résultats d'extraction sur corpus comparables .....	88
Tableau 4.3. Tableau récapitulatif des étapes d'extraction sur corpus comparable .....	89
Tableau 5.1. Types de document dans les deux corpus .....	98
Tableau 5.2. Analyse comparative des années de parution des documents des deux corpus .....	99
Tableau 5.3. Abréviation et nom des organismes apparaissant dans les deux corpus .....	103
Tableau 5.4. Nombre et pourcentage de CT par parties du discours .....	111
Tableau 5.5. La fiche du terme <i>permafrost</i> accompagnée de ses informations .....	115
Tableau 5.6. Données chiffrées de la validation des 343 termes anglais .....	117
Tableau 5.7. La fiche du terme <i>permafrost</i> accompagnée de toutes ses informations.....	121
Tableau 5.8. Exemple de coquilles ou d'erreurs dans le corpus parallèle .....	124
Tableau 5.9. L'homographe sémantique anglais <i>plant</i> et ses candidats équivalents français .....	125
Tableau 5.10. Équivalents validés pour le terme anglais <i>plant</i> -NOM ('végétal') .....	126
Tableau 5.11. Exemples de coquilles et d'erreurs dans le corpus comparable .....	128
Tableau 5.12. Fiche du terme anglais <i>uptake</i> -NOM .....	129
Tableau 5.13. Équivalents validés pour le terme anglais <i>uptake</i> -NOM.....	133
Tableau 5.14. Exemples d'équivalents validés dans le corpus comparable .....	134
Tableau 6.1. Les 5 termes présents dans les 38 textes .....	146

Tableau 6.2. Les 32 termes de la colonne 8 de la Figure 6.2 .....	147
Tableau 6.3. Les 14 termes de la colonne 22 de la Figure 6.2 .....	147
Tableau 6.4. Nombre de termes validés en fonction du rang (par tranche de 200) .....	149
Tableau 6.5. Les 5 termes les plus fréquents et les 5 termes les moins fréquents.....	152
Tableau 6.6. Nombre de termes par nombre d'équivalents dans le corpus parallèle .....	153
Tableau 6.7. Répartition des équivalents par ÉqT et ÉqNT dans le corpus parallèle.....	155
Tableau 6.8. Nombre d'équivalents dans le corpus parallèle par partie du discours et en pourcentage.....	155
Tableau 6.9. Résultats de la validation des équivalents dans le corpus comparable .....	157
Tableau 6.10. Répartition des équivalents par ÉqT et ÉqNT dans le corpus comparable ..	159
Tableau 6.11. Pourcentage en parties du discours des équivalents .....	159
Tableau 7.1. Cinq termes anglais et leurs équivalents dans le CP et le CC .....	165
Tableau 7.2. Exemples d'ÉqT ( <i>région boisée</i> -ADJ et <i>surface boisée</i> -NOM) .....	166
Tableau 7.3. Exemple d'ÉqNT ( <i>préserver</i> -VER).....	167
Tableau 7.4. Exemple d'ÉqPF ( <i>tourbeux</i> -ADJ) .....	167
Tableau 7.5. Exemple d'ÉqF ( <i>capture</i> -NOM) .....	168
Tableau 7.6. Exemple d'emprunt fautif trouvé dans le CC ( <i>mitigation</i> -NOM) .....	169
Tableau 7.7. Exemples d'emprunts fautifs trouvés dans le CP ( <i>anthropogène</i> -ADJ, <i>anthropogénique</i> -ADJ).....	169
Tableau 7.8. Exemple de faux sens ( <i>refléter</i> -VER).....	170
Tableau 7.9. Exemple de cognat critiquable ( <i>permafrost</i> -NOM).....	171
Tableau 7.10. Exemple de cognat critiquable ( <i>capture</i> -NOM).....	171
Tableau 7.11. Résultat de l'analyse comparative quantitative .....	177
Tableau 7.12. Équivalents absents du corpus comparable .....	177
Tableau 7.13. Répartition des équivalents par ÉqT et ÉqNT dans les deux corpus .....	178
Tableau 7.14. Répartition des termes anglais par parties du discours .....	178
Tableau 7.15. Répartition des termes anglais par critère de sélection a), b), c) et d) .....	179
Tableau 7.16. Termes anglais et distribution par parties du discours et par nombre d'équivalents.....	181
Tableau 7.17. Les 67 équivalents différents répartis par ±, ≠CP, ≠CC .....	184
Tableau 7.18. Pourcentage des équivalents différents sur les 339 termes anglais .....	184

Tableau 7.19. Répartition des 67 termes par partie du discours .....	186
Tableau 7.20. Équivalents classés en fonction des critères ÉqNT et ÉqT .....	188
Tableau 7.21. Informations générales sur les ≠CP .....	190
Tableau 7.22. Informations générales sur les ≠CC .....	191
Tableau 8.1. Exemples de termes anglais et de leurs équivalents .....	201
Tableau 8.2. Typologie des équivalents français associés à un même terme anglais .....	202
Tableau 8.3. Quelques exemples de synonymes relevés dans le corpus.....	203
Tableau 8.4. Exemples de quasi-synonymes tirés du corpus .....	205
Tableau 8.5. Exemples d'hyperonymes trouvés dans le corpus .....	207
Tableau 8.6. Exemples d'hyponymes trouvés dans le corpus .....	208
Tableau 8.7. Exemples d'holonymes trouvés dans le corpus.....	209
Tableau 8.8. Trois exemples de méronymes relevés dans le corpus .....	210
Tableau 8.9. Trois exemples de termes anglais dont le système conceptuel présente une divergence par rapport à celui de l'autre langue .....	211
Tableau 8.10. Exemples d'équivalents présentant un point de vue différent par rapport à l'ÉqT.....	212
Tableau 8.11. Exemples d'équivalents transcatégoriels par rapport à l'ÉqT et morphologiquement reliés à celui-ci .....	214
Tableau 8.12. Exemples d'équivalents transcatégoriels par rapport à l'ÉqT .....	217
Tableau 8.13. Répartition des ÉqT et des ÉqNT .....	219
Tableau 8.14. Résultats de l'analyse des équivalents .....	220
Tableau 8.15. Répartition des ÉqT associés à des équivalents transcatégoriels.....	220
Tableau 8.16. Nombre de passage d'une partie du discours à une autre .....	221
Tableau 8.17. Typologie des équivalents transcatégoriels .....	222
Tableau 8.18. ÉqT dont la partie du discours est différente du terme anglais .....	223
Tableau 8.19. Équivalents de 7 termes anglais adverbes .....	223
Tableau 8.20. Éléments et composés chimiques et équivalence .....	225
Tableau D.1. Équivalents complexes du corpus parallèle.....	XX
Tableau D.2. Équivalents complexes du corpus comparable .....	XXI
Tableau F.1. Terme anglais avec différence d'équivalence peu marquée (±).....	XXXVII
Tableau F.2. Équivalents dont la différence est marquée dans le CP (≠CP).....	XLVI



Tableau F.3. Équivalents dont la différence est marquée dans le CC ( $\neq$ CC).....XLVII

## LISTE DES FIGURES

Figure 2.1. Spectre des couleurs illustrant le non-isomorphisme (d'après Lyons 1970 : 46)	17
Figure 2.2. Système conceptuel dont la structure française diffère de celle de l'anglais ....	19
Figure 2.3. Équivalence exacte, équivalence partielle et non-équivalence .....	21
Figure 2.4. Équivalence par supériorité ou par infériorité .....	22
Figure 2.5. Convergence (de la langue A vers la langue B) (d'après Svensén 2009 : 259).	23
Figure 2.6. Divergence (de la langue A vers la langue B) (d'après Svensén 2009 : 259) ...	23
Figure 2.7. Équivalence par intersection.....	24
Figure 2.8. Équivalence par supériorité ou par infériorité .....	33
Figure 2.9. Illustration des champs sémantiques de <i>nuts</i> et de <i>noix</i> .....	33
Figure 2.10. Deuxième exemple d'équivalence par supériorité ou par infériorité (d'après Van Campenhoudt 1996 : 289).....	34
Figure 2.11. Équivalence par intersection (d'après Felber 1987 : 129).....	34
Figure 4.1. Exemple de non-correspondance entre mots de textes source et cible .....	66
Figure 4.2. Illustration du fonctionnement prototypique de l'extraction sur corpus comparable.....	72
Figure 5.1. Corpus parallèle et corpus comparables .....	96
Figure 5.2. Exemple de coquilles apparaissant lors de la conversion au format texte .....	101
Figure 5.3. Illustration du problème posé par les en-têtes et les numéros de page .....	102
Figure 5.4. Taille des sous-corpus .....	104
Figure 5.5. Échantillon au format HTML de l'alignement au niveau phrastique.....	107
Figure 5.6. Échantillon de page html montrant des résultats de l'extraction avec TermoStat .....	109
Figure 5.7. Image agrandie des onglets de la page des résultats html de TermoStat .....	110
Figure 5.8. Extraction des cooccurrents du terme anglais <i>oscillation</i> dans l'onglet <i>Collocates</i> d'AntConc .....	119
Figure 5.9. Recherche des cooccurrents de l'homographe sémantique <i>current</i> dans le concordancier d'AntConc .....	120
Figure 5.10. Repérage automatique des équivalents en corpus parallèle.....	122

Figure 5.11. Échantillon de l'extraction des correspondances lexicales avec Alinea .....	123
Figure 5.12. Repérage semi-manuel des équivalents en corpus parallèle.....	128
Figure 6.1. Calcul de la répartition du terme <i>biota</i> dans AnctConc .....	143
Figure 6.2. Répartition des termes anglais en fonction du nombre de textes.....	145
Figure 6.3. Nombre de termes validés en fonction du rang (score) attribué par TermoStat .....	150
Figure 6.4. Fréquence des termes validés .....	151
Figure 6.5. Nombre de termes en fonction du nombre d'équivalents (corpus parallèle)....	154
Figure 6.6. Nombre de termes en fonction du nombre d'équivalents (corpus comparable) .....	158
Figure 7.1. Analyse comparative des équivalents identifiés dans les deux corpus .....	164
Figure 7.2. Différence d'équivalence unilatérale avec un équivalent différent par la forme ( <i>glaciel</i> -ADJ) .....	173
Figure 7.3. Différence d'équivalence unilatérale avec deux cognats critiquables ( <i>mitigation</i> - NOM et <i>mitiger</i> -VER).....	173
Figure 7.4. Différence d'équivalence bilatérale avec des équivalents différents par la forme ( <i>houiller</i> -NOM et <i>charbonnier</i> -NOM).....	174
Figure 7.5. Différence d'équivalence bilatérale avec équivalent différent par la forme ( <i>pergélisolé</i> -ADJ) et cognat critiquable ( <i>permafrost</i> -NOM).....	174
Figure 7.6. Nombre de termes en fonction du nombre d'équivalents (corpus parallèle et corpus comparable) .....	176
Figure 7.7. Termes avec équivalents identiques dans le CP et le CC .....	181
Figure 7.8. Répartition des termes anglais en fonction de leur partie du discours et de leur nombre d'équivalents .....	182
Figure 7.9. Répartition des termes anglais avec équivalents identiques en fonction du critère de validation.....	183
Figure 7.10a. Catégorisation des termes avec équivalents différents en fonction de la partie du discours (corpus parallèle). .....	185
Figure 7.10b. Catégorisation des termes avec équivalents différents en fonction de la partie du discours (corpus comparable). .....	186

Figure 7.11. Répartition des termes avec équivalents différents du CP et du CC en fonction du critère de validation .....	187
--	-----

# LISTE DES SYMBOLES, ABRÉVIATIONS ET CONVENTIONS

## TYPOGRAPHIQUES

±	terme anglais avec équivalent différent non critiquable dans le CP ou le CC
≠CP	terme anglais avec équivalent différent critiquable dans le CP
≠CC	terme anglais avec équivalent différent critiquable dans le CC
ADJ	adjectif
ADV	adverbe
AN	anglais
CC	corpus comparable
CE	candidat équivalent
CP	corpus parallèle
CT	candidat terme
ÉqF	équivalent fréquent (60 et plus occurrences)
ÉqPF	équivalent peu fréquent (15 et moins occurrences)
ÉqT	équivalent typique du terme anglais
ÉqNT	équivalent non typique du terme anglais
FR	français
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
LOC	locution
NOM	nom
TleCP	occurrences dans tout le corpus parallèle
VER	verbe

**Conventions graphiques**

- Les parenthèses en exposant indiquent le sens d'un terme, par exemple : *wood-NOM* (matière ligneuse)
- Dans les contextes tirés du corpus, le terme qui fait l'objet d'une analyse apparaît en lettres majuscules (ex. Les aérosols ANTHROPIQUES formés de poussières sont également mal quantifiés).

*Je dédie cette thèse à la mémoire de mon père et de ma mère.*

## REMERCIEMENTS

Je souhaite tout d'abord remercier ma directrice de thèse, Marie-Claude L'Homme, pour l'intérêt qu'elle a porté à mon travail et pour sa constante disponibilité. Par les nombreuses lectures et relectures de ma thèse, j'ai pu profiter de sa très précieuse expérience. Enfin, je la remercie tout particulièrement d'avoir toujours su m'encadrer et me conseiller tout en me laissant libre de mes choix.

Je tiens également à offrir mes sincères remerciements à Patrick Drouin qui a eu la gentillesse d'accepter le rôle de président-rapporteur. Je suis aussi redevable à Louise Dagenais et Lynne Bowker d'avoir accepté de faire partie de mon jury. Leurs remarques et commentaires me sont précieux.

J'aimerais remercier l'équipe du laboratoire de l'Observatoire de linguistique Sens-Texte (OLST) de l'université de Montréal de m'avoir accueillie. Au sein de ce laboratoire, j'ai beaucoup appris. J'adresse des remerciements aux doctorants avec qui j'ai eu le plaisir de travailler ou de discuter au fil des années, Iveth Carreño, Janine Pimentel, Daniel Labonia, Elizabeth Marshman, Sébastien Marengo. Je pense en outre à tous les autres membres de l'équipe de l'OLST, dont Benoît Robichaud et Marie-Éve Laneville.

Je tiens à remercier le Fonds québécois de la recherche sur la société et la culture (FQRSC) et le Programme de renforcement du secteur langagier au Canada. Je leur suis redevable du soutien financier qu'ils m'ont accordé pour réaliser mon projet de recherche. Je suis également redevable à l'Université de Montréal pour les bourses qu'elle m'a octroyées durant toute ma scolarité.

Je remercie ma famille et mes amis pour leur soutien moral. Je remercie tout particulièrement Georges pour sa patience et son appui inconditionnel. Ensemble, nous avons vécu ces années d'études comme une aventure passionnante et enrichissante.



# 1 INTRODUCTION

Notre travail porte sur l'équivalence terminologique en corpus et se divise en deux volets. Le premier fait une étude comparative des équivalents français de plus de 300 termes simples anglais en corpus parallèle et en corpus comparable. Le deuxième vise à comparer les équivalents associés à un même terme anglais et à les décrire sous un jour nouveau, c'est-à-dire en fonction de ce que nous aurons observé en corpus.

Par cette thèse, nous espérons contribuer à l'avancement des connaissances sur l'équivalence en terminologie afin de présenter aux terminologues, aux terminographes, aux traductologues et aux spécialistes en traitement automatique du langage (TAL) des pistes de réflexion sur l'équivalence terminologique.

## 1.1 Problématique

Le nombre de plus en plus important de textes électroniques disponibles sur Internet constitue une source riche de données linguistiques. Pour exploiter efficacement cette masse grandissante de documents, il est nécessaire de la traiter partiellement ou entièrement à l'aide d'outils informatiques. Les lexiques et les terminologies bilingues extraits automatiquement figurent parmi les types de données linguistiques très recherchées. Ces ressources sont de plus en plus employées dans l'élaboration de dictionnaires, de bases de données terminologiques et d'ontologies. De plus, elles sont largement utilisées en traduction automatique. Toutefois, certaines techniques d'extraction bilingue ne sont pas encore suffisamment au point, notamment lorsqu'elles sont réalisées à partir de corpus comparables (ensemble de textes originaux traitant d'un même domaine dans deux langues). Or, nous pensons qu'une connaissance approfondie de l'équivalence terminologique en corpus permettrait d'améliorer ces techniques d'extraction, sans oublier que d'autres domaines d'applications, telle la terminologie, pourraient également en tirer profit.

L'équivalence terminologique n'est pas un sujet inconnu pour nous. Dans des études précédentes (Le Serrec 2008; Le Serrec *et al.* 2010), nous avons étudié dans un corpus parallèle (ensemble de textes sources et de textes cibles) les équivalents anglais de

50 termes simples français. Ces termes appartenaient aux parties du discours du nom et de l'adjectif<sup>1</sup>. L'analyse a révélé des mécanismes de l'équivalence peu décrits en terminologie classique. Nous avons observé, entre autres, que de nombreux termes étaient rendus par plusieurs équivalents. Au terme de ce travail, nous en sommes arrivée à la conclusion que le sujet méritait d'être étudié plus en détail et à plus grande échelle afin d'analyser les différents équivalents associés à un même terme anglais.

Pour ce type d'étude, les corpus parallèles sont particulièrement intéressants sur plusieurs plans. Tout d'abord, la similarité du contenu est assurée puisqu'ils sont constitués de textes originaux et de leur traduction. Ensuite, l'alignement des textes phrase à phrase facilite la fouille du corpus. En effet, depuis le début des années 1980, les techniques d'alignement sont arrivées à maturité et permettent d'aligner des textes avec un haut degré de précision – dans les meilleurs des cas on obtient plus de 98,5 % (Véronis 2000 : 161). Enfin, l'extraction bilingue de termes simples sur corpus parallèle donne des résultats<sup>2</sup> satisfaisants – 90 % d'appariements corrects pour 88 % des unités possédant des correspondances sur un corpus français anglais (Kraif et Chen 2004).

Toutefois, l'utilisation des corpus parallèles est critiquée par de nombreux auteurs (Déjean et Gaussier 2002; Dubuc 2002; McEnery et Xiao 2007). Ces auteurs soutiennent que la langue produite dans les textes traduits n'est pas aussi idiomatique que celle des textes originaux, car, disent-ils, le traducteur est inévitablement influencé par la langue source. Par conséquent, ils préconisent l'emploi de corpus comparables dans lesquels il n'y aurait pas de traduction pour étudier et extraire les équivalents des termes, étant donné que ces corpus présentent l'avantage d'être composés de textes originaux traitant d'un même domaine dans deux langues. En outre, il est fréquemment souligné que les corpus comparables pour des couples de langues ne faisant pas intervenir l'anglais sont plus faciles à constituer que les corpus parallèles (Morin *et al.* 2004). Cela dit, les corpus comparables posent un problème majeur : ils sont difficiles à exploiter dans une perspective d'extraction bilingue. Tout d'abord, comme il n'est pas possible de les aligner phrase à phrase, la fouille

---

<sup>1</sup> Les verbes et les adverbes ne faisaient pas partie des termes étudiés, car aucun n'était présent parmi les 50 premiers termes validés de la liste d'extraction.

<sup>2</sup> Les techniques d'extraction bilingue de termes sont expliquées plus en détail dans le Chapitre 4.

du corpus en est complexifiée. Ensuite, l'extraction bilingue de termes, même simples, ne donne pas encore des résultats exploitables par les terminologues ou les traducteurs.

## 1.2 Objectifs et hypothèse

Devant les critiques négatives exprimées à l'endroit des corpus parallèles, il nous paraît maintenant essentiel de vérifier jusqu'à quel point l'équivalence terminologique observée dans les traductions des corpus parallèles se distingue de celle que l'on trouve dans les corpus comparables. Par conséquent, cette étude qui au départ n'avait qu'un objectif, à savoir, comparer les équivalents associés à un même terme et relevés dans un corpus, s'est tout naturellement divisée en deux volets.

1) Dans le premier, nous effectuons une analyse comparative de l'équivalence dans les deux types de corpus afin de répondre à la question suivante : est-ce que l'équivalence terminologique observable dans les corpus parallèles se distingue de celle observée dans les corpus comparables?

2) Dans le deuxième, nous comparons en détail les équivalents associés à un terme anglais, afin de les répertorier et de les décrire pour en dégager une typologie.

En ce qui concerne la question du premier volet, à l'instar de Baker (1993 : 233–234), nous croyons que les traductions jouent un rôle important dans notre façon d'appréhender et de comprendre le monde et qu'elles ne sont ni inférieures ni supérieures aux textes originaux. Les traductions font tout autant partie du langage naturel que les originaux (Mauranen 2002). Plus spécifiquement, en ce qui concerne la terminologie employée dans les traductions françaises de textes spécialisés, nous émettons l'hypothèse qu'elle ne se démarque pas autant qu'on le suppose généralement de celle des textes originaux français. Effectivement, dans de nombreux domaines de spécialité, afin de produire des textes originaux, les auteurs s'appuient bien souvent sur des ouvrages écrits dans une autre langue que la leur, notamment l'anglais, comme l'attestent les bibliographies. Autrement dit, la rédaction de textes spécialisés implique inévitablement une part de traduction, même dans les textes dits *originaux*. Le traducteur, en raison de son métier, aura tendance à être plus conscient des interférences entre langues que les auteurs

de textes originaux et tentera, selon nous, d'éviter certains pièges dans ses choix terminologiques.

Dans le deuxième volet de notre thèse, notre objectif est de comparer les équivalents d'un même terme anglais. Pour cette analyse, nous tiendrons compte de ce que nous aurons observé dans les deux types de corpus sans les mettre en opposition. Une fois que nous aurons validé dans les corpus français les équivalents de plus de 300 termes anglais, nous proposerons une classification de l'équivalence qui permettra d'en montrer la richesse. Nous pensons qu'une étude comme la nôtre mettra en lumière des phénomènes peu décrits jusqu'ici.

Afin de réaliser nos deux objectifs, nous nous proposons d'étudier l'équivalence terminologique dans des textes spécialisés appartenant au domaine du changement climatique. L'analyse sera unidirectionnelle et se fera de l'anglais vers le français sur des termes simples appartenant aux parties du discours du nom, de l'adjectif, du verbe ou de l'adverbe.

### **1.3 Cadres théoriques**

Avant d'annoncer les cadres théoriques sur lesquels nous nous appuyons pour mettre en place une méthodologie qui réponde à nos objectifs, il convient tout d'abord de dresser un portrait succinct de l'évolution récente de la terminologie afin de mieux situer notre choix.

La terminologie, discipline au carrefour de la linguistique, de la logique, de l'ontologie, de l'informatique et des sciences des choses (Wüster : 1981), suscite depuis le début du 20<sup>e</sup> siècle l'intérêt grandissant des chercheurs. Certains, à l'exemple de Wüster, entrevoient la terminologie comme une discipline autonome dont l'objet central est le concept. D'après Wüster, le concept scientifique est universel et ne dépend ni des cultures, ni des langues (*ibid.* : 194). Sous cette perspective, en terminographie, la démarche onomasiologique<sup>3</sup> est largement privilégiée et l'utilisation des textes pour constituer des ressources terminologiques n'est pas considérée comme essentielle. De plus, étant donné que l'on décrit des réalités concrètes ou des concepts abstraits, les termes de nature nominale sont largement privilégiés. Enfin, pour les tenants de la Théorie générale de la

---

<sup>3</sup> Démarche onomasiologique : on part de la notion pour aller vers le terme.

terminologie (TGT)<sup>4</sup>, la terminologie se veut, dans une large mesure, un instrument de normalisation des termes afin de faciliter la communication entre experts.

Toutefois, certains chercheurs (Bourigault et Slodzian 1999; Cabré 2000; Diki-Kidiri 2008; Gaudin 1993; L'Homme 2004; Temmerman 2000) considèrent ces points de vue idéalistes et peu applicables. Ces auteurs soutiennent que l'objet d'étude de la terminologie est le terme et préconisent une démarche sémasiologique<sup>5</sup> pour l'étudier. Selon ces chercheurs, les termes en tant que signes linguistiques sont des unités du langage naturel faisant partie du vocabulaire de la langue qui subissent « les contraintes syntaxiques et discursives de l'énonciation » (Ramavonirina *et al.* 2007 : 40).

Ce changement de position est en grande partie attribuable à l'apparition massive des documents sous format numérique. La possibilité d'interroger, de consulter et d'analyser des quantités de données à l'aide d'outils informatiques a permis de mettre en lumière des phénomènes largement sous-estimés jusqu'ici (Bourigault et Slodzian 1999 : 29). De ce fait, en terminologie, la fouille de corpus a eu pour effet de largement modifier la pratique. À tel point que, selon plusieurs auteurs, les méthodologies s'apparentent maintenant de plus en plus à la lexicographie (Cabré 2005) ou à la linguistique de corpus (Condamines 2005).

On aura compris que notre étude ne peut prendre appui sur la Théorie générale de la terminologie, car notre approche fait appel aux corpus et s'intéresse aux termes simples appartenant à plusieurs parties du discours. Les modèles que nous avons adoptés sont la terminologie textuelle (Bourigault et Slodzian 1999) et l'optique lexico-sémantique (L'Homme 2005). Selon la terminologie textuelle, « le texte est le point de départ de la description lexicale à construire » (*ibid.* : 31). Ainsi, le corpus constitue son objet d'étude (Condamines 2005) dans lequel le terme est « observé et analysé en contexte discursif » (Cabré 2005 : 206). La terminologie textuelle incorpore des démarches linguistiques en s'appuyant notamment sur la linguistique de corpus en tant qu'instrument méthodologique. Elle considère le terme comme un construit choisi par le linguiste-terminologue en fonction de la place qu'il occupe dans le texte et de sa pertinence par rapport au projet

---

<sup>4</sup> Théorie inspirée des travaux de Wüster (1981).

<sup>5</sup> Démarche sémasiologique : on part du terme pour aller vers la notion.

terminologique visé (Bourigault et Slodzian 1999). De son côté, l'optique lexico-sémantique (L'Homme 2005) considère également le texte comme point de départ pour sélectionner les termes et recueillir des renseignements sur ces termes afin de les décrire. Selon cette optique, les termes sont des unités lexicales ayant un sens spécialisé, plus précisément, ils dépendent « de la délimitation qu'on a faite d'un domaine spécialisé » (L'Homme 2004 : 33). Par ailleurs, l'approche lexico-sémantique accueille les termes appartenant aux parties du discours du nom, de l'adjectif, du verbe et de l'adverbe et envisage les termes simples tout comme les termes complexes.

## **1.4 Plan de la thèse**

Nous avons organisé notre thèse en trois parties principales. La première est constituée par les chapitres communs aux deux volets de la recherche. La deuxième est consacrée au premier volet de notre recherche, soit l'analyse comparative de l'équivalence. La troisième partie présente le deuxième volet, c'est-à-dire la comparaison des équivalents français associés à un même terme anglais.

### *Chapitres communs aux deux volets de la recherche*

Les Chapitres 2, 3 et 4 font la synthèse des principaux travaux portant sur les aspects reliés à cette étude : équivalence, corpus et outils informatiques.

Ainsi, le Chapitre 2, qui se divise en trois grandes sections, se penche sur les fondements théoriques de l'équivalence. Dans la première section, nous faisons un rapide tour d'horizon de l'équivalence en traductologie, en nous intéressant principalement à l'équivalence envisagée par les théoriciens se fondant sur une approche linguistique de la traduction. Dans la deuxième section, nous abordons l'équivalence en lexicographie. Dans la troisième section, nous nous penchons sur l'équivalence en terminologie, notamment l'équivalence conceptuelle, l'équivalence en corpus et l'équivalence structurelle.

Dans le Chapitre 3, nous définissons et décrivons les corpus parallèles et les corpus comparables, afin de mieux élaborer nos corpus. Ensuite, nous effectuons une analyse des critiques touchant les corpus parallèles. Puis, nous poursuivons par une analyse dans

laquelle nous exposons notre point de vue sur les questions entourant les corpus parallèles et les corpus comparables.

Le Chapitre 4 est consacré à l'extraction automatique de lexiques<sup>6</sup> et de terminologies. La première partie traite de l'extraction des termes dans les corpus monolingues. En deuxième partie, nous décrivons l'extraction bilingue lexicale ou terminologique sur corpus parallèles. Enfin, dans une troisième partie, nous effectuons une description détaillée de l'extraction bilingue à partir des corpus comparables. Cette troisième partie est beaucoup plus développée que les deux autres étant donné que l'extraction à partir de ce type de corpus pose encore de nombreuses difficultés (*cf.* 1.1). Par cette description, nous visons à bien circonscrire les points sur lesquels notre étude serait susceptible de contribuer à la recherche en extraction automatique sur corpus comparable.

Le Chapitre 5 présente en six sections la méthodologie commune aux deux volets de la recherche. La première décrit la constitution des corpus et l'alignement du corpus parallèle (5.1). Les sections 5.2 et 5.3 expliquent respectivement le mode d'extraction des candidats terme anglais (CT) et leur validation. La section 5.4 présente la collecte des cooccurrents des termes anglais. Enfin, les sections 5.5 et 5.6 sont dédiées à l'identification des équivalents dans le corpus parallèle et le corpus comparable.

Le Chapitre 6 vient clore la section commune aux deux volets de la recherche. Afin de mieux connaître la nature des termes anglais analysés, nous présentons dans le détail leur répartition, leur score et leur fréquence (6.1). Puis, les résultats de la validation des équivalents dans le corpus parallèle et dans le corpus comparable font l'objet des sections 6.2 et 6.3.

#### *Chapitre consacré au premier volet de la recherche*

Le Chapitre 7 consacré au premier volet de la recherche se divise en trois sections principales. La première (7.1) expose la méthodologie adoptée pour comparer les équivalents du corpus parallèle aux équivalents du corpus comparable. La deuxième (7.2)

---

<sup>6</sup> Extraction de toutes les unités lexicales simples d'un texte.

présente les résultats de l'analyse comparative répartis en trois sous-sections. La sous-section 7.2.1 montre des résultats quantitatifs généraux. La sous-section 7.2.2 traite des termes anglais ayant des équivalents identiques dans les deux corpus et la sous-section 7.2.3 porte sur les résultats de l'analyse comparative des termes anglais dont les équivalents sont différents dans les deux corpus. Enfin, la section 7.3 résume les points saillants de ce volet de la recherche.

### *Chapitre du deuxième volet de la recherche*

Le Chapitre 8 décrit la méthodologie élaborée pour comparer les équivalents associés à un même terme anglais et présente les résultats de cette comparaison. Dans notre méthodologie (section 8.1), nous expliquons tout d'abord la terminologie qui nous sert de point de comparaison, puis nous décrivons la typologie permettant de caractériser les divers équivalents associés à un même terme anglais – les équivalents sur lesquels porte l'analyse sont validés au Chapitre 5. La section 8.2 concerne les résultats de l'analyse de ces équivalents. La section 8.3 traite de certains phénomènes de l'équivalence non examinés dans les sections précédentes.

Enfin, le Chapitre 9 vient clore cette étude en offrant une conclusion générale. Nous présentons une synthèse des points théoriques et applicatifs abordés dans cette thèse et mettons de l'avant notre contribution dans le domaine de l'équivalence. Nous dégageons également des pistes de recherches futures visant à faire avancer les connaissances dans ce domaine.



CHAPITRES COMMUNS AUX DEUX  
VOLETS  
DE LA RECHERCHE

## **2 FONDEMENTS THÉORIQUES DE L'ÉQUIVALENCE**

Dans le cadre d'une analyse approfondie de l'équivalence, il importe de bien cerner cette notion. Dans cette thèse, bien que l'objet de notre étude soit l'équivalence en terminologie, nous élargissons notre horizon en examinant l'équivalence en traductologie et en lexicologie. En effet, selon l'angle sous lequel on se place, l'équivalence est perçue sous un nouveau jour, chacun de ces points de vue apportant un éclairage complémentaire sur le sujet.

Ce chapitre est organisé en quatre sections. De la première à la troisième section, nous examinons, les unes après les autres, l'équivalence telle que perçue par les trois disciplines citées ci-dessus (2.1 à 2.3). De cet examen, nous tirons quelques conclusions sur l'équivalence et précisons où nous nous situons d'un point de vue théorique et en quoi nous espérons contribuer à une meilleure compréhension du phénomène (2.4).

### **2.1 L'équivalence en traductologie**

En traductologie, les théories sur l'équivalence ont principalement été élaborées au cours des années 1960 et 1970. Elles proviennent généralement des courants traditionnels, notamment les courants s'appuyant sur des approches linguistiques. Mais ces théories ne font pas l'unanimité, car le concept « équivalence », bien que nécessaire en tant qu'instrument de travail pour décrire les textes et leur traduction, est grandement controversé et a fait l'objet de nombreux débats, surtout à partir des années 1980. Les auteurs ne s'entendent pas sur sa définition, sur sa pertinence ou sur son applicabilité (Leonardi 2000). Par exemple, Toury (1995) estime que les traductions ne peuvent être étudiées hors de leur contexte historique et définit l'équivalence comme une notion relative :

Rather than being a single relationship, denoting a recurring type of invariant, it comes to refer to any relation which is found to have characterized translation under a specified set of circumstances. (Toury 1995 : 61)

Pour d'autres, l'équivalence est même tombée complètement en disgrâce, notamment pour Genzler (1993) et Snell-Hornby (1988), qui vont jusqu'à dire que le concept serait inutile, voire dommageable à la traductologie. En outre, de nouveaux courants théoriques apparus à la même période, en particulier le courant féministe et le courant postcolonial, remettent en question la primauté du texte source et prônent la visibilité du traducteur, ce qui a pour conséquence de rendre caduques les théories traditionnelles de l'équivalence.

Né dans les pays où les revendications des femmes prennent le devant de la scène, le courant féministe (Chamberlain 1988; Flotow 1991; Godard 1984; Lotbinière-Harwood 1991; Simon 1996; etc.) poursuit plusieurs objectifs : dénoncer le sexisme dans les langues, faire connaître les ouvrages des femmes passés sous silence, retraduire et féminiser les textes. Certaines traductrices, par un travail de réécriture, pratiquent une traduction interventionniste afin de souligner leur différence. La traduction devient alors une activité politique qui vise à rendre visible la femme et à remettre en question la fidélité à l'original.

Dans le même ordre d'idée, les auteurs du courant postcolonial (Bassnett et Trivedi 1999; Niranjana 1992; Spivak 2004; etc.), qui souhaitent libérer la traduction des valeurs imposées par les cultures dominantes, occidentales principalement, s'entendent pour dire que la traduction n'est jamais neutre, qu'elle est le reflet des tensions qui existent entre les cultures. Certains auteurs examinent en quoi la traduction a servi les intérêts des colonisateurs et s'emploient à en identifier les mécanismes. Alors que d'autres expliquent que la traduction peut être détournée au profit des colonisés. Selon les plus radicaux du mouvement, le traducteur est en droit de s'approprier ce que le colonisateur a voulu lui imposer et de le réinventer à sa convenance, par la parodie, la transcréation ou la transculturation (Vieira 1999). De ce fait, la traduction devient un site de résistance où idéologie et politique s'affrontent (Bastin *et al.* 2004). Par les prises de position du courant féministe et du courant postcolonial, on comprendra que l'équivalence n'est plus au centre de leurs préoccupations.

### **2.1.1 Typologies de l'équivalence en traductologie**

Cela dit, bien que les points de vue du courant féministe et du courant postcolonial soient légitimes et importants, il n'en demeure pas moins que pour mener à bien des études

empiriques sur l'équivalence, il est nécessaire de s'appuyer sur une théorie qui fasse consensus au sein de la majorité des chercheurs. Du nombre des auteurs qui se sont penchés sur l'équivalence, on cite volontiers Catford (1965), Jakobson (1959), Koller (1979), Nida (1964) et Vinay et Darbelnet (1958). Dans les paragraphes qui suivent, nous présentons par ordre chronologique les approches théoriques de chacun des auteurs cités.

Suivant Vinay et Darbelnet (1958), traduire c'est transférer le sens d'une langue source vers une langue cible, tant au niveau du lexique, de l'agencement (morphologie, syntaxe) ou du message (*ibid.* : 36–45). Leur méthode de traduction, basée sur la stylistique comparée de l'anglais et du français, comporte sept « procédés techniques » (*ibid.* : 46). Ces procédés se divisent en deux catégories : la traduction directe (littérale) et la traduction oblique. Les procédés directs ont recours à l'emprunt (ex. *bulldozer*), au calque (ex. *guerre froide* => *cold war*) et à la traduction littérale (ex. *le livre est sur la table* => *the book is on the table*). Selon Vinay et Darbelnet, ces trois procédés reposent sur les parallélismes (structuraux, métalinguistiques) qui existent entre deux langues. Les procédés obliques, quant à eux, font appel à « des moyens équivalents » qui devront rendre la même impression globale dans les deux messages (*ibid.* : 46). Ces procédés sont la transposition (ex. *expéditeur* => *from*), la modulation (ex. *peu profond* => *shallow*), l'équivalence (ex. *château de cartes* => *hollow triumph*) et l'adaptation (*cyclisme* => *cricket*). Selon les auteurs, l'équivalence, en tant que procédé technique, permet de « rendre compte d'une même situation en mettant en œuvre des moyens stylistiques et structuraux entièrement différents » (*ibid.* : 52). Par exemple, le proverbe anglais *like a bull in a china shop* est généralement rendu en français par *comme un chien dans un jeu de quilles*. On notera au passage que Vinay et Darbelnet entretiennent un flou terminologique par l'emploi du terme *équivalence* et l'expression *moyen équivalent*. Ces sept procédés, souvent cités et pris en exemple dans l'enseignement de la traduction, sont cependant fortement critiqués, principalement en raison de l'approche prescriptive adoptée par les auteurs.

Le linguiste, Jakobson (1959) explique dans « On linguistic aspects of translation » :

Equivalence in difference is the cardinal problem of language and the pivotal concern of linguistics. [...]. No linguistic specimen may be interpreted by the

science of language without a translation of its signs into other signs of the same system or into signs of another system. Any comparison of two languages implies an examination of their mutual translatability. (*ibid.* : 114)

D'un point de vue sémiotique, Jakobson distingue trois types de traduction : *intralinguistique*, *interlinguistique* et *intersémiotique*. L'intralinguistique concerne la reformulation de signes linguistiques par d'autres signes linguistiques dans la même langue, par exemple, par des synonymes ou des périphrases. Toutefois, l'auteur précise qu'en règle générale les synonymes ne peuvent être de parfaits équivalents. La traduction interlinguistique fait appel, le plus souvent, à la substitution de messages dans une langue par des messages dans une autre langue. En effet, dans ce type de traduction, on traduit rarement les unités linguistiques séparément. Pour ce type de traduction, comme le souligne l'auteur, l'équivalence est rarement totale. La traduction intersémiotique est l'interprétation de signes linguistiques au moyen de signes non linguistiques. Les exemples de systèmes non linguistiques donnés par Jakobson sont la musique, la danse, le cinéma et la peinture. Ici, il est intéressant de rappeler qu'en langue de spécialité, on a également recours à la traduction intersémiotique par le biais d'illustrations, de graphiques ou de formules mathématiques (Kocourek 1991 : 10–11).

S'inspirant des travaux de Chomsky (1957) et abordant l'équivalence du point de vue de la sociolinguistique, Nida (1964) déclare : « we are not content to look upon a language as some fixed corpus of sentences, but as a dynamic mechanism capable of generating an infinite series of different utterances » (*ibid.* : 9). Nida propose de distinguer l'équivalence formelle de l'équivalence dynamique. Dans le contexte de l'équivalence formelle, le traducteur s'attache à reproduire autant la forme que le message du texte source, c'est-à-dire respecter au plus près les structures lexicales et grammaticales du texte source. D'après Nida, ce type d'équivalence aboutit à des textes difficiles à lire et à comprendre. Dans le contexte de l'équivalence dynamique, le traducteur s'efforce de susciter sur le récepteur du texte cible le même effet que celui ressenti par le récepteur du texte source. En tant que linguiste et traducteur de la Bible, Nida privilégie l'équivalence dynamique, car : « [...] it is inconceivable to a Maya Indian that any place should not have vegetation unless it has been cleared for a maize-field » (Nida 1945 : 197). À titre d'exemple, Nida suggère de traduire *desert* par *endroit abandonné*. Avec ce raisonnement,

Nida veut nous faire comprendre que pour jeter des ponts entre deux réalités culturelles très différentes, le traducteur qui privilégie l'équivalence dynamique accorde la primauté à l'esprit du message du texte source au moyen de la paraphrase. De nombreux auteurs, dont Gentzler (1993), critiquent fortement l'usage de l'équivalence dynamique, ils soutiennent, entre autres, que ce type d'équivalence permet au traducteur de manipuler les textes sources, le plaçant de ce fait en position d'omnipotence.

Selon Catford (1965), l'équivalence est le problème central de la traduction et la définition de sa nature revêt une importance capitale (*ibid.* : 21). L'approche de Catford diffère radicalement de celle de Nida, en ce sens que l'auteur examine l'équivalence sous une perspective linguistique. Catford distingue deux types d'équivalence : la correspondance formelle et l'équivalence textuelle. Il y a correspondance formelle lorsqu'une partie du discours (ex. préposition) occupe la même place en langue cible que celle occupée par la même catégorie dans la langue source (*ibid.* : 27). Il y a équivalence textuelle lorsqu'un texte cible (ou une partie d'un texte cible) est considéré comme l'équivalent d'un texte source (ou une partie d'un texte source) (*ibid.* : 27). L'équivalence textuelle, quant à elle, constitue un écart (shift) par rapport à la correspondance formelle (*ibid.* : 73). Il existerait deux grands types d'écarts : *les écarts de niveau* et *les écarts de catégorie*. Les écarts de niveaux se produisent, entre autres, lorsque dans le texte cible une expression est exprimée par un élément appartenant à une forme grammaticale, alors que dans le texte source l'équivalent est représenté par une forme lexicale. Ainsi, dans l'expression *This text is intended for...* peut être rendue par *Le présent manuel s'adresse à...* L'écart de catégorie se divise en quatre : *écart de structure*, *écart de classe*, *écart d'unité* et *écart intra-systémique* (*ibid.* : 76). Les écarts de structure touchent les différences dans l'ordre des mots entre la langue source et la langue cible, par exemple, *The man is in the boat* donne en gaélique *Tha (is) an duine (the man) anns a' bhata (in the boat)*. Les écarts de classe concernent les changements de partie du discours des mots lors du passage d'une langue à l'autre, de cette manière l'adjectif dans *medical student* se traduit par un nom dans *étudiant en médecine*. Les écarts d'unité concernent les changements de rang, par exemple, dans *white house* l'adjectif précède le nom et dans *maison blanche* l'adjectif suit le nom. Dans les cas d'écarts intra-systémiques, les systèmes des deux langues restent les mêmes, mais présentent des différences, par exemple, des mots en langue source peuvent

prendre le pluriel dans la langue cible et vice-versa, ainsi *the dishes* se rend par *la vaisselle*. Les travaux de Catford sur l'équivalence sont ceux qui ont été le moins bien reçus par les traductologues qui lui reprochent une conception de l'équivalence ne correspondant pas à la pratique de la traduction (Yoda 2005 : 1009). Enfin, dernier point non négligeable à souligner, Anderman (2007) fait remarquer que les shifts de Catford s'apparentent aux recherches sur les correspondances formelles dans le domaine de la traduction assistée par ordinateur.

De son côté, Koller (1979) dresse une typologie de l'équivalence en cinq catégories : *dénotative*, *connotative*, *stylistique* (normative), *pragmatique* et *esthétique* (formelle). Il y a équivalence dénotative lorsque les mots de la langue source et les mots de la langue cible font référence au même objet (concret ou abstrait), par exemple, *apple* et *pomme* (fruit du pommier). L'équivalence connotative se réalise lorsque les mots de la langue source et les mots de la langue cible produisent la même valeur communicative sur les récepteurs, par exemple, *have kittens* et *piquer une crise*. L'équivalence stylistique (normative) consiste à respecter le genre textuel, par exemple, traduire une recette de cuisine de la langue source comme on la formulerait dans la langue cible. L'équivalence pragmatique (communicative) se produit lorsque les mots de la langue source et les mots de langue cible provoquent le même effet sur les récepteurs, par exemple, *no parking* et *défense de stationner*. Enfin, l'équivalence esthétique (formelle) produit sur les récepteurs des deux langues le même effet esthétique, ce type d'équivalence est employé pour traduire des poèmes, des chansons, des bandes dessinées, etc., par exemple :

**Épidemais** : Oh non, il l'a vendu. Il était meilleur marchand que marin.

**Ekonomikrisis** : No, he sold it. He was a better salesman than sailsman.

(Gosciny et Uderzo : *Astérix gladiateur*)

La typologie de Koller est toutefois mal reçue par certains : « Koller's five equivalence types, for example (*denotativ, konnotativ, textnormativ, pragmatisch, formal*), [...], represents little more than a reshuffling of other equivalence types... » (Snell-Hornby 1988 : 20–21).

### 2.1.2 Récapitulation

Pour résumer, la diversité des typologies de l'équivalence montre les différents aspects sous lesquels le concept peut être envisagé en traductologie. Si certains voient la nécessité de le définir, en particulier les théoriciens se fondant sur des approches linguistiques, d'autres s'y refusent pour des raisons idéologiques. Dans tous les cas, il est aujourd'hui admis que la traduction ne peut être définie uniquement en terme d'équivalence (Baker 2004 : 63), ne serait-ce que pour éviter une définition circulaire, comme ce fut le cas par le passé :

Translation may be defined as follows:

The replacement of textual material in one language (SL) by equivalent textual material in another language (TL). (Catford 1965 : 20)

Les définitions de l'équivalence reproduites ci-après évitent ce problème et parlent d'une *identité de sens* ou de *discours* entre deux langues :

L'équivalence désigne une identité de sens des segments de textes ou des textes entiers, quelles que soient les différences de choix aux plans grammatical ou lexical. (Truffaut 2004, vol 2 : 35)

Relation d'identité établie dans le discours entre deux unités de traduction de langues différentes, dont la fonction discursive est identique ou presque identique. (Delisle 2003 : 39)

Même si ces définitions ne font pas l'unanimité au sein des traductologues, il est particulièrement intéressant de remarquer que l'équivalence en traductologie touche des *unités lexicales*, des *expressions*, des *segments de texte* et même des *textes entiers*. Delisle (2003) précise que les unités de traductions « peuvent être de nature phonétique (ex. allitération), lexicale, stylistique (ex. registre), narrative (ex. mise en relief, mode ou temps des verbes), rhétorique (ex. anaphore, enchaînement argumentatif), intertextuelle (ex. allusion), etc. » (*Ibid.* : 67–68).

## 2.2 L'équivalence en lexicographie

Les dictionnaires bilingues de langue générale font partie des ouvrages élaborés par les lexicographes. La nomenclature de ces dictionnaires comprend, d'une part, toutes les parties du discours d'une langue – les mots lexicaux comme les mots grammaticaux.



D'autre part, toutes les significations d'une même forme graphique (vocable) y sont, en principe, décrites afin que les utilisateurs comprennent, dans des proportions inégales, une culture qui diffère de la leur (Al-Kasimi 1977 : 59). La mise en équivalence des unités lexicales constitue une étape cruciale et délicate :

Il est rare, même lorsque deux langues sont proches (ex. le français et l'anglais), que la traduction terme à terme de l'une à l'autre soit possible, autrement dit qu'il existe des équivalences parfaites pour la majeure partie de leur lexique. Ce qui est réalisable s'agissant de lexiques spécialisés, par exemple dans un vocabulaire géographique ou médical, ne l'est pas pour le vocabulaire courant; ce sont les mots les plus simples et les plus souvent employés qui, paradoxalement, ne trouvent pas de traduction exacte, si proches soient les deux langues [...]. (Bernot 2000 : 53)

La difficulté fondamentale de traduire une unité lexicale provient du non-isomorphisme des systèmes linguistiques (Zgusta 1971 : 294). C'est-à-dire que le découpage des réalités n'est pas le même d'une langue à l'autre pour diverses raisons : historiques, géographiques, sociales, culturelles ou économiques (Svensén 1993 : 140). Lyons (1970 : 46–47) illustre le non-isomorphisme dans son tableau du spectre des couleurs (Figure 2.1). Par exemple, si nous examinons l'unité lexicale *jaune* du français par rapport à des langues hypothétiques A, B et C, nous voyons que l'unité lexicale *c* de la langue A recouvre exactement le sens de *jaune*, alors que l'équivalent *h* de la langue B ne recouvre pas tout à fait ce sens.

<b>Français</b>	<i>rouge</i>	<i>orange</i>	<i>jaune</i>	<i>vert</i>	<i>bleu</i>
<b>A</b>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
<b>B</b>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>
<b>C</b>	<i>p</i>	<i>q</i>	<i>r</i>	<i>s</i>	

Figure 2.1. Spectre des couleurs illustrant le non-isomorphisme (d'après Lyons 1970 : 46)

D'autre part, le nombre de concepts à l'intérieur d'un champ conceptuel peut différer d'une langue à l'autre<sup>7</sup> (Svensén 2009 : 254). Ainsi, Liberman (2004) répertorie dans la langue somalie pas moins de 46 unités lexicales à l'intérieur du champ conceptuel ('grand mammifère ongulé (camélidés) à une ou deux bosses dorsales, à pelage laineux'), alors que dans d'autres langues on n'en compte que quelques-unes (Tableau 2.1). Selon l'auteur, ce phénomène serait attribuable au fait que les camélidés constituent une part importante de la culture somalie.

Tableau 2.1. Liste partielle des unités lexicales proposées par Liberman (2004).

Langue somalie	Champ conceptuel	Langue française
<i>baarqab</i> 'stud camel' <i>geel</i> 'camel' <i>gool</i> 'fat camel' <i>hal</i> 'female camel' <i>irmaan</i> 'dairy camel' <i>labakursle</i> 'two humped camel' <i>nirig</i> 'camel foal' <i>rati</i> 'male camel' <i>qaalin</i> 'young camel' <i>qurbac</i> 'young male camel' <i>rakuub</i> 'riding camel' etc.	('grand mammifère ongulé (camélidés) à une ou deux bosses dorsales, à pelage laineux')	<i>dromadaire</i> <i>méhari</i> <i>chameau</i> <i>chamelle</i> <i>chamelon</i> <i>chameau de selle</i>

Il arrive aussi que le système conceptuel d'une langue présente une différence de structure par rapport à celui d'une autre langue (Svensén 2009 : 254). À la Figure 2.2, nous voyons que le système conceptuel ('courant d'eau dans un chenal naturel') est désigné par une seule unité lexicale en anglais (*river*), alors qu'en français il est représenté par deux unités lexicales (*rivière*, *fleuve*). Le concept ('courant d'eau dans un chenal naturel'), qu'il se jette dans un autre cours d'eau ou qu'il aboutisse à la mer se nomme toujours *river* en anglais. Par contre, en français, la même notion est désignée par des unités lexicales différentes selon l'endroit où il se jette (Figure 2.2).

<sup>7</sup> Parmi les deux exemples présentés par Svensén (2009), nous évitons de reprendre l'exemple du champ conceptuel *neige* des Inuit, car il est fort critiqué (Pullum 1991). Par contre, nous utilisons celui du champ conceptuel *chameau*, confirmé par Liberman (2004).

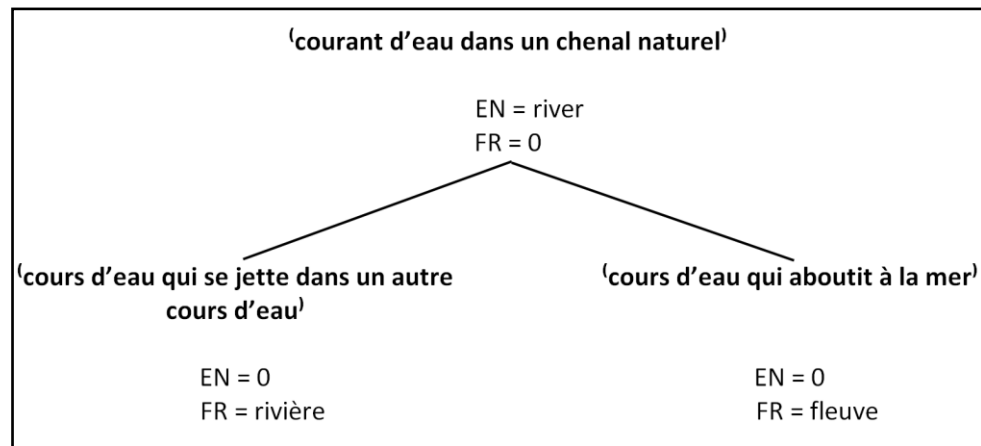


Figure 2.2. Système conceptuel dont la structure française diffère de celle de l'anglais

Même lorsque deux unités lexicales appartenant à des langues différentes sont parfaitement équivalentes du point de vue sémantique, elles présentent généralement des écarts sur les plans culturel, historique ou social. Par exemple, les couleurs sont porteuses d'une symbolique différente selon les cultures ou les époques. Le jaune est la couleur de l'empereur en Chine, alors que pour l'Islam il représente la trahison et la déception. En Occident, le blanc est lié à la pureté, à la virginité. Cependant, il est couleur de deuil pour certaines cultures en Asie. Les noms de couleurs sont loin d'être les seules unités lexicales à être marqués culturellement ou chargés d'histoire, en fait pratiquement l'ensemble du lexique est touché par ce phénomène.

Quelquefois, les concepts n'évoluent pas au même rythme d'un pays à l'autre, on songera, entre autres, au vocabulaire touchant les domaines techniques. Les changements culturels sont aussi susceptibles de progresser plus ou moins rapidement d'une culture à l'autre (Zgusta 1971 : 295). Par conséquent, les unités lexicales de deux langues désignant ces réalités peuvent être relativement équivalentes à une époque et ne plus l'être à une autre ou vice-versa. Par exemple, en 1971, la notion américaine de *drugstore* n'avait pas d'équivalent en Europe (Zgusta 1971 : 294). Toutefois, depuis ce temps-là sont apparus en France des établissements qui pratiquent approximativement ce même type de commerce sous le nom de *drugstore*.

La différence de structure entre deux systèmes linguistiques est un facteur qui influence également sur la mise en relation des équivalents. Il arrive que le nombre de parties du

discours varie d'une langue à l'autre; ou encore, que les catégories grammaticales et le genre grammatical ne se correspondent pas, par exemple, en arabe *lune* est masculin et *soleil* est féminin, alors qu'en français c'est l'inverse (Al-Kasimi 1977 : 63–64).

Enfin, un concept présent dans une culture peut être complètement absent dans une autre. D'ailleurs, selon Zgusta (1971 : 294), il n'est pas nécessaire que ces cultures soient très différentes l'une de l'autre pour que cela se produise. De nombreuses dénominations appartenant à la gastronomie en sont de savoureux exemples. Par exemple, le mot *poutine*, nom d'un populaire mets québécois, n'a pas d'équivalent en anglais.

Pour décrire les différents phénomènes que nous venons d'évoquer, les lexicographes ont établi diverses typologies de l'équivalence. La plus commune mesure le degré d'isomorphisme entre deux unités lexicales de langues différentes (Cruse 1986 : 87; Svensén 2009 : 257). Dans ce genre de typologie, on distingue généralement trois degrés d'équivalence : *équivalence exacte*, *équivalence partielle* et *non-équivalence* (Figure 2.3). Il est à souligner cependant que les frontières entre ces trois catégories ne sont pas figées (Svensén 2009 : 257) et que selon les auteurs l'équivalence partielle fait l'objet de différentes sous-divisions.

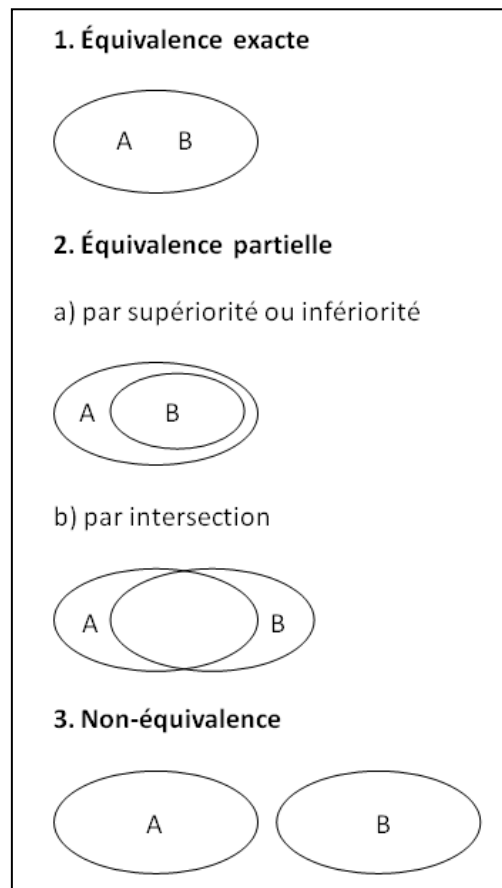


Figure 2.3. Équivalence exacte, équivalence partielle et non-équivalence

### 2.2.1 Équivalence exacte

L'équivalence est dite « exacte » entre deux unités lexicales appartenant à deux codes linguistiques différents lorsque le contenu et le registre affichent une correspondance parfaite (n° 1 de la Figure 2.3). Svensén (1993 : 150) donne pour exemple le nom français *Norvège* et le nom anglais *Norway*. Zgusta (1971 : 312) précise que les équivalents appartenant à cette catégorie doivent être identiques à tous les points de vue : même connotation, même valeur stylistique, etc. Ce type d'équivalence, selon de nombreux auteurs, est peu courant, voire rarissime (Al-Kasimi 1977; Bernot 2000; Svensén 1993; Zgusta 1971).

## 2.2.2 Équivalence partielle

L'équivalence partielle, de loin la plus fréquente et la plus complexe à décrire, se manifeste lorsqu'une unité lexicale d'une langue ne couvre pas tout à fait le même le même sens que celui d'une unité lexicale d'une autre langue. Dans ces conditions, plusieurs cas de figure sont possibles. Nous les avons divisés en trois grands groupes : *équivalence par supériorité ou infériorité*, *équivalence par intersection*, *équivalence et disparité d'usage*.

### 2.2.2.1 Équivalence par supériorité ou par infériorité

Une unité lexicale dans une langue couvre un sens plus étendu ou moins étendu que celui d'une unité lexicale dans une autre langue (Figure 2.4).

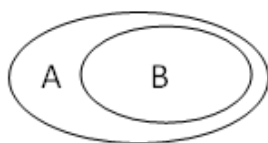


Figure 2.4. Équivalence par supériorité ou par infériorité

Dans l'exemple du Tableau 2.2, l'unité lexicale *singe* se rend soit par *ape*, soit par *monkey* en anglais selon le type de queue, alors qu'en français on ne fait pas cette distinction (Svensén 1993 : 145).

Tableau 2.2. Exemple d'équivalence partielle (*singe*)

Sens	Anglais	Français
{mammifère primate [...] à queue courte ou sans queue [...]}	<i>ape</i>	<i>singe</i>
{mammifère primate [...] à longue queue [...]}	<i>monkey</i>	

Dans le deuxième exemple (Tableau 2.3), l'italien distingue deux phases en amour (*innamoramento* et *amore*), ce que le français ne fait pas (Celotti 2002 : 457).

Tableau 2.3. Exemple d'équivalence partielle (*amour*)

Sens	Français	Italien
'le moment, la période pendant laquelle on tombe amoureux'	<i>amour</i>	<i>innamoramento</i>
'inclination envers une personne, souvent à caractère passionnel'		<i>amore</i>

Certains auteurs sentent la nécessité d'effectuer un découpage plus fin à l'intérieur de ce groupe. Par exemple, Svensén (2009 : 259) distingue la convergence de la divergence. La convergence se produit lorsqu'une unité lexicale de la langue A contient un plus grand nombre de traits sémantiques qu'une unité lexicale de la langue B. Par exemple, en anglais (A), le sens de *boyhood* est caractérisé par un plus grand nombre de traits sémantiques que celui d'*enfance*, en français (B) (Figure 2.5). La divergence survient lorsqu'une unité lexicale dans la langue A contient un moins grand nombre de traits sémantiques qu'une unité lexicale de la langue B. Par exemple, en anglais (A), *river* ne fait pas de distinction entre les sens des unités lexicales *rivière* et *fleuve* du français (B) (Figure 2.6).

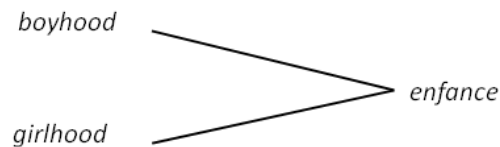


Figure 2.5. Convergence (de la langue A vers la langue B) (d'après Svensén 2009 : 259)

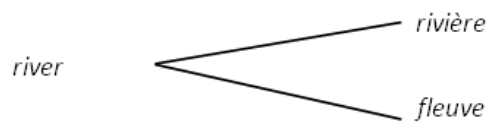


Figure 2.6. Divergence (de la langue A vers la langue B) (d'après Svensén 2009 : 259)

Les notions de convergence et de divergence sont utiles si l'on considère l'équivalence à partir d'une même langue, par exemple, d'une langue A vers une langue B

– lors de la construction d'un dictionnaire. Cependant, comme elles sont l'image miroir l'une de l'autre, nous croyons qu'il n'est pas toujours nécessaire de les distinguer et qu'elles peuvent être regroupées, soit la définition donnée au début de cette section (2.2.2.1) : une unité lexicale dans une langue couvre un sens plus étendu ou moins étendu que celui d'une unité lexicale dans une autre langue.

### 2.2.2.2 Équivalence par intersection

Deux unités lexicales appartenant à des langues différentes partagent de nombreux traits sémantiques. Par contre, chacune de ces unités lexicales possède des traits sémantiques qui lui sont propres (Figure 2.7). Par exemple, dans le domaine de l'architecture paysagère, le terme anglais *gazebo* et le terme français *belvédère* dénotent la même idée générale, mais leurs sens ne se recouvrent pas tout à fait, en effet ('summerhouse from which one can gaze at the surrounding scenery') et plus ou moins égal à ('kiosque isolé placé sur une éminence') (ITWG 1999).

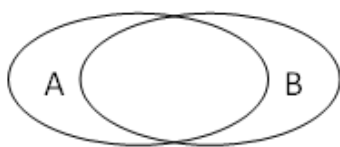


Figure 2.7. Équivalence par intersection

### 2.2.2.3 Équivalence et disparité d'usage

Enfin, l'équivalence partielle provient parfois du fait qu'une unité lexicale d'une langue soit d'un registre différent de celui d'une unité lexicale d'une autre langue. Par exemple, selon le Robert & Collins (2003), le mot français qui se rapproche le plus de l'expression britannique *goggle-box* est *télé*. Cependant, étant donné que *goggle-box* est plus familier que *télé*, il faut spécifier la marque d'usage des deux entrées.

## 2.2.3 Non-équivalence

On parle de non-équivalence lorsque pour une unité lexicale dans une langue donnée, on ne trouve pas d'équivalent dans une autre langue. Il y a non-équivalence lorsque : a) un



concept existe dans une culture cible, mais n'a pas encore été nommé; b) un concept est absent dans une culture cible. Selon Al-Kasimi (1977 : 61), deux types d'unités lexicales sont plus susceptibles d'être touchées par la non-équivalence : les unités liées à la culture de la langue source (ex. onomatopées, interjections, mots grammaticaux invariables) et les notions culturelles, scientifiques ou technologiques. Nous pensons à des mots tels que *saudade* en portugais. Cette expression qui exprime une tristesse empreinte de nostalgie, quand une personne se sent dépossédée de son passé, ne trouve pas d'équivalent dans les autres cultures.

## 2.2.4 Autres typologies de l'équivalence en lexicographie

Outre la typologie de l'équivalence exprimant le degré d'isomorphisme entre deux unités lexicales de langues différentes, il existe des typologies qui envisagent l'équivalence sous d'autres angles. Plusieurs sont décrites dans la littérature, mais nous n'en présentons que deux dans cette section. La première nous vient de Svensén (2009 : 256), *l'équivalence sémantique et l'équivalence pragmatique*; la deuxième de Zgusta (1987 : 147–154), *l'équivalence traductionnelle* (translational) et *l'équivalence explicative* (explanatory).

### 2.2.4.1 Équivalence sémantique et équivalence pragmatique

L'*équivalence sémantique* se dit d'un équivalent dont le sens est identique à celui de l'unité lexicale source, par exemple, l'unité lexicale *citrouille* est l'équivalent de l'unité lexicale anglaise *pumpkin* lorsque ces dernières désignent le même concept ('courge arrondie et volumineuse de couleur jaune orangé'). L'équivalence sémantique est consignée dans les dictionnaires bilingues. Toutefois, dans certains contextes, *citrouille* ne sera plus l'équivalent adéquat de *pumpkin*. Par exemple, aux États-Unis, pour signifier son affection à un enfant, il est fréquent d'employer le mot *pumpkin* lorsqu'on s'adresse à lui : *You are my little pumpkin*. La phrase anglaise traduite mot à mot en français donnerait *Tu es ma petite citrouille*. En français, cette traduction ne produirait évidemment pas le même effet qu'en anglais, surtout du point de vue esthétique. Par conséquent, il est nécessaire de recourir à l'*équivalence pragmatique* afin de trouver un équivalent convenant au contexte. Dans le cas présent, cela pourrait être *chou*, ce qui donnerait *Tu es mon petit chou*.

Dans un dictionnaire bilingue, les équivalents sémantiques d'une unité lexicale sont idéalement tous répertoriés, alors que les équivalents pragmatiques ne le sont pas tous. Par exemple, à l'entrée *pumpkin*, la version électronique du Robert & Collins (2003) n'offre pas l'équivalent *chou* quand le terme anglais est utilisé dans un contexte tel que celui décrit ci-dessus. Les équivalents pragmatiques étant très nombreux, les dictionnaires bilingues doivent faire un choix et ne présenter que ceux qu'ils jugent utiles. Ainsi, à l'entrée *sugar* (sucre), le Robert & Collins (2003) offre un équivalent pragmatique *chérie* : « come here *sugar* » => « Viens ici *chérie* ».

#### 2.2.4.2 Équivalence traductionnelle et équivalence explicative

Dans les dictionnaires bilingues, les équivalents sont généralement de deux ordres : *traductionnels* ou *explicatifs*. L'*équivalent traductionnel* est une unité lexicale de la langue cible, qui, une fois insérée dans le texte cible, produit une traduction conforme aux usages établis dans la langue cible (Tableau 2.4). Idéalement, les dictionnaires bilingues devraient toujours proposer des équivalents traductionnels (Zgusta 1987 : 147). Or, comme expliqué par l'auteur et comme nous avons pu le constater au début de cette section, il est parfois impossible d'offrir un équivalent traductionnel. Dans ces conditions, le lexicographe doit recourir à l'équivalence explicative. L'*équivalent explicatif* est un syntagme non figé de la langue cible qui permet à l'utilisateur du dictionnaire de comprendre la notion exprimée par l'unité lexicale source (Tableau 2.4). Ces syntagmes ne sont pas des définitions et ils peuvent selon le contexte être insérés dans le texte cible. Toutefois, ils ne possèdent pas les mêmes qualités que les équivalents traductionnels.

Tableau 2.4. Exemple d'équivalent traductionnel et d'équivalent explicatif

	Anglais	Français
Équivalent traductionnel	<i>tundra</i>	<i>toundra</i>
Équivalent explicatif	<i>tundra</i>	<i>steppe arctique</i>

#### 2.2.5 Récapitulation

Pour récapituler, l'équivalence en lexicographie ne se présente pas toujours comme une relation symétrique, c'est-à-dire : A = B et B = A. Bien au contraire, les relations d'équivalence sont très souvent caractérisées par des écarts. Afin de décrire avec plus de

précision les équivalents, les dictionnaires bilingues en lexicographie comportent généralement deux parties, celle qui décrit les unités lexicales de la langue A vers la langue B et celle qui les décrit de la langue B vers la langue A (Van Campenhoudt 2003 : 84).

Pourtant, bien que l'on connaisse depuis toujours le non-isomorphisme des systèmes linguistiques et qu'il soit bien décrit dans la littérature, les dictionnaires bilingues continuent de présenter presque exclusivement les équivalences de même catégorie grammaticale, à moins qu'ils n'y soient absolument forcés (Zgusta 1971 : 313), ce qui pose bien des problèmes à l'utilisateur non averti. Ainsi, pour traduire *changement climatique* en anglais, on trouvera respectivement dans le Robert & Collins (2003) comme équivalence *change* et *climatic*, ce qui donnerait *climatic change* (ADJ + NOM) – terme, qui même s'il n'est pas totalement fautif, est beaucoup moins idiomatique que *climate change* (NOM + NOM). Ici, le dictionnaire bilingue n'a pas informé l'utilisateur qu'en fonction de la base de la collocation les adjectifs français sont souvent rendus par des noms en anglais.

Ensuite, d'après les deux définitions suivantes, nous constatons qu'en lexicographie l'équivalence touche principalement les unités lexicales ou les expressions – mais pas les phrases, comme c'était le cas en traductologie.

Equivalence can be defined as the relationship between a source-language expression and a target-language expression with regard to meaning (**SEMANTIC EQUIVALENCE**) and usage (**PRAGMATIC EQUIVALENCE**). (Svensén 2009 : 253)

We call equivalent such a lexical unit of the target language which has the same lexical meaning as the respective lexical unit of the source language. (Zgusta 1971 : 312)

Pour terminer, comme le souligne Adamska-Salaciak (2010 : 388–389), l'utilisation de plus en plus fréquente des corpus pour étudier l'équivalence invite le lexicographe à considérer la langue non seulement en tant que système, mais aussi en tant que texte. Ainsi, les dictionnaires introduisent de plus en plus des équivalents intertextuels (ex. l'équivalence pragmatique de Svensén (2009 : 255)). Par ailleurs, dans ce même article, Adamska-Salaciak aborde une question importante « Is equivalence discovered or created? ». Par cette question, l'auteure distingue les équivalents qui préexistent (en principe répertoriés dans les dictionnaires) de ceux qui sont créés au moment du processus de traduction.

## 2.3 L'équivalence en terminologie

Comme décrite dans la section intitulée Cadres théoriques (1.3), l'approche conceptuelle, qui est principalement onomasiologique, envisage le terme comme une étiquette ayant une fonction essentiellement dénominative qui sert à désigner la représentation mentale générale et abstraite d'un objet, le concept. L'approche lexico-sémantique, quant à elle, qui part nécessairement des textes – donc sémasiologique, considère les termes comme des unités lexicales et les décrits dans leur fonctionnement linguistique.

Ces différents points de vue influent sur la manière de définir l'équivalence. Sous l'angle conceptuel, l'équivalence est envisagée comme une « relation établie entre deux termes de langues différentes qui désignent des notions correspondantes » (Boutin *et al.* 1985 : 20). Alors que du point de vue de la lexico-sémantique, on dira que « Des termes sont équivalents lorsqu'ils ont les mêmes composantes sémantiques » (L'Homme 2004 : 115). Ces deux définitions présentent l'équivalence terminologique sous sa forme idéale. Toutefois, comme le soulignent de nombreux auteurs (Dubuc 2002 : 73–74; L'Homme 2004 : 115; Rondeau 1981 : 33; Van Campenhoudt 2001), il n'est pas toujours possible d'établir une équivalence parfaite entre deux termes parce que le découpage de la réalité diffère d'une langue à l'autre, ou, autrement dit, parce que les champs de signification ne se superposent pas exactement (Vandaele et Raffo 2008).

Dans les sections 2.3.1 à 2.3.3, nous présentons *l'équivalence d'un point de vue conceptuel, telle qu'observée dans les corpus* et *l'équivalence d'un point de vue structurel* (Le Serrec *et al.* 2010). Au point 2.3.4, nous abordons le traitement des termes polysémiques et des homonymes. Enfin, au point 2.3.5, nous procédons à une récapitulation de l'équivalence en terminologie.

Mais avant de poursuivre sur le sujet de l'équivalence, nous jugeons à propos de clarifier maintenant ce que nous entendons par *terme simple* et *terme complexe*, puisque ce sont des notions centrales à cette étude, notamment dans cette section. Pour définir ces termes, deux approches peuvent être envisagées : celle qui se base sur des critères linguistiques et celle qui s'appuie sur des indices graphiques. Dans cette thèse, nous avons adopté cette dernière approche. Autrement dit, un *terme simple* est composé d'une chaîne

de caractères délimitée par des espaces (ex. *adaptivity, cloud, coastline*) et un *terme complexe* est constitué de plusieurs chaînes de caractères séparés par des espaces ou des traits d'union (ex. *greenhouse effect, palaeo-lake*). Parmi les exemples du terme simple, on trouve un terme formé d'une seule base (*cloud*), un dérivé (*adaptivity*) et un composé<sup>8</sup> (*coastline*). Les *dérivés* résultent de « la combinaison d'un radical et d'un affixe » (Polguère 2003 : 62); alors que les *composés* « sont issus de l'association de deux radicaux autonomes ou non » (Jousse 2002 : 64). Pour écrire un composé, trois graphies sont possibles : les éléments sont accolés (*groundwater*), reliés par un trait d'union (*human-induced*) ou séparés par un blanc (*sea rise*) (Ahronian 2005 : 14). Parfois, pour une même dénomination, la graphie fluctue (*seawater, sea-water* ou *sea water*).

Dans la littérature, les auteurs qui se fondent sur des critères linguistiques définissent le *terme simple* et le *terme complexe* en fonction des mécanismes linguistiques qui président à leur formation. Ci-après, nous présentons des définitions de trois auteurs :

**simple term**

**term** (3.4.3) containing only one root

NOTE Examples of simple terms are: *sound, light*.

**complex term**

**term** (3.4.3) containing two or more roots

NOTE Examples of complex terms are: *bookmaker, know-how, fault recognition circuit*. (ISO 1087-1 2000 : 7)

Selon les types de morphèmes qui interviennent dans la formation d'un terme complexe, les termes peuvent être dérivés (formés en ajoutant des affixes à un radical) ou composés (formés en combinant des bases lexicales, actuelles ou historiques, avec la possibilité d'y adjoindre des affixes). Exemples : *dirigeable, fruitier, téléphonique, terminologie, ulcéreux, tourne-disque*. (Cabré 1998 : 155)

Dorénavant, *terme simple* servira à désigner les unités lexicales composées d'une seule entité graphique. Ce premier groupe comprend les termes formés d'une seule

---

<sup>8</sup> La composition est plus fréquente en anglais qu'en français. Cependant, elle existe aussi chez cette dernière (ex. *abribus, autoroute, betterave, chèvrefeuille, chiendent, Kilowatt, portefeuille, portemanteau, tournevis, vélomoteur, vinaigre*).

base (ex. ROBINET, CLÉ, MARCHÉ) ainsi que les *dérivés*, c'est-à-dire les termes comprenant un radical et un ou plusieurs *morphèmes dérivationnels* (ex. MARCHANDIS-AGE, ANTI-CHAR, NAVIG-ATION, MICRO-ORDINATEUR). *Terme complexe* sera utilisé pour désigner les termes constitués de plusieurs entités graphiques séparées par des blancs ou par des diacritiques comme le trait d'union ou l'apostrophe (ex. SYSTÈME-EXPERT, POISSON-CLOWN, MÉMOIRE NON RÉMANENTE, SERRE-JOINT). (L'Homme 2004 : 59)

Selon l'ISO 1087-1 (2000 : 7) et Cabré (1998 : 155), les composés, même écrits en une seule chaîne de caractères, figurent parmi les termes complexes. De son côté, L'Homme (2004 : 59) ne les mentionne pas dans ses définitions. En outre, Cabré place les dérivés parmi les termes complexes alors que l'ISO et L'Homme ne le font pas. À la lumière de ces trois exemples, il apparaît que les définitions qui s'appuient sur des critères linguistiques ne font pas consensus. Par ailleurs, comme l'observent Ahronian (2005 : 13) et Jousse (2002 : 64–66), les auteurs ne classifient pas toujours de la même manière certaines unités lexicales de constructions similaires. C'est le cas notamment pour les unités formées à partir de bases grecques ou latines. En effet, pour Kocourek (1991 : 109) la formation à l'aide de ces bases s'appelle la confixation (ex. *poly* et *mère* qui donne *polymère*). Pour cet auteur, la confixation est apparentée à la dérivation, alors que chez Tournier (1985 : 107), la formation savante des termes avec des confixes – qu'il appelle quasi-morphèmes – est associée à la composition.

Afin d'éviter les différents problèmes de classification observés dans les définitions linguistiques, nous avons choisi de suivre les définitions qui se basent plus spécifiquement sur la forme graphique des termes pour départager les termes simples des termes complexes :

Les *termes simples* ont la structure d'un seul mot<sup>9</sup>, formé par composition ou par dérivation, qui est délimité des autres mots dans la phrase par deux espaces blancs. Les *termes complexes* comprennent deux ou plusieurs mots séparés par des espaces blancs ou liés par des traits d'union, qui forment une expression à sens unique, aussi appelée « syntagme terminologique » (Le Bureau de la traduction : 2007)

---

<sup>9</sup> Bien que le terme *mot* soit ambigu, dans cette définition, nous comprenons qu'un mot est une chaîne de caractères délimitée par des espaces.

À ces dernières définitions, les auteurs ajoutent que les mots-valises (ex. *logithèque*), les acronymes (ex. *ovni*) et les sigles (ex. *PDF*) sont traités comme des termes simples.

D'autre part, les définitions du Bureau de la traduction s'harmonisent avec celles que l'on retrouve généralement dans le domaine du TAL (Drouin 2002 : 210–216; Zhang et Fang 2010). Effectivement, en informatique un terme simple (en anglais : « single word term ») est généralement constitué d'une seule unité graphique délimitée par des espaces, alors que les termes complexes (en anglais « multi word term ») sont composés de plusieurs unités graphiques séparées par des espaces ou par des traits d'union<sup>10</sup>. Ces distinctions résolument pragmatiques s'imposent dans ce domaine, car distinguer les unités lexicales dont les éléments sont soudés des unités de base poserait de grandes difficultés. Enfin, étant donné que la terminographie accorde de plus en plus de place aux outils de traitement de textes, il nous semble justifié d'uniformiser ces définitions avec celles du domaine du TAL.

### 2.3.1 L'équivalence d'un point de vue conceptuel

En terminologie, comme en lexicographie, on distingue différents types d'équivalence dont le spectre s'étend de l'*équivalence exacte* à la *non-équivalence* en passant par divers degrés d'*équivalence partielle*. Dans le but de répertorier les différents cas de figure de l'équivalence, plusieurs typologies ont été avancées (Dubuc 2002; Felber 1987; Rondeau 1981; etc.). Ci-après, nous proposons une synthèse de ces typologies.

#### 2.3.1.1 Équivalence exacte

Selon Rondeau (1981 : 33), la relation d'équivalence est dite exacte quand deux termes,  $T_1$  et  $T_2$ , de langues différentes,  $L_1$  et  $L_2$ , affichent un rapport identique entre le concept (C) et la dénomination (D), ce qui donne :

$$T_1(L_1) = T_2(L_2) = \left( \frac{D}{C} \right)_{L_1} = \left( \frac{D}{C} \right)_{L_2}$$

(Rondeau 1981 : 33)

---

<sup>10</sup> Il arrive que certains auteurs considèrent les chaînes de caractères contenant un trait d'union comme des termes simples (ex. *re-hearing*) (Kit et Liu 2008).

Pour illustrer ce type d'équivalence, nous empruntons au domaine de l'hydrologie le terme anglais *eutrophication* et son équivalent français *eutrophisation*. Dans ce domaine<sup>11</sup>, ces termes entretiennent un rapport d'équivalence exacte, car chacun d'eux correspond en tout point au concept ('Enrichissement de l'eau, qu'elle soit douce ou saline, par des nutriments, [...]'). En outre, ils ne « présentent pas de disparité quant à leurs modalités d'utilisation (niveaux de langue, usages géographiques ou professionnels, etc.) » (Dubuc 2002 : 73).

### 2.3.1.2 Équivalence partielle

L'approche conceptuelle de la terminologie privilégie l'équivalence exacte puisque cette dernière s'accorde avec ses visées de normalisation. Toutefois, deux termes provenant de deux langues différentes peuvent ne pas recouvrir tout à fait la même réalité conceptuelle, ce qui pose des difficultés lors de leur mise en équivalence. De façon générale, il y a équivalence partielle lorsqu'un terme  $T_1$  d'une langue  $L_1$  ne partage pas la même réalité avec un terme  $T_2$  d'une langue  $L_2$ . Rondeau présente cette relation de la façon suivante :

$$T_{(L_1)} \neq T_{(L_2)}$$

(D'après Rondeau 1981 : 33)

À l'intérieur de l'équivalence partielle, plusieurs cas de figure sont généralement envisagés. Ici, comme en lexicographie, nous les avons divisés en trois groupes : *équivalence par supériorité ou par infériorité*, *équivalence par intersection* et *équivalence et disparité d'usage*.

#### 2.3.1.2.1 Équivalence par supériorité ou par infériorité

Une dénomination en langue  $L_1$  couvre une réalité plus étendue ou moins étendue qu'une dénomination de la langue  $L_2$  (Rondeau 1981 : 34) (Figure 2.8).

---

<sup>11</sup> Il est à remarquer que le terme *eutrophisation* appartient également à d'autres domaines (ex. écologie) sans pour autant changer de sens.



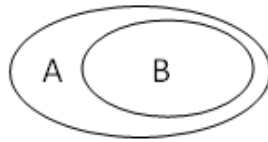


Figure 2.8. Équivalence par supériorité ou par infériorité

Dans un premier exemple (Figure 2.9), nous voyons que le terme anglais *nuts*, dans le domaine de l'alimentation, couvre un champ sémantique plus étendu que le terme français *noix*<sup>12</sup> puisqu'il inclut des fruits à écales qui ne sont pas compris dans le champ sémantique du terme français (Kraif 2001 : 117).

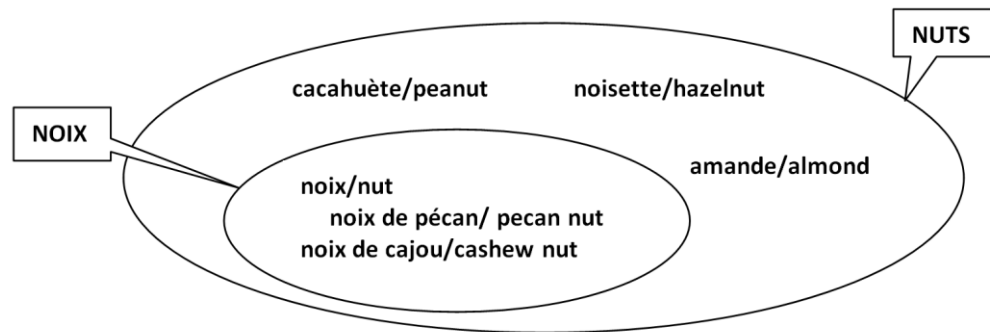


Figure 2.9. Illustration des champs sémantiques de *nuts* et de *noix*

Dans un deuxième exemple tiré du sous-domaine de la voile (Figure 2.10), nous voyons que le champ conceptuel ('petite voile carrée au-dessus du perroquet') est désigné par un seul terme en français (*cacatois de perruche*), alors qu'en anglais il est associé à deux termes (*mizen-royal, jigger-royal*) (d'après Van Campenhoudt 1996 : 289). Cette voile, qu'elle soit située sur un trois-mats carré ou un quatre-mats carré se nomme toujours

<sup>12</sup> *Noix* désigne d'abord le fruit du noyer, mais il a un second sens, plus étendu, soit « se dit d'autres fruits qui ont quelque ressemblance avec la noix; *noix de pécan, noix d'acajou, de coco, de kola, du Brésil, etc.* ». L'équivalent français *amandes diverses* nous semble donc inexact. *Amande*, fruit de l'amandier, se dit en anglais *almond*, pris dans le sens générique qu'on veut lui attribuer dans l'expression *amandes diverses*, *amande* signifie « toute graine contenu dans leur noyau » et se dit en anglais *kernel* et non pas *nut*. Les Français utilisent *fruits secs* pour désigner ces fruits. Nous n'avons pas retenu cette expression étant donné qu'il existe des fruits secs autres comme les pruneaux, les raisins, etc. En français, il n'y a pas vraiment de terme générique pour désigner l'ensemble des fruits à écale (comme les noisettes, les arachides, les amandes et les noix). L'énumération des termes exacts peut alors être nécessaire, surtout dans des mises en garde destinées à protéger les personnes souffrant d'allergies alimentaires (GDT 2009).

*cacatois de perruche* en français. Par contre, en anglais, le même concept est désigné par des unités lexicales différentes selon le type de voilier, *mizen-royal* sur un trois-mâts carré et *jigger-royal* sur un quatre-mâts carré. Dans ce dernier exemple, *cacatois de perruche* possède un sens plus large que *mizen-royal* ou que *jigger-royal* alors que ces derniers ont un sens plus spécifique.

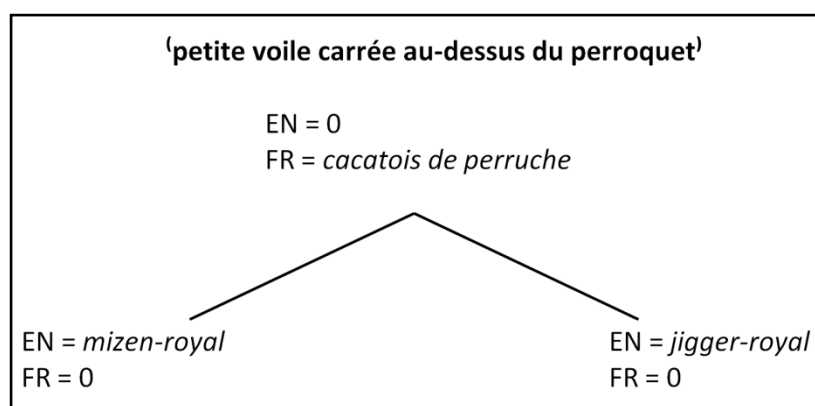


Figure 2.10. Deuxième exemple d'équivalence par supériorité ou par infériorité (d'après Van Campenhoudt 1996 : 289)

### 2.3.1.2.2 Équivalence par intersection

Une dénomination en langue  $L_1$  et une dénomination en langue  $L_2$  ne partagent que certains de leurs traits sémantiques (Felber 1987 : 130) (Figure 2.11).

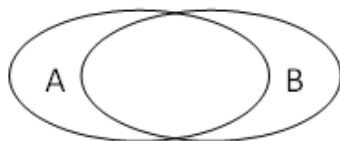


Figure 2.11. Équivalence par intersection (d'après Felber 1987 : 129)

Pour ce type d'équivalence, Felber (1987 : 130) donne pour exemple le terme anglais *cricket* et le terme allemand *Schlagball*, termes appartenant au domaine du sport. Ces deux activités utilisent les mêmes instruments, mais ne partagent pas tout à fait les mêmes règles de jeu.

### 2.3.1.2.3 *Équivalence et disparité d'usage*

La disparité d'usage n'est pas exclusive aux unités lexicales, elle peut également toucher les termes. Par *usage*, on entend tout ce qui concerne les modalités d'utilisation d'un terme : registre de langue, usage géographique ou professionnel, etc. (Dubuc 2002 : 73–76). Il arrive qu'un concept dans une langue soit désigné par plusieurs termes de registres différents, alors que dans l'autre langue il n'existe pas le même découpage. Ainsi, le terme anglais *zoom* ('type d'objectif') a pour équivalent français *objectif à focale* lorsque employé dans le registre technico-scientifique et *zoom* dans le jargon des studios (Dubuc 2002 : 74–75). De même, un terme dans une langue donnée sera utilisé de préférence dans une aire géographique alors que le terme désignant le même concept dans l'autre langue ne sera pas soumis à la même restriction. Par exemple, le terme anglais *permafrost* ('sol gelé en permanence') a pour équivalents français *pergélisol*, *permagel* ou *permafrost*. Cependant le terme français *permafrost* ne peut être utilisé partout dans la francophonie, car il est fortement déconseillé au Québec.

### 2.3.1.3 Non-équivalence

La non-équivalence se produit : a) soit parce qu'un concept dans une langue  $L_1$  est inconnu dans une langue  $L_2$ ; b) soit parce qu'un nouveau concept nommé dans une langue  $L_1$  n'a pas encore reçu de dénomination dans une langue  $L_2$  (Rondeau 1981 : 34). Certains domaines sont plus susceptibles que d'autres de présenter des cas de non-équivalence (ex. juridique, gastronomique, titre de poste, aménagement du territoire). Dans le domaine juridique, *common law* et *Equity* ne trouvent pas d'équivalents en français (Gémar 2002 : 170–171). Le titre de poste anglais *Lord Chancellor* n'existe pas en France alors que le terme français *arrondissement* ('circonscription administrative') n'a pas d'équivalent en anglais, etc.

## 2.3.2 Équivalence terminologique telle qu'observée en corpus

Comme nous l'avons vu dans la section 2.3.1, l'équivalence en terminologie est principalement abordée sous l'angle conceptuel. Les contextes dans lesquels le terme s'insère sont rarement pris en considération. Par conséquent, certains phénomènes courants

qui peuvent être observés en corpus sont passés sous silence. Pourtant, comme le souligne Bédard (1986) :

Les termes employés dans le texte de départ doivent faire l'objet d'une analyse non pas intrinsèque, mais contextuelle : ils ne doivent pas simplement être analysés pour eux-mêmes, mais dans une certaine mesure **être redécouverts en contexte**. (Bédard 1986 : 42)

En outre, le fait de considérer les termes seulement sous l'angle conceptuel implique que les études de l'équivalence en terminologie se limitent généralement aux noms ou aux syntagmes nominaux. Or, même en terminologie, les autres parties du discours jouent un rôle important.

Ci-après, en quatre points, nous présentons les principaux phénomènes rapportés par les chercheurs qui se sont penchés sur l'équivalence ou sur la variation terminologique en corpus<sup>13</sup> bilingue.

a) « Un terme dans une langue A a plusieurs équivalents dans le corpus de la langue B » (L'Homme 2004 : 208). Plusieurs causes peuvent expliquer ce cas de figure, la synonymie, la quasi-synonymie, la variation, l'omission d'un élément de sens d'un terme complexe ou encore la variation graphique dans la langue d'arrivée (Carreño 2004; Le Serrec 2008; Sager 1994). À titre d'exemple, l'omission d'un (ou plusieurs) élément de sens du terme complexe de la langue source se produit lorsqu'on ne veut pas répéter au long un terme, ainsi, *témoin lumineux de contrôle de rotation* peut dans certaines circonstances être réduit à *témoin de rotation* (Bédard 1986 : 66). Dans une étude anglais-espagnol portant sur la variation terminologique en corpus parallèle<sup>14</sup>, Carreño montre que la variation touche presque tous les termes étudiés (simples et complexes) et que cette variation se présente sous différentes formes : sémantique (ex. le terme anglais *abatement* donne en espagnol *reducción* ou *combate*), syntaxique (ex. omission du terme dans la traduction ou substitution par une anaphore, tel *éstos*) et morphosyntaxique (ex. changement de catégorie grammaticale *abatement* en anglais à *reducir* en espagnol).

<sup>13</sup> La plupart des phénomènes traités dans cette section sont énumérés dans L'Homme (2004 : 207–208).

<sup>14</sup> Les études de la variation ou de l'équivalence sur corpus bilingue étudient les mêmes phénomènes.

b) « Un syntagme nominal s'exprime dans l'autre langue par un syntagme d'une autre nature » (L'Homme 2004 : 208). De nombreux cas de ce type ont été relevés dans Carreño (2004) et Le Serrec *et al.* (2010). Dans les exemples suivants, nous voyons que les syntagmes *carbon capture*, *carbon storage* et *climate change* sont parfois traduits par *piéger le carbone*, *stocker le carbone* et *modifier le climat* :

Currently the world's ecosystems, instead of maintaining and enhancing nature's CARBON CAPTURE and STORAGE capacity, are being depleted at an alarming rate.

Actuellement, au lieu de maintenir et d'améliorer les capacités de la nature à PIÉGER et à STOCKER LE CARBONE, les écosystèmes du monde s'appauvrissent à une vitesse alarmante. (PNUE\_bioseq\_2009)

If this were to occur, it could lead to a rapid regional CLIMATE CHANGE in the North Atlantic region, ...

Une telle situation serait susceptible de MODIFIER rapidement LE CLIMAT dans l'Atlantique Nord... (IPCC\_bilan\_2001)

c) « Il arrive qu'une phrase d'un texte source comporte une mention explicite à un terme complexe, mais que la phrase du texte cible utilise plutôt une anaphore. Par exemple, *...the disk drive is identified...* peut se rendre par *...ce dispositif est identifié...* ou encore *...celui-ci est identifié...* » (L'Homme 2004 : 208). Même dans les textes techniques, les traducteurs ont souvent recours à l'anaphore afin d'éviter d'alourdir le texte.

d) Un terme simple ou complexe de la langue de départ n'est pas traduit dans la langue d'arrivée. Le terme est volontairement omis dans la traduction, généralement pour éviter une répétition dans la langue cible. Selon Carreño (2004 : 103) ce type d'omission toucherait plus souvent les termes simples que les termes complexes. L'auteure donne pour exemple le cas du terme *recycling* qui figure deux fois dans une phrase anglaise et une seule fois dans la traduction espagnole.

A CEC effort could determine whether the same concerns are applicable to RECYCLING of products containing lead or RECYCLING of lead.

La CCA podrá emprender una acción para determinar si las mismas consideraciones son aplicables al RECICLADO de productos con contenido de plomo o del propio plomo. (Carreño 2004 : 86)

### 2.3.3 Équivalence terminologique d'un point de vue structurel

Dans le but de répondre à la demande grandissante de terminologies bilingues, les informaticiens s'efforcent de trouver des moyens d'extraire automatiquement les termes et leur équivalent des corpus parallèles (Gaussier 2001). Les outils informatiques ne pouvant pas tirer parti des typologies sur l'équivalence présentées dans la section 2.3.1, des chercheurs ont étudié les différences structurelles entre les termes complexes et leur équivalent (Carl *et al.* 2004; Gaussier 2001; Tran *et al.* 2003)<sup>15</sup>. Par exemple, le terme anglais *climate change* constitué sur le patron NOM<sub>2</sub> + NOM<sub>1</sub> se rend en français selon le patron NOM + ADJ, soit *changement climatique*. Les règles de transformation des patrons touchent les éléments des termes selon divers scénarios : a) leur nature grammaticale, b) leur ordre ou c) leur nombre.

a) Un terme complexe dont l'un des éléments est un nom dans une langue peut voir l'équivalent de cet élément rendu par un adjectif, *energy balance* (NOM<sub>2</sub> + NOM<sub>1</sub>) donne *bilan énergétique* (NOM + ADJ).

b) L'ordre des éléments d'un terme complexe dans une langue n'est pas le même dans l'équivalent, *écologie profonde* (NOM + ADJ) donne *deep ecology* (ADJ + NOM).

c) Le nombre d'éléments d'un terme complexe dans une langue est différent de celui de son équivalent, *portable life support system* (ADJ + NOM + NOM + NOM) est composé d'un adjectif et de trois noms alors que son équivalent *équipement de survie* (NOM + prép + NOM) ne comporte que deux noms et une préposition (L'Homme 2004 : 208). Un terme complexe peut aussi se rendre par un terme simple, *palm tree* (NOM + NOM) donne *palmier* (NOM).

Toutefois, ainsi que le souligne Gaussier (2001 : 174), on ne peut se fier uniquement à ces patrons pour extraire des paires d'équivalents, car, pour un même patron dans une

<sup>15</sup> Les auteurs cités se sont penchés sur le couple de langue français-anglais.

langue donnée, il peut exister plusieurs règles de transformation vers l'autre langue (Tableau 2.5).

Tableau 2.5. Exemples de transformations du patron français NOM + prép/dét + NOM vers l'anglais

Terme français ayant un patron NOM + prép/dét + NOM	Patron anglais	Terme anglais
<i>effet de serre</i> (changement climatique)	NOM <sub>2</sub> + NOM <sub>1</sub>	<i>greenhouse effect</i>
<i>roche de profondeur</i> (pétrographie)	ADJ + NOM	<i>abyssal rock</i>
<i>désistement d'action</i> (juridique)	NOM <sub>1</sub> + prép + NOM <sub>2</sub>	<i>abandonment of action</i>
<i>mise en garde</i> (escrime)	VER + prép + NOM	<i>putting on guard</i>
<i>ancre à jas</i> (navigation)	NOM + poss + NOM	<i>fisherman's anchor</i>
<i>durée de vie</i> (changement climatique)	NOM	<i>lifetime</i>

À quelques exceptions près et pour les raisons évidentes que nous venons d'évoquer, les recherches sur l'équivalence d'un point de vue structurel se sont surtout concentrées sur les termes complexes. Parmi les rares exceptions, on peut compter l'étude de Tran *et al.* (2003). Dans leur article, les auteurs présentent deux règles de transformations structurelles touchant les termes simples : l'*isomorphisme* et le *calque morphologique*. Ces règles de transformation font référence à différents degrés de cognation<sup>16</sup>, phénomène largement utilisé en TAL. L'*isomorphisme* se rapporte aux cognats dont la ressemblance est complète, alors que le *calque morphologique* concerne les cognats dont la ressemblance est partielle à complète (Tableau 2.6).

<sup>16</sup> Cognats : « (de l'anglais *cognate*), des mots qui se traduisent l'un par l'autre et qui présentent une ressemblance graphique » (Kraif 1999 : 205).

Tableau 2.6. Différents degrés de cognation

Terme français	Terme anglais	Degré de cognation
<i>ozone</i> <i>CO<sub>2</sub></i>	<i>ozone</i> <i>CO<sub>2</sub></i>	ressemblance complète (isomorphisme)
<i>insecte</i> <i>régénération</i>	<i>insect</i> <i>regeneration</i>	ressemblance partielle à complète (calque morphologique)
<i>polaire</i> <i>aquatique</i>	<i>polar</i> <i>aquatic</i>	ressemblance partielle à insuffisante
<i>culture</i> <i>localement</i>	<i>cultivation</i> <i>locally</i>	ressemblance insuffisante
<i>dépérissement</i> <i>boisement</i>	<i>dieback</i> <i>afforestation</i>	aucune ressemblance

Avant de conclure cette section, il est important de signaler que les règles de transformations structurelles ne sont valides que pour les couples de langues analysées (ex. anglais-français) et le sens de l'analyse (ex. de la langue A vers la langue B). Autrement dit les observations faites ne peuvent pas toujours être étendues à d'autres langues et ne sont pas nécessairement réversibles.

### 2.3.4 Traitement des termes polysémiques et des homonymes

Dans les dictionnaires lexicographiques bilingues, il est d'usage d'envisager les unités lexicales d'une même forme graphique soit comme des unités polysémiques, soit comme des homonymes. En revanche, dans les dictionnaires de terminologie multilingue, on aura tendance à considérer les termes d'une même forme graphique comme des homonymes (Pic 2007–2008 : 66; Van Campenhoudt 1996 : 284). En terminologie, le dégroupement homonymique facilite le traitement des formes graphiques possédant plusieurs sens distincts. Van Campenhoudt (2001 : 9) précise que ce mode de représentation simplifie la création de ressources terminologiques multilingues. En effet, il arrive souvent qu'une même forme graphique partage plusieurs sens dans une langue  $L_1$ , alors que dans une langue  $L_2$  chacun de ces sens est représenté par une forme graphique différente. Dans le Tableau 2.7, trois termes anglais du domaine de l'hydrologie de même forme graphique (FLOOD), mais non synonymes, ont pour équivalent respectif *crue*, *inondation* et *flux* (les termes français ne sont pas synonymes entre eux non plus).



Tableau 2.7. Représentation de la forme graphique *flood* et de ses équivalents

Sens	Anglais	Français
(montée du niveau de l'eau au-dessus de la normale)	<i>flood</i> <sub>1</sub>	<i>crue</i>
(débordement d'eau)	<i>flood</i> <sub>2</sub>	<i>inondation</i>
(marée montante)	<i>flood</i> <sub>3</sub>	<i>flux</i>

### 2.3.5 Récapitulation

Dans cette section, nous avons constaté que l'équivalence en terminologie est beaucoup plus complexe que ne le présente généralement la terminologie classique – cette dernière, dans un souci de normalisation, cherche à évacuer les ambiguïtés, comme la polysémie et la synonymie (L'Homme 2004 : 27). Cependant, comme nous l'avons constaté, les typologies de l'équivalence en terminologie et en lexicographie se recoupent sur de nombreux points. Enfin, nous avons vu comment les sources terminologiques traitent les termes de même forme graphique.

En introduction de la section 2.3, nous avons présenté une définition de l'équivalence issue de l'optique conceptuelle et une autre de l'optique lexico-sémantique. La première met l'accent sur la notion, alors que la deuxième parle de « composante sémantique ». La première a pour conséquence de privilégier une approche extralinguistique, alors que la deuxième accorde l'avantage à la forme linguistique.

## 2.4 Conclusion du chapitre

Dans les sections 2.1 à 2.3, nous avons vu que la traductologie, la lexicographie et la terminologie nous offrent trois perspectives différentes sur l'équivalence, la première s'opposant plus nettement aux deux autres disciplines. En effet, des trois disciplines, la traductologie s'intéresse à une plus grande variété d'unités de traduction – de l'unité lexicale au texte, en passant par les phrases. Cette discipline couvre ainsi un plus grand spectre du phénomène, ce qui nous permet d'appréhender la richesse des langues et d'apprécier leurs différences. En outre, elle nous situe au-delà du système linguistique et nous permet également de comprendre l'importance du message. Le point de vue de la lexicographie est riche d'enseignements sur la complexité du phénomène de l'équivalence. À titre d'exemple, la typologie non exhaustive des différences entre les langues nous incite

à être plus attentifs devant cette réalité, même en terminologie. Les lexicographes décrivent abondamment les problèmes liés à la polysémie et à la synonymie lors de l'élaboration de dictionnaires bilingues de langue générale – une problématique souvent évacuée en terminologie. Enfin, les différents points de vue de la terminologie nous montrent que l'établissement de l'équivalence pose aussi des problèmes au niveau des unités terminologiques, qu'il suffise de penser aux phénomènes observés en corpus.

Nous constatons qu'il n'existe pas une définition de l'équivalence qui fasse consensus et que chacune des disciplines en offre plusieurs. Devant ces diverses visions, nous avons décidé de choisir parmi les approches de la traductologie, les sept procédés de traduction de Vinay et Darbelnet (1958) et la typologie des écarts (shift) de Catford (1965). Ces deux approches présentent de l'intérêt pour notre recherche dans ce sens qu'elles abordent plusieurs phénomènes qui peuvent être observés en corpus. De la lexicographie, nous retenons la leçon sur le non-isomorphisme des systèmes linguistiques, problème peu abordé en terminologie. Enfin, de la terminologie, nous prendrons comme point de départ l'équivalence terminologique telle qu'observée en corpus décrite à la section 2.3.2 (Carreño 2004 : 103; L'Homme 2004 : 208). Les choix théoriques que nous venons d'énumérer impliquent que notre démarche est essentiellement sémasiologique. Ce qui veut dire que notre étude vise à étudier l'équivalence telle qu'observée en corpus. Il en découle également que nous nous attendons à découvrir des phénomènes d'équivalence sans *a priori*.

Pour terminer, nous rappelons que notre étude se distingue des précédentes du fait que nous effectuons une analyse de l'équivalence à partir de termes simples anglais appartenant aux quatre parties du discours suivantes : nom, adjectif, verbe et adverbe, car à notre connaissance, en terminologie, aucune étude ne s'est intéressée uniquement aux termes simples et encore moins aux termes simples appartenant à ces quatre parties du discours.

### **3 Textes électroniques et corpus**

L'étude de l'équivalence que nous effectuons repose sur la construction d'un corpus parallèle et d'un corpus comparable. Afin de mieux élaborer ces corpus, il est nécessaire d'en connaître plus à fond les caractéristiques. Dans ce chapitre, nous nous penchons également sur un certain nombre de textes pour analyser le type de critiques formulées à l'endroit des traductions.

Dans ce chapitre, nous voyons en quoi les textes électroniques sont devenus des sources incontournables dans de nombreuses disciplines, dont la terminologie (3.1). Nous donnons une définition générale des corpus (3.2). Puis, nous présentons les corpus parallèles (3.2.1) et les corpus comparables (3.2.2). Nous effectuons une analyse approfondie des documents traitant des deux types de corpus, afin de répertorier les jugements portés sur les traductions (3.3). Dans la section Conclusion de l'analyse critique (3.4), nous présentons notre point de vue sur l'emploi des traductions en extraction de termes. Enfin, la section 3.5 constitue la conclusion de ce chapitre.

#### **3.1 Textes électroniques**

À l'ère de l'informatique, les documents électroniques augmentent de façon exponentielle et de nombreuses disciplines ont vite compris l'intérêt que présente l'utilisation des documents électroniques pour l'étude des phénomènes qui les intéressent. L'informatique, l'intelligence artificielle, le traitement automatique du langage (TAL), la linguistique de corpus, la terminologie, la traduction, les sciences de l'information, les sciences humaines (ex. science politique, sociologique, criminelle, littéraire) figurent parmi les disciplines qui se sont intéressées aux textes sur support électronique. L'analyse des textes par ordinateur comprend les méthodologies et les opérations informatiques effectuées sur ce type de support. Roberto Busa est considéré comme le pionnier dans le domaine grâce au Corpus Thomisticum<sup>17</sup>. Le projet, commencé en 1946, avait pour but de présenter un système de gestion de données capable de retrouver, de rassembler et de mettre en ordre les phrases, les

---

<sup>17</sup> Adresse du site Corpus Thomisticum : <http://www.corpusthomisticum.org/>.

citations, les ressemblances, les corrélations et les données statistiques des œuvres complètes de Saint Thomas d'Aquin. L'indexation des données dans le Corpus Thomisticum a débuté au moyen de fiches cartonnées, elle est ensuite passée à l'usage de cartes perforées, pour enfin passer à l'utilisation des ordinateurs.

En terminologie, les chercheurs ont, eux aussi, rapidement su tirer profit des outils informatiques. Ainsi, depuis que l'on dispose d'ordinateurs suffisamment puissants et d'une quantité formidable de documents électroniques à explorer, de nombreux chercheurs en terminologie ont constitué des corpus dans le but de les analyser dans tous les sens. Ces études sur le comportement des termes en corpus ont permis de mettre en évidence des phénomènes largement sous-estimés jusqu'ici. De nouvelles approches théoriques sont nées, en particulier la terminologie textuelle, approche qui, rappelons-le, critique l'approche wüstérienne et considère les textes comme le point de départ de l'étude, de l'analyse et de la collecte des termes (Bourigault et Slodzian 1999). Ainsi, ces nouvelles approches théoriques accordent une importance centrale aux corpus.

## 3.2 Corpus

Souvent mentionné jusqu'ici, le terme *corpus* n'a pas encore été défini comme tel dans notre thèse. En outre, lorsque nous en avons parlé, nous n'avons fait allusion qu'aux corpus électroniques. Il est à remarquer que cette notion n'est pas toujours appréhendée de la même manière chez les auteurs qui s'y intéressent, et c'est la raison pour laquelle, nous nous penchons sur cette question dans la présente section. Tout d'abord, il est important de rappeler qu'un corpus ne se présente pas obligatoirement sur support électronique et pourrait être constitué par un ensemble de textes, d'images, de graphiques, d'enregistrements sonores ou de vidéos. Cela dit, voyons maintenant, une définition propre à la linguistique, en l'occurrence celle de Sinclair (1996) :

A *corpus* is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language. (Sinclair 1996 : 4)

Cette définition nous indique que les corpus sont constitués uniquement de données linguistiques, qui peuvent se présenter sous forme d'œuvres complètes ou non. La notion de « critère » est également un élément clé dans cet énoncé – à titre d'exemple, le thème et le

genre font partie des critères que Sinclair évoque. Notons au passage que la définition ne fait pas référence au type de support. Maintenant, lorsque nous regardons du côté de la communauté du TAL, nous constatons qu'elle envisage le corpus à peu près sous le même angle que la linguistique, mais, cette fois-ci, seuls les corpus électroniques sont considérés. Une des définitions souvent employée en TAL nous vient à nouveau de Sinclair : « *A computer corpus is a corpus which is encoded in a standardized and homogenous way for open-ended retrieval tasks. Its constituent pieces of language are documented as to their origins and provenance* » (Sinclair 1996 : 4).

Des définitions qui précèdent, nous pouvons dégager un dénominateur commun essentiel : un corpus n'est pas un ensemble hétéroclite. En effet, les éléments d'un corpus sont sélectionnés selon des critères bien circonscrits (ex. domaine, genre, taille, registre) et dans un but précis. Zweigenbaum et Habert (2006 : 32) soulignent l'importance de connaître en quoi un corpus est représentatif afin de mieux interpréter les résultats. Il est à noter toutefois que, dans le monde du TAL, selon les différents objectifs de recherche ou selon le type d'analyse, les critères de sélection seront plus ou moins pris en compte et parfois ils seront complètement ignorés. Surtout lorsqu'il est nécessaire de travailler avec des corpus de très grande taille, par exemple pour améliorer la performance des outils. Dans ce dernier cas, plusieurs considèrent ce type de corpus comme un ensemble de textes, sans plus (Sinclair 1996 : 5). Point de vue important dont il faut tenir compte, surtout lorsque nous étudions les travaux en extraction sur corpus comparables (*cf.* 4.3).

La définition de Sinclair (1996 : 4) nous a appris qu'il existe une quantité importante de types de corpus. Parmi les corpus constitués de textes, nous avons également l'embaras du choix : unilingue ou multilingue; ouvert ou fermé; synchronique ou diachronique; en langue spécialisée ou générale; etc. Dans cette thèse, nous nous intéressons aux corpus multilingues électroniques, plus précisément, aux corpus parallèles et aux corpus comparables spécialisés.

### **3.2.1 Corpus parallèles**

Les corpus parallèles sont constitués de paires de textes source (original) et cible (traduction). Les textes parallèles existent depuis fort longtemps, bien avant l'apparition de

l'informatique. Le plus connu est sans conteste la pierre de Rosette. Ce fragment de stèle en granite noir sur lequel figuraient trois systèmes d'écriture (hiéroglyphe, démotique et grec) a permis à Champollion de déchiffrer l'écriture hiéroglyphique. En dehors de cette stèle de granite, on retrouve de nombreux types de textes parallèles que ce soit des traités, des contrats, des textes sacrés, des ouvrages littéraires ou des documents scientifiques. L'informatique ayant facilité l'exploitation des textes parallèles : on a pensé les utiliser notamment en traduction automatique, car, comme le note Isabelle (1992), ils constituent une mine de solutions de traduction réutilisables.

De façon générale, un corpus parallèle<sup>18</sup> se compose d'un ensemble de couples de textes, chaque couple étant constitué d'un texte source et de sa traduction. L'intérêt particulier de ce type de corpus, c'est qu'ils peuvent être alignés, le plus souvent, au niveau des phrases. En général, ces corpus sont bilingues<sup>19</sup>, mais plus de deux langues peuvent intervenir. À cette définition de base, Sinclair (1996 : 12) ajoute qu'à l'intérieur d'un même corpus il n'est pas nécessaire que la direction de la traduction soit toujours la même, par exemple, des textes dans une langue  $L_1$  ou  $L_2$  peuvent être soit des textes source soit des traductions. Selon nous, cette précision revêt une grande importance, car elle signifie que, lorsqu'on travaille avec des corpus parallèles dont on ne connaît pas le sens de la traduction, on attribue la même valeur qualitative aux textes sources et aux textes cibles. De fait, la présence des mémoires de traduction montre que l'utilisation des textes cibles n'est pas toujours condamnée (Bourigault *et al.* 1999 : 70). D'autre part, il existe des textes parallèles où « [l]es deux langues font également foi et [où] il est impossible de distinguer une langue source et une langue cible » (*ibid* : 72), comme c'est le cas pour de nombreux textes de loi, dont la Convention de sauvegarde des droits de l'homme et des libertés fondamentales dans laquelle il est spécifié : « Fait à Rome, le 4 novembre 1950, en français et en anglais, les deux textes faisant également foi, en un seul exemplaire qui sera déposé dans les archives du Conseil de l'Europe »<sup>20</sup>.

<sup>18</sup> De nombreux auteurs (Harris 1988; Kraif 2001; L'Homme 2005a) utilisent le terme *bi-texte*, *bitexte* ou encore *textes alignés* pour désigner le corpus parallèle.

<sup>19</sup> Deleger (2009 : 17) cite également les corpus parallèles dans un contexte monolingue, par exemple, on peut aligner les différentes versions d'un même texte source.

<sup>20</sup> Convention de sauvegarde des droits de l'homme et des libertés fondamentales. <http://conventions.coe.int/Treaty/fr/treaties/html/005.htm> (page consultée le 29 août 2012).

Plusieurs auteurs critiquent l'emploi des corpus parallèles, notamment lorsqu'il s'agit d'extraire des termes ou des unités lexicales (Déjean et Gaussier 2002; Maia 2003; McEnery et Xiao 2007; Teubert 1996). De façon générale, ils invoquent le fait que le traducteur, sous l'influence de la langue source, produit des textes dans une langue moins riche que celle que l'on trouve dans les textes originaux. Ces critiques ayant soulevé notre intérêt, nous estimons qu'elles méritent d'être examinées de plus près (*cf.* 3.3). Cependant, pour avoir une meilleure vue d'ensemble, passons en premier lieu à la définition des corpus comparables.

### 3.2.2 Corpus comparables

Par définition, les corpus comparables sont constitués de textes traitant du même sujet, mais écrits dans différentes langues<sup>21 22</sup>. Par extension, on peut présumer que les textes qui les composent comportent des termes dont les usages sont comparables entre les langues.

D'un point de vue théorique, la majorité des chercheurs jugent préférable l'emploi des corpus comparables aux corpus parallèles, car :

L'utilisation de ces textes permet en effet d'avoir un accès à une langue dont on peut supposer que sa phraséologie sera plus riche que celle d'une langue traduite que d'aucuns nomment « translationese » pour en souligner le caractère non idiomatique (Maniez 2001 : 552).

À part le fait que certains auteurs discréditent les corpus parallèles, d'un point de vue plus pratique, on dit aussi que les corpus parallèles « de bonne qualité »<sup>23</sup> sont des ressources moins faciles à construire que les corpus comparables (Chiao 2004; Déjean et Gaussier 2002; Fung 1998), en particulier pour certains domaines spécialisés et pour des couples de langues dans lesquelles ne figure pas l'anglais (Morin 2009 : 1). De plus, selon Fung et Yee (1998 : 414–415), les corpus comparables présentent un autre avantage lorsqu'on recherche

---

<sup>21</sup> Un corpus comparable peut également être monolingue, par exemple, « composés de textes destinés pour les uns à des spécialistes et pour les autres au grand public » (Deléger 2009 : 3), ou encore, de textes originaux en langue L<sub>1</sub> et de traductions dans la langue L<sub>1</sub> et dans des domaines apparentés afin d'en comparer les différences (Baker 1993 : 2).

<sup>22</sup> Dans notre thèse, le sous-corpus français contenant les traductions et celui contenant les originaux est considéré comme un corpus comparable monolingue (*cf.* Figure 5.1).

<sup>23</sup> « De bonne qualité », expression tirée du texte de Déjean et Gaussier (2002 : 3).

les néologismes, par exemple, les nouveaux noms propres y figurent plus rapidement que dans les textes parallèles, tels les noms des personnalités dans les journaux.

Toutefois, à l'étape de la constitution des corpus comparables, contrairement aux corpus parallèles, se pose le problème de la comparabilité ou de la similarité des textes. Pour Maia (2003 : 2–4), la comparabilité est assurée par la forme et le contenu du corpus. La forme touche le nombre de textes et le nombre de mots de l'ensemble du corpus ainsi que la nature des différents textes (format, images, etc.). Le contenu concerne la structure des textes, le registre de langue, les sujets abordés, etc. De leur côté, Déjean et Gaussier (2002 : 2) se basent sur deux types de critères pour constituer les corpus comparables : qualitatifs et quantitatifs. Le premier rejoint dans les grandes lignes ceux de Maia (2003). Le deuxième s'appuie sur le nombre plus ou moins important du vocabulaire commun des deux sous-parties du corpus. Enfin, d'après Gœuriot (2009 : 123), les corpus comparables spécialisés doivent être constitués de textes appartenant au même domaine et au même type de discours. Pour créer automatiquement ce genre de corpus, Gœuriot sélectionne en premier des textes du domaine choisi avec des mots clés. Ensuite, pour retenir les textes ayant le type de discours voulu, elle applique divers filtres : structurels (format, structure), modaux (présence de locuteur et d'interlocuteur), lexicaux (vocabulaire, unité de mesure, symbole).

Comme nous venons de le voir, pour créer des corpus comparables plusieurs paramètres doivent être pris en compte. Cependant, ainsi que le fait remarquer Maia (2003 : 3), l'application de tous ces paramètres, qui semblent facilement maîtrisables sur le plan théorique, se fait beaucoup plus difficilement dans la pratique, d'autant plus avec des corpus de très grande taille.

### **3.3 Analyse des critiques touchant les corpus parallèles**

Comme nous avons déjà eu l'occasion de le signaler plus haut, l'utilisation de textes traduits (donc de corpus parallèles) fait l'objet de nombreuses critiques dans les textes traitant des corpus. Étant donné que ces critiques sont formulées par des chercheurs



œuvrant dans des domaines différents, nous avons divisé les documents<sup>24</sup> en deux groupes. Le premier groupe, que nous nommerons A (33 documents<sup>25</sup>), est très homogène et renferme les textes décrivant des techniques d'extraction de terminologies ou de lexiques sur corpus comparables. De son côté, le deuxième groupe, nommé B (23 documents<sup>26</sup>), est très hétérogène, en ce sens que les textes abordent, généralement d'un point de vue théorique, divers sujets touchant les corpus parallèles et/ou comparables.

Avant tout, que cela touche le groupe A ou le groupe B, il est important de noter que la terminologie entourant les corpus parallèles et les corpus comparables diffère parfois d'un texte à l'autre (Pearson 1998 : 47–48). Ainsi, les définitions s'appliquant aux corpus comparables ou aux corpus parallèles ne sont pas encore tout à fait normalisées, d'où l'importance de savoir ce que chaque auteur entend par *parallèle* et *comparable*. À titre d'exemple, un corpus comparable tel que décrit au point 3.2.2 est nommé *non-parallel corpora* chez Fung (1995) et *monolingual corpora* chez Haghghi *et al.* (2008). Baker (1993) définit un corpus comparable comme un ensemble de textes originaux en langue  $L_1$  et un ensemble de traductions en langue  $L_2$ , c'est-à-dire que tous les textes sont écrits dans une même langue (*cf.* note 21). Hartmann (1996) distingue trois types de corpus parallèles : a) des originaux en  $L_1$  et des traductions en  $L_2$ , b) des originaux en  $L_1$  et des adaptations en  $L_2$ , c) des originaux en  $L_1$  et en  $L_2$ , mais de mêmes genres ou de mêmes domaines<sup>27</sup>.

### 3.3.1 Analyse du groupe A

Des 33 textes du groupe A (textes décrivant les techniques d'extraction sur corpus comparables), 10 portent un jugement négatif sur les corpus parallèles, 22 ne font de

---

<sup>24</sup> Nous parlons à dessein de documents et non pas d'auteurs, en effet, il peut arriver qu'un auteur émette une opinion différente d'un texte à un autre.

<sup>25</sup> Documents du groupe A : Chiao 2004; Chiao et Zweigenbaum 2002; Chiao *et al.* 2004; Déjean et Gaussier 2002; Déjean *et al.* 2005; Deléger 2009; Fung 1995, 1998; Fung et McKeown 1997; Fung et Yee 1998; Gamallo 2008; Garera *et al.* 2009; Gaussier *et al.* 2004; Haghghi *et al.* 2008; Koehn et Knight 2002; Langlais *et al.* 2009; Laroche et Langlais 2010; Laws *et al.* 2010; Léon 2008; Morin 2009; Morin *et al.* 2004; Morin et Daille 2006; Prochasson 2008, 2009, 2009a; Prochasson et Morin 2009; Rapp 1999; Sadat *et al.* 2003; Shahzad *et al.* 1999; Shao et Ng 2004; Tanaka et Iwasaki 1996; Zweigenbaum 2006; Zweigenbaum et Habert 2006.

<sup>26</sup> Documents du groupe B : Baker 1993, 1995, 1996, 1998; Bekavac *et al.* 2004; Biber 1993; Bowker et Pearson 2002; Gæuriot 2009; Hartmann 1996; Johansson 2000; Laviosa 1997, 1998; Maia 2003; Maniez 2001; Mauranen 2002; McEnery et Xiao 2007; Pearson 1998; Philip 2009; Sinclair 1996; Teubert 1996, 2001; Véronis 2000a; Zanettin 1998.

<sup>27</sup> Pour nous, le corpus de la catégorie c) est un corpus comparable comme défini à la section 3.2.2.

commentaires ni dans un sens ni dans l'autre, 1 précise tout au plus que les corpus parallèles sont utiles, mais qu'il est nécessaire d'exploiter les corpus comparables (Fung et McKeown 1997 : 192–193). Parmi les textes critiquant les traductions, tous sont de source française à l'exception de celui de Fung (1995).

Comme on peut s'y attendre, la préoccupation principale des auteurs porte sur les mots, plus précisément le choix des équivalents. Les autres aspects de la traduction sont peu ou pas abordés : syntaxe, contexte de production, questions sociales et culturelles, etc. Bien entendu, nous ne devons pas oublier que, dans le domaine du TAL, ce sont les mots (suites graphiques délimitées par des blancs) qui sont principalement manipulés. Les critiques à l'endroit des traductions sont présentées comme des faits admis par tous et les auteurs, à l'exception de Prochasson (2009) et de Léon<sup>28</sup> (2008), ne jugent pas nécessaire de citer des sources pour justifier leurs affirmations. Curieusement, aucun des auteurs ayant critiqué les traductions ne présente, dans leur méthodologie, les moyens mis en œuvre pour écarter les traductions des corpus. A fortiori, aucun ne dit s'il est difficile ou non de vérifier le statut d'un texte (original ou traduction). À titre d'exemple, étant donné que Léon (2008) critique les traductions, nous aurions aimé savoir s'il s'était assuré que les 4 160 000 pages Web de son corpus provenaient réellement de textes originaux.

De façon plus spécifique, parmi les auteurs qui critiquent les traductions, plusieurs estiment qu'elles risquent de ne pas être représentatives de la langue cible<sup>29</sup>, autrement dit, selon ces auteurs, les traducteurs sous l'influence de la langue source produiraient des textes non idiomatiques (Déjean et Gaussier 2002 : 3; Deléger 2009 : 18; Léon 2008 : 83; Prochasson 2009 : 6). Cela se traduirait principalement par le manque d'exactitude du vocabulaire trouvé dans les traductions. Pour Deléger (2009 : 18) et Chiao (2004 : 33), la langue source influence le choix des équivalents. Déjean et Gaussier (2002 : 3) vont dans le même sens et précisent que les traducteurs sont susceptibles de privilégier les cognats<sup>30</sup>. Zweigenbaum et Habert (2006 : 22), de leur côté, parlent des risques d'y trouver des

---

<sup>28</sup> Prochasson (2009 : 11–12) fait mention de Fung (1995), et Léon (2008 : 80–83) cite Déjean et Gaussier (2002), Fung (1998), Maniez (2001) et Véronis (2000a). Les textes de ces auteurs font l'objet d'une analyse soit dans le groupe A, soit dans le groupe B.

<sup>29</sup> De Chesterman (2004 : 36–39), voir le chapitre « Beyond the particular » dans lequel il se penche sur les différentes approches utilisées pour décrire les traductions.

<sup>30</sup> Les cognats ne constituent pas des fautes de langue.

calques<sup>31</sup> et des « biais de traduction ». À titre d'exemple, les auteurs expliquent que le mot anglais *consistent* (qui veut dire *cohérent*) risque d'être traduit par le mot français *consistant*<sup>32</sup>.

### 3.3.2 Analyse du groupe B

Le groupe B (23 documents), comme nous l'avons dit, est beaucoup plus hétérogène que le groupe précédent. Le plus souvent, les auteurs décrivent les corpus parallèles et/ou les corpus comparables d'un point de vue théorique. Néanmoins, en ce qui a trait aux jugements touchant les traductions, il est possible de les diviser en quatre catégories : *négatif*, *nuancé*, *aucun jugement de valeur*, *positif*. Des 23 documents du groupe B, nous en avons recensé 9 dans la catégorie *négatif*, 1 dans la catégorie *nuancé*, 8 dans *aucun jugement de valeur* et 5 dans la catégorie *positif*. Pour ce dernier groupe, remarquons au passage que 4 proviennent du même auteur (Baker 1993, 1995, 1996, 1998).

Des 9 textes de la catégorie *négatif* (Bekavac *et al.* 2004; Gœuriot 2009; Hartmann 1996; Maia 2003; McEnery et Xiao 2007; Philip 2009; Sinclair 1996; Teubert 1996; Zanettin 1998), nous présentons deux citations qui nous semblent les plus représentatives de la catégorie. La première, celle de Sinclair (1996), est tirée d'un texte très connu et souvent cité, en l'occurrence par Bekavac (2004) :

The possibilities of a comparable corpus are to compare different languages or varieties in similar circumstances of communication, but avoiding the inevitable distortion introduced by the translations of a parallel corpus. (Sinclair 1996 : 12)

Le jugement de Sinclair est sans appel et présenté comme une vérité qui s'impose à tous. La deuxième citation nous vient de Teubert (1996) et va dans le même sens, de plus elle est particulièrement accablante et mérite pour cela d'être citée au complet :

---

<sup>31</sup> Le calque est la traduction littérale d'une langue à l'autre d'un mot. Par exemple, *honeymoon* donne *lune de miel*. À l'inverse des cognats, de nombreux calques sont à éviter, comme *sand paper* qui est parfois rendu par *papier sablé* au lieu de *papier de verre*.

<sup>32</sup> L'exemple donné par Zweigenbaum et Habert (2006 : 22) s'appelle un faux ami. Les faux amis sont des mots qui présentent une ressemblance formelle, mais dont les sens sont différents, un des exemples souvent cités est le mot anglais *library* (en français *bibliothèque*) et le mot français *librairie* (en anglais *bookstore*). Les faux amis sont à proscrire.

There is one essential objection to parallel corpora. Translations, however good and near-perfect they may be (but rarely are), cannot but give a distorted picture of the language they represent. Linguists should never rely on translations when they are describing a language. That is why translations have no place in reference corpora. Rather than representing the language they are written in, they give a mirror image of their source language. [...] This is true both for vocabulary and for syntax. (Teubert 1996 : 247)

Cet extrait nous rappelle le fameux jeu de mots « traduttore, traditore ». Quoi que fasse le traducteur, il sera toujours critiqué. Néanmoins, Teubert est quand même d'avis que les corpus parallèles peuvent rendre certains services et être utilisés en complémentarité avec les corpus comparables composés à partir de textes originaux. Dans un même ordre d'idée, mais sur un ton plus mesuré, Maia (2003 : 2) critique l'usage des corpus parallèles pour extraire des termes, car leur fiabilité dépend largement de la qualité des traductions.

Dans l'unique texte de la catégorie *nuancé* (Maniez 2001), l'auteur présente ses arguments avec prudence. Il reconnaît que les corpus bilingues ont leur place dans la recherche des équivalents, mais propose de compléter les analyses avec des corpus parallèles bidirectionnels ou mieux, des corpus comparables pour les raisons suivantes :

Toutefois, dans la mesure où les traducteurs choisissent souvent la traduction littérale quand elle est possible et évitent de traduire certains verbes par de stricts équivalents syntaxiques afin de contourner certaines difficultés inhérentes à leur traduction, la partie du corpus bilingue qui correspond à la langue d'arrivée est parfois d'une moins grande richesse lexicale que celle correspondant à la langue de départ. (Maniez 2001 : 562)

Nous avons remarqué que Maniez utilise toujours le même ton mesuré, que ce soit dans la citation apparaissant à la section 3.2.2 ou dans d'autres parutions, par conséquent il est bien important de ne pas lui prêter de jugements définitifs sur la traduction. De plus, dans la citation qui précède, son évaluation sur la richesse lexicale de la traduction est comparée au texte source et non aux textes appartenant à la même langue que la traduction, comme le font de nombreux auteurs.

Dans les 8 documents de la catégorie *aucun jugement de valeur* (Biber 1993; Bowker et Pearson 2002; Johansson 2000; Laviosa 1997; Laviosa 1998; Pearson 1998;

Teubert 2001; Véronis 2000a), les auteurs ne portent pas de jugement de valeur sur les traductions lorsqu'ils décrivent les corpus parallèles ou comparables. Dans son texte, Teubert (2001) fait valoir les apports de la linguistique de corpus en lexicographie. L'auteur s'attarde plus particulièrement sur les corpus parallèles et le repérage des équivalents. Par ailleurs, contrairement à son article de 1996, il ne critique plus les traductions. Dans son ouvrage, Pearson (1998 : 47–48) se penche sur l'utilité des corpus dans des applications terminologiques. L'auteure décrit les corpus parallèles et les corpus comparables sans porter de jugement de valeur sur la traduction. Dans son texte, Véronis (2000a) espère que les corpus parallèles soient plus utilisés malgré les réticences des lexicographes :

First of all, there is some reluctance on the part of traditional lexicographers to use translation, which are often not regarded as authentic discourse acts but artefacts (see Knowles, 1996) – but given that the purpose of bilingual dictionaries is precisely, or at least partly, to produce such artefacts, this reluctance should diminish in the future. (Véronis 2000a : 10)

Dans cette citation, ce n'est pas Véronis qui porte un jugement négatif à l'endroit des traductions, comme pourraient le laisser sous-entendre certains auteurs (*cf.* note 28). En fait, Véronis ne fait que rapporter le sentiment des lexicographes. De son côté, Laviosa (1997), qui s'intéresse aux *universaux traductionnels*, présente une méthodologie pour construire des corpus comparables monolingues et affirme avoir mis au jour quelques universaux. D'après certains auteurs, les universaux traductionnels sont des régularités linguistiques apparaissant dans les textes traduits quelles que soient les langues considérées. L'objectif de leur découverte n'est pas de critiquer négativement ou positivement les traductions, mais de les distinguer des originaux, ce qui permettrait entre autres de justifier l'existence de la traductologie en tant que science à part entière. En 1993, Baker (1996) voyait, avec la linguistique de corpus, un moyen de vérifier à grande échelle l'existence des universaux. Ces derniers seraient de deux ordres, les *S-universal* et les *T-universal* – S pour *source* (source) et T pour *target* (cible) (Chersterman 2004 : 39–47). Les S-universal font référence aux caractéristiques qui distinguent les traductions des textes source, par exemple, l'explicitation, la normalisation, la réduction des répétitions, la désambiguïsation. Les T-universal seraient des différences systématiques entre les traductions et les originaux de la langue cible, par exemple, une variété lexicale moins riche et une sous-représentation

des traits spécifiques de la langue cible dans les traductions. Quelques travaux semblent confirmer l'existence des deux types d'universaux (Gellerstam 1996; Laviosa 1997, 1998), alors que d'autres les infirmeraient (Corpas Pastor *et al.* 2008; House 2008). Par exemple, dans une étude portant sur les universaux, Tirkkonen-Condit (2002) conclut que leur identification n'est pas évidente :

[...] translations are not readily distinguishable from original writing on account of their linguistic features. (Tirkkonen-Condit 2002: 210)

De nombreux auteurs mettent en doute leur existence ou les réfutent carrément (Chesterman 2008; Tirkkonen-Condit 2002; Tymoczko 1998). Pour la grande majorité des théoriciens en traduction, sinon tous, les universaux ne sont pas des formes fautives de traduction. Toutefois, selon Chesterman (2004 : 39), les universaux pourraient être perçus par certains de façon négative, en particulier les T-universal<sup>33</sup>.

La catégorie *positif* est représentée par cinq articles, quatre de Baker (1993, 1995, 1996, 1998) et un de Mauranen (2002). Deux passages de Baker (1993; 1998) ont particulièrement attiré notre attention. Dans le premier, Baker constate avec regret le peu de crédit accordé à la traduction :

Given that translated texts play such an important role in shaping our experience of life and our view of the world, it is difficult to understand why translation has traditionally been viewed as a second-rate activity, not worthy of serious academic enquiry, and why translated texts have been regarded as no more than second-hand and distorted versions of 'real' texts. (Baker 1993 : 233)

Dans le deuxième passage, Baker (1998) clarifie encore plus sa position sur les traductions :

Il n'est pas question ici de parler de «jargon de traduction» (*translationese*) avec toutes ses connotations péjoratives. La traduction crée un troisième code parce qu'elle est une forme de communication unique, et non parce qu'elle est une forme de communication fautive, déviante ou non conforme à la norme. (Baker 1998 : 3)

---

<sup>33</sup> Le terme anglais pour *universaux traductionnels* est *translationese*.

Mauranen (2002 : 161–185), quant à elle, s’en prend à la citation<sup>34</sup> de Teubert (1996) et conclut son texte par le raisonnement suivant : « [...] translated language constitutes a natural and important part of any language and therefore should find a place in a general corpus which aims at representativeness of the language as a whole » (Mauranen 2002 : 181).

### 3.3.3 Récapitulatif des analyses des groupes A et B

Tableau 3.1. Catégories offrant un jugement de valeur des groupes A et B

Groupe	Catégorie	Nombre
A	négatif	10
B	négatif	9
B	nuancé	1
B	positif	5

Si l’on regroupe les textes (A et B) dans lesquels un jugement de valeur est porté, le nombre des jugements négatifs (19) est largement supérieur à celui des positifs (5). Cela ne nous étonne pas puisque les critiques à l’endroit des traductions existent depuis toujours, ainsi que le rappelle Mauranen (2002 : 161) « Their [translation] value as suitable data has tended to be controversial throughout ». Dans les groupes A et B, on s’inquiète en particulier du risque d’employer des traductions de mauvaise qualité. Cette inquiétude est justifiée, mais selon nous les raisons invoquées pour exclure en bloc les traductions des corpus sont trop souvent basées sur des *a priori*. De façon plus spécifique, on reproche aux traducteurs de prioriser les cognats, de commettre des calques et de tomber dans le piège des faux amis. Une autre raison parfois avancée pour écarter les traductions serait qu’elles sont coupables de *translationese*<sup>35</sup>.

## 3.4 Conclusion de l’analyse critique

Comme nous l’avons dit plus haut, il est légitime de vouloir sélectionner des textes de bonne qualité pour constituer des corpus. Pour ce faire, deux stratégies principales peuvent

<sup>34</sup> Celle incluse ci-dessus, dans la section *négatif*.

<sup>35</sup> Pour certains le terme *translationese* a une connotation péjorative (Baker 1998; Tirkkonen-Condit 2002). On se souviendra que Baker lui donne pour équivalent *jargon de traduction*. Pour d’autres, ce terme est plus ou moins synonyme d’*universaux traductionnels*<sup>35</sup> (Gœuriot 2009; Hartmann 1996; McEnery et Xiao 2007).

être adoptées. La première stratégie, généralement employée par ceux qui tiennent pour acquis que les traductions ne sont pas fiables, les écarte systématiquement de leur corpus. La deuxième stratégie, celle que nous privilégions, consiste à choisir les traductions et les textes originaux à partir de critères de sélection rigoureux.

Dans la première stratégie, les traductions sont écartées sous prétexte qu'elles sont plus susceptibles de contenir des fautes linguistiques (calques, faux amis, etc.) ou qu'elles seraient différentes des originaux. Cette stratégie, quoique radicale, peut sembler valable. Cependant, lorsque cette solution est adoptée, il est essentiel de présenter dans la méthodologie les moyens pris pour ne pas inclure de traductions dans le corpus<sup>36</sup>. À ce sujet, il n'est pas toujours facile de savoir si un ouvrage est une traduction ou pas, cette information n'est pas toujours mise en évidence; parfois elle est complètement absente. Un autre aspect peu envisagé est que les ouvrages collectifs sont souvent hybrides – certains chapitres ou articles sont des traductions alors que d'autres sont des originaux. D'autre part, les corpus construits avec des articles de journaux posent un problème majeur. Comment peut-on s'assurer que ce sont réellement des originaux lorsqu'on sait que les journalistes tirent souvent leurs informations des communiqués des agences de presse, Thomson Reuters entre autres? Toutes ces vérifications prennent du temps et constituent un inconvénient majeur lors de la construction des corpus, surtout pour ceux de très grande taille. Maintenant, à supposer que l'on soit en mesure de procéder à toutes ces vérifications, il reste un autre problème. Certaines traductions ont autant de poids sinon plus, du point de vue terminologique, que certains originaux. Par exemple, dans le domaine du changement climatique, les rapports du GIEC sont des traductions. Cependant, ces textes sont la source de très nombreux articles scientifiques et la terminologie qu'ils contiennent est reprise par beaucoup<sup>37</sup>. Que faire dans ces cas-là? Tolérer une entorse à la méthodologie de la première stratégie ou se priver d'une ressource de premier plan?

La deuxième stratégie, celle que nous privilégions, consiste à choisir les traductions avec des critères de sélection rigoureux (auteur, provenance, registre, thème, etc.), comme

---

<sup>36</sup> Nous rappelons qu'aucun des textes cités faisant une critique négative des traductions ne présente dans sa méthodologie de stratégies à cet effet.

<sup>37</sup> Les traductions ayant une importance majeure dans divers domaines scientifiques sont très courantes, à commencer par le texte sur l'évolution des espèces de Darwin.



on le fait en principe pour les originaux. Nous convenons que cette méthode n'est pas infaillible, mais elle a le mérite de permettre l'inclusion de traductions incontournables comme celles du GIEC dans les corpus et d'exclure celles qui ne remplissent pas les critères de sélection. D'ailleurs, cette méthode est également employée avec les originaux afin d'éviter, entre autres, les textes de mauvaise qualité. Ici, nous aimerions attirer l'attention sur les deux points suivants. Premièrement, étant donné que les auteurs de textes spécialisés lisent et citent fréquemment des textes de langues étrangères, nous serions en droit de nous demander s'ils ne risquent pas, comme les traducteurs, d'introduire dans la langue de nouvelles caractéristiques linguistiques et ainsi dénaturer « l'usage réel de la langue » comme le suggèrent Sanz Vicente et García Palacios (2010 : 195–196) et Vandaele (2001 : 117). Enfin, deuxième point et non le moindre, les auteurs réputés dont ce n'est pas la langue maternelle peuvent produire des originaux dans lesquels ils glisseraient involontairement un vocabulaire ou des tournures inhabituelles qui ne serait pas « représentatifs de la langue ». Devrait-on exclure ces textes des corpus également, surtout si ces textes répondent aux critères de sélection?

Comme on le voit, le concept exprimé par les expressions *usage réel de la langue*, *représentatif de la langue* ou encore *langue originale* est assez vague. Premièrement, on pourrait se demander en quoi consiste une « langue originale ». Par rapport à qui ou selon quelles références se base-t-on? On devine que les réponses ne sont pas définitives et encore moins partagées par tout le monde. Cependant, une chose est certaine, nous vivons quotidiennement avec des traductions. La plupart du temps, nous ne nous en rendons même pas compte et nous ne nous préoccupons pas de savoir si un texte est un original ou pas. Si les traductions venaient à disparaître, notre culture aurait un tout autre visage comme l'explique Confiant (2003) :

Je ferai d'abord deux remarques : tout d'abord qu'entre 70 et 90 % de notre savoir nous est acquis grâce à des traductions. Peu d'entre nous ont lu, ou sont capables de lire dans le texte Platon, Aristote, Spinoza, Marx, Freud, Bakhtine ou Chomsky. Or, dans le même temps, la traduction jouit d'une considération fort médiocre à nos yeux. (Confiant 2003 : 1)

En plus des grands auteurs cités par Confiant, songeons à ce que nous apportent les traductions des films, des émissions de télévision, des nouvelles et des ouvrages en tout genre (scientifiques, techniques, guides, littéraires, etc.).

D'autre part, plusieurs auteurs affirment que les corpus parallèles sont plus difficiles à construire dans certains domaines que les corpus comparables parce que les textes parallèles sont en moins grand nombre que les textes originaux. Cela est vrai, surtout pour les langues des cultures possédant une expertise technique et scientifique développée. Par contre, c'est moins vérifié pour les cultures maîtrisant peu les nouvelles technologies. En effet, les langues de ces cultures ne possèdent pas leur propre vocabulaire pour désigner toutes les réalités modernes – scientifique, technique, commercial (Cronin 2003). On estime qu'il existe actuellement plus de 6 000 langues parlées à travers le monde et que plus de la moitié sont en danger de disparition. La survie de ces langues dépend de nombreux facteurs, dont le recours à la traduction, afin de rester fonctionnelles, comme l'exprime si bien Confiant (2003) :

La traduction semble être la voie royale pour que les langues accèdent à ce que Jean Bernabé appelle la souveraineté scripturale. C'est qu'elle oblige la langue à sortir de son cocon, des réalités qu'elle a l'habitude de désigner et la force à actualiser ses potentialités cachées. (Confiant 2003 : 6)

Par conséquent, nous pensons que de montrer du doigt les traductions est préjudiciable. Les traductions autant que les originaux, avant d'entrer dans un corpus, devraient être sélectionnés sur des critères objectifs – texte entier ou extrait, écrit ou oral, thème, registre, année de publication, type de destinataire, expertise de l'auteur, etc. Nous ne disons pas que les traductions doivent toujours apparaître dans un corpus, mais qu'on ne peut les écarter sur de simples idées reçues. Qu'elles soient utilisées ou non relève du cas par cas : du type de recherche, de leur disponibilité dans un domaine particulier, du couple de langue étudié, etc.

Pour être représentatif d'une langue, Biber (1993) dit qu'un corpus devrait contenir différents registres. Dans le même ordre d'idée, nous avançons que si nous voulons qu'un corpus soit vraiment représentatif d'une langue, les traductions ne devraient pas d'emblée en être exclues.

### **3.5 Conclusion du chapitre et contribution à la recherche**

Dans ce chapitre, nous avons vu qu'en terminologie on peut tirer avantage des documents électroniques pour l'étude, l'analyse et la collecte des termes. Nous avons expliqué en quoi consistent les corpus parallèles et les corpus comparables. Nous avons vu qu'il n'est pas toujours aussi facile de construire un corpus comparable et abordé certains des aspects à considérer lors de leur élaboration. Nous avons constaté qu'en extraction terminologique, l'usage des corpus parallèles est critiqué par de nombreux auteurs et qu'on lui préfère les corpus comparables. Dans l'analyse des critiques touchant la traduction, nous avons observé qu'elles sont le plus souvent négatives. Les auteurs accusent notamment les traducteurs de faire des emprunts, des faux sens et d'utiliser des cognats plus souvent que dans les textes originaux. Dans la conclusion de l'analyse critique, nous avons pris la défense des traductions en évoquant le fait qu'elles font autant partie de l'univers textuel d'une langue que les originaux. Nous avons également avancé que lors de la construction d'un corpus on ne devrait pas exclure les traductions.

Dans notre thèse, un de nos objectifs est de mesurer les différences et les ressemblances terminologiques entre les corpus parallèles et les corpus comparables afin de jeter un éclairage nouveau sur ce débat. Notre étude, qui se veut objective, met en œuvre des outils informatiques et analyse un nombre conséquent de données afin de vérifier les observations négatives à l'endroit des traductions formulées par les auteurs, mais à notre connaissance jamais démontrées.

## **4 TECHNIQUES D'EXTRACTION LEXICALE ET TERMINOLOGIQUE**

Pour tirer parti des corpus, il existe de nombreux outils informatiques : concordanciers, étiqueteurs de textes, extracteurs de termes ou de lexiques, aligneurs de textes, analyseurs syntaxiques, etc. Parmi les nombreux outils qui s'offrent aux chercheurs, certains sont suffisamment performants alors que d'autres n'en sont qu'à leurs balbutiements. Ainsi, les extracteurs terminologiques et lexicaux sur corpus parallèles sont maintenant passés du domaine de la recherche au domaine commercial, par contre, il n'en va pas de même pour l'extraction de termes à partir de corpus comparables.

Dans ce chapitre, nous décrivons tout d'abord les principales caractéristiques des extracteurs de termes en corpus monolingue (4.1) et les techniques d'extraction bilingue (lexicale et terminologique) en corpus parallèle (4.2), afin de mieux interpréter les listes obtenues à l'aide de ces outils (*cf.* Chapitre 5 et Chapitre 6). Ensuite, nous procédons à une analyse détaillée des approches adoptées pour identifier les équivalents sur corpus comparable (4.3). Cette dernière description vise à tirer parti de certaines techniques de mise en équivalence sur corpus comparables et à circonscrire les points sur lesquels notre étude serait susceptible de contribuer à la recherche en extraction automatique sur corpus comparable.

### **4.1 Extraction de termes en corpus monolingue**

L'extraction de ressources terminologiques à partir de corpus unilingues est maintenant un classique qui a fait ses preuves, même s'il reste de la place à l'amélioration (Vintar 2010 : 142). Sur des corpus de grande taille, l'usage des extracteurs de termes permet de gagner du temps et de systématiser le dépouillement (Estopà 2001 : 219).

Le principal défi que doivent relever ces outils repose sur l'identification des termes contenus dans un corpus. Contrairement à un dépouilleur lexical, qui dresse la liste complète des suites ininterrompues de caractères d'un texte – c'est-à-dire les mots

simples<sup>38</sup> – l’extracteur terminologique a pour tâche de ne sélectionner que les termes. En théorie cela peut sembler relativement facile, cependant dans la pratique il n’en est rien, car départager les termes des autres unités lexicales n’est pas aussi simple qu’il n’y paraît. Au reste, n’oublions pas que même entre experts la notion de « terme » ne fait pas consensus.

Malgré ces difficultés, il est possible, dans une certaine mesure, d’identifier les termes en se basant, entre autres, sur leurs caractéristiques formelles et sur leur fréquence relative. La plupart des systèmes d’extraction de termes ne s’intéressent qu’aux syntagmes nominaux (Wong *et al.* 2008) principalement pour deux raisons. D’une part, les unités terminologiques complexes sont celles qui possèdent les caractéristiques formelles les plus facilement reconnaissables<sup>39</sup>; de l’autre, elles seraient plus fréquentes que les unités simples (Nakagawa and Mori 2002). Bien entendu, les listes de candidats termes (CT) proposées par les extracteurs de termes contiennent du bruit et doivent être nettoyées par un expert ou un terminologue. En outre, il faut tenir compte du fait que ces outils extraient la terminologie propre à des corpus spécialisés dans lesquels il n’est pas rare de voir plusieurs domaines coexister.

De façon classique, l’extraction de termes fait appel à deux approches : linguistique et statistique. Les outils faisant intervenir les deux approches de façon équilibrée donnent en principe les meilleurs résultats. Dans la littérature, ces outils sont souvent nommés *hybrides*. L’ordre dans lequel les stratégies linguistiques et statistiques sont appliquées varie d’un outil à l’autre. Les sections 4.1.1 et 4.1.2 qui suivent donnent un aperçu des principales stratégies employées par ces deux approches.

### 4.1.1 Approches linguistiques

Les approches linguistiques reposent sur l’observation que, dans une langue donnée, les termes complexes se construisent selon des patrons morphosyntaxiques typiques. En français, de nombreux termes sont formés sur les patrons suivants : NOM + prép/dét +

---

<sup>38</sup> L’extraction lexicale en corpus monolingue ne fait pas l’objet d’une description dans ce chapitre puisque cette technique repose sur des stratégies relativement simples à gérer, comme la reconnaissance des frontières entre chaînes de caractères, du moins pour les langues non agglutinantes.

<sup>39</sup> Premièrement, il a été observé que les termes complexes empruntent des patrons morphosyntaxiques typiques (*cf.* 4.1.1). Deuxièmement, les termes complexes apparaissent généralement plusieurs fois dans un texte (*cf.* 4.1.2).

NOM (*dioxyde de carbone, absorption de la chaleur*), NOM + ADJ (*énergie fossile, calotte glaciaire*). Les termes français adoptent également des patrons moins fréquents : NOM + dét + éponyme (*force de Coriolis, cycle de Calvin*), etc. En anglais, les patrons se présentent généralement sous la forme de NOM<sub>2</sub> + NOM<sub>1</sub> (*carbon dioxide, heat absorption, fossil energy, ice cap*), ADJ + NOM (*algal bloom, interglacial period*). Les termes anglais peuvent également adopter des patrons moins courants : éponyme + NOM (*Coriolis force, Calvin cycle*), etc. Les patrons sont plus ou moins complexes selon les langues. Par exemple, en français, les termes sont plus souvent construits à l'aide de prépositions et de déterminants qu'en anglais. Par ailleurs, pour aller chercher des termes plus longs, les patrons de base peuvent être combinés, par exemple, NOM + prép/dét + NOM + prép/dét + NOM (*gaz à effet de serre, réduction des émissions de carbone*) ou encore NOM + ADV + ADJ (*organisme génétiquement modifié*). En règle générale, les extracteurs de termes ayant recours à des patrons complexes produiront des listes ayant un taux de rappel élevé<sup>40</sup>, par contre leur taux de précision sera bas<sup>41</sup>. Inversement, les outils n'utilisant que quelques patrons présenteront des listes dont le taux de précision est élevé, mais ayant un taux de rappel bas.

Plutôt que de rechercher des patrons morphosyntaxiques, une stratégie inverse peut être adoptée. Elle consiste à repérer les marqueurs de frontière entre les groupes nominaux, comme les verbes conjugués, les conjonctions, les prépositions + adjectifs possessifs, les pronoms, etc. L'hypothèse sous-jacente est que ces unités sont peu susceptibles de figurer dans les termes (Bourigault 1993).

Les approches linguistiques nécessitent le prétraitement du corpus. Celui-ci doit être étiqueté et lemmatisé, cette dernière étape permettant de regrouper les formes fléchies. Divers filtres sont ensuite appliqués pour affiner les résultats. Les principaux désavantages des approches linguistiques sont la complexité et la lourdeur du traitement et la faible portabilité des systèmes d'une langue à l'autre. De plus, elles ne permettent pas l'extraction des termes simples. Par contre, elles présentent l'avantage de pouvoir s'appliquer sur de petits corpus.

---

<sup>40</sup> Le rappel est égal au nombre de propositions correctes identifiées par un extracteur par rapport à l'ensemble des termes contenus dans le corpus.

<sup>41</sup> La précision est égale au nombre de propositions correctes identifiées par un extracteur par rapport à l'ensemble proposé par ce même extracteur.

### 4.1.2 Approches statistiques

Pour extraire les unités terminologiques, les approches statistiques peuvent mettre en œuvre plusieurs stratégies. Une des plus simples consiste à extraire les segments répétés (Choueka *et al.* 1983). Si des segments composés de deux chaînes de caractères ou plus apparaissent plus de deux fois à l'intérieur d'un corpus, ils sont identifiés comme des termes potentiels. Cette approche repose sur l'hypothèse que les termes apparaissent plus d'une fois dans les textes spécialisés et qu'ils sont plus fréquemment complexes. Une deuxième stratégie consiste à mesurer la force d'association entre deux occurrences. Si un mot apparaît plus souvent avec un autre mot (ou plusieurs) qu'isolément, alors l'ensemble est susceptible d'être un terme. Par exemple, dans notre corpus du changement climatique, le mot *puits* apparaît plus souvent avec *carbone* que seul, par conséquent *puits de carbone* sera considéré comme un terme. Les approches statistiques peuvent également calculer la fréquence relative d'une chaîne de caractères par rapport à un corpus de référence, généralement de très grande taille et en langue générale. Cette dernière technique offre l'avantage d'extraire des termes simples.

Utilisées seules, les stratégies statistiques n'exigent généralement pas le traitement linguistique du corpus. Par contre ils donnent de meilleurs résultats sur des corpus de très grande taille. Leur portabilité est leur principal atout.

## 4.2 Extraction des équivalents en corpus parallèles

Dans cette section, nous décrivons brièvement les techniques d'extraction bilingue lexicale et terminologique en corpus parallèle. Sur ce type de corpus, l'extraction de lexiques bilingues (extraction de mots simples et alignement des équivalents) donne de meilleurs résultats que l'extraction de terminologies bilingues. Dans le premier cas, les outils n'ont pas pour tâche de distinguer un terme d'un mot de la langue courante, comme ont à le faire les outils qui extraient les termes. Par ailleurs, les extracteurs de termes s'intéressent avant tout aux syntagmes nominaux, ce qui complexifie la tâche, comme expliqué au point 4.2.2. Malgré tout, quelques outils d'extraction de termes sont désormais suffisamment

performants pour être commercialisés (Déjean et Gaussier 2002 : 1), comme Synchronterm<sup>42</sup>, ou encore la suite ITools (Foo et Merkel 2010).

En corpus parallèles, que ce soit pour le dépouillement lexical ou pour l'extraction terminologique, la première étape consiste généralement à aligner le texte au niveau de la phrase. Puis, dans la deuxième étape, on procède à l'alignement des correspondances, autrement dit des équivalents. La section 4.2.1 présente brièvement les méthodes d'alignement au niveau de la phrase, et la section 4.2.2 aborde l'extraction des équivalents au niveau des d'unités lexicales ou terminologiques.

### **4.2.1 Alignement de textes parallèles au niveau de la phrase**

L'alignement au niveau des phrases permet de réduire le champ de recherche des équivalents. Au lieu de chercher l'équivalent d'un mot source dans tout le texte cible, on le repère dans le segment qui lui correspond.

Pour aligner les textes parallèles, il ne suffit pas de s'appuyer uniquement sur les frontières naturelles de la phrase : majuscules en début de phrase et ponctuations de fin de phrase (. | ? | ! | ¶ | etc.), car il arrive souvent qu'une phrase en langue source se traduise par plusieurs phrases en langue cible (ou inversement). Afin d'optimiser l'alignement, deux méthodes de base utilisant des informations uniquement internes aux textes sont employées. La première, mise au point par Kay et Röscheisen (1988), exploite les chaînes de caractères qui sont identiques dans les deux textes (noms propres, dates, etc.). Dans un premier temps, on extrait ces repères, ou points d'ancrage, en se basant sur leur similarité de distribution à l'intérieur de zones (par exemple, le début et la fin du texte) qui se correspondent dans les deux textes. Puis, dans un deuxième temps, à l'aide de ces points d'ancrage, on aligne les textes au niveau des phrases. La deuxième méthode repose sur la longueur relative des phrases. Par exemple, les phrases françaises sont en principe plus longues que les phrases anglaises; de plus, cette caractéristique est généralement régulière. La longueur des phrases peut être calculée en fonction du nombre de caractères contenus dans une phrase (Gale et Church 1991) ou en fonction du nombre de mots présents dans une phrase (Brown *et al.* 1991). D'autre part, les deux méthodes de base peuvent être

---

<sup>42</sup> Adresse Internet de Synchronterm : <http://www.terminotix.com/?lang=fr>.



combinées et affinées par la prise en compte des cognats (Church 1993; Kraif 1999; Langlais 1997; Simard *et al.* 1992).

Les techniques d'alignement dont on vient de parler donnent de très bons résultats pour les langues comme l'anglais et le français, ils se situent aux alentours de 98,5 % (Véronis 2000 : 10), à condition bien sûr que les textes ne présentent pas de différences structurelles (paragraphe manquant ou inversé, etc.). Cependant certaines paires de langues sont plus difficiles à aligner. Par exemple, « le système d'écriture du japonais ne dispose pas de séparateur graphique indiquant les frontières entre les mots » (Nakamura-Delloye 2005). Par conséquent, pour ces langues, il est plus difficile de cerner les points d'ancrage qui permettent l'alignement optimal des textes parallèles. Pour pallier ces difficultés, les chercheurs ont dû recourir à des ressources externes en intégrant des analyseurs morphologiques et des dictionnaires bilingues dans les systèmes d'alignement.

#### **4.2.2 Extraction des équivalents en corpus parallèle**

Avant d'aborder les techniques d'alignement à un degré de granularité plus fin que celui de la phrase, il convient de rappeler que pour de nombreuses raisons tous les mots d'un texte source ne peuvent être alignés avec tous les mots d'un texte cible. Pour illustrer quelques-uns des problèmes qui peuvent survenir lors de la mise en relation des mots, nous avons pris dans notre corpus parallèle une paire de phrases (anglais/français) qui, à première vue, semble présenter peu de difficultés (Figure 4.1). Cependant, en observant attentivement les alignements de la Figure 4.1, nous constatons rapidement que le terme anglais *event* n'est pas traduit en français. D'ailleurs, dans notre corpus, il est assez fréquent que le mot anglais *event* soit traduit en français. Ensuite, nous voyons que les équivalents des termes *extreme* et *precipitation* ne suivent pas le même ordre dans la phrase. Enfin, nous remarquons que *is projected* et *to increase* se rendent chacun par un seul mot en français *devrait* et *progresser*.

The	→	La
frequency	→	fréquence
of	→	des
extreme	↘	précipitations
precipitation	↗	extrêmes
events	→	∅
is projected	→	devrait
to increase	→	progresser
almost	→	presque
everywhere	→	partout
.	→	.

Figure 4.1. Exemple de non-correspondance entre mots de textes source et cible

En ce qui touche les deux derniers exemples, *is projected* et *to increase*, nous aurions pu effectuer un découpage différent de la phrase et considérer dans son entier l'expression : *is projected to increase*. Toutefois, le problème n'en serait pas simplifié. Premièrement, le logiciel aurait à effectuer un découpage adéquat dans le texte source. Deuxièmement, il devrait identifier les divers équivalents dans le texte cible, qui peuvent emprunter divers patrons (unité lexicale, syntagme, syntagme avec insertion, etc.) (Tableau 4.1). Sans oublier que, si l'expression source est plus longue qu'initialement prévu, d'autres traductions sont envisageables (dernière ligne du Tableau 4.1). Tous les exemples du Tableau 4.1 sont tirés du corpus parallèle.

Tableau 4.1. *Is projected to increase* et quelques-uns de ses équivalents possibles

Expression anglaise	Patron	Traduction
<i>is projected to increase</i>	1 mot	<i>augmenterait</i>
		<i>passerait (de ... à ...)</i>
	2 mots	<i>pourrait atteindre</i>
		<i>devrait atteindre</i>
		<i>devrait augmenter</i>
		<i>devrait progresser</i>
	+ forme pronominale	<i>devrait passer (de ... à ...)</i>
	4 mots	<i>devrait s'accroître</i>
		<i>on prévoit une hausse</i>
	+ insertion	<i>on prévoit une augmentation</i>
<i>on estime que (les températures) vont augmenter</i>		
<i>pourrait rapidement augmenter</i>		
<i>is projected to increase threefold.</i>	expression plus longue que prévue	<i>devrait légèrement augmenter</i>
		<i>devrait tripler</i>

#### 4.2.2.1 Extraction lexicale en corpus parallèle

Les méthodes d'extraction de lexiques bilingues se divisent en deux grandes familles : celle à dominante statistique et celle qui combinent des calculs statistiques simples et des indices linguistiques (Ozdowska et Claveau 2005; Tiedemann 2003).

Les méthodes statistiques qui s'appuient notamment sur les modèles IBM (Brown *et al.* 1993) sont généralement utilisées en traduction automatique. Les modèles de traduction mot à mot d'IBM, au nombre de 5, implémentent des techniques probabilistes. Ces modèles de complexité croissante, mais présentant des caractéristiques communes, nécessitent des corpus de très grande taille alignés phrase à phrase. Ils peuvent tenir compte, entre autres, de l'ordre des mots dans la phrase (*distortion*) et de l'alignement de terme simple à terme complexe (*fertility*). Les méthodes IBM obtiennent de bons résultats, mais elles sont difficiles à réaliser. Comme ces méthodes concernent principalement la traduction automatique, nous ne les décrivons pas plus avant dans cette thèse<sup>43</sup>.

<sup>43</sup> On trouvera une bonne description des modèles IBM, entre autres, dans Déleger (2009) et Ozdowska (2006).

Les méthodes mettant en œuvre des calculs statistiques simples et des indices linguistiques sont particulièrement utiles pour extraire des lexiques bilingues et sont mieux adaptées que les méthodes statistiques (du type IBM) pour traiter les petits corpus. Ces méthodes nous intéressent sur deux plans. Premièrement, de nombreuses stratégies employées en extraction lexicale sont également utilisées en alignement bilingue de termes. Deuxièmement, les terminologues peuvent tirer avantage des extracteurs lexicaux pour extraire les candidats équivalents des termes simples<sup>44</sup>.

En règle générale, les méthodes utilisant des calculs statistiques simples s'appuient sur la comparaison de la distribution d'un mot source à celle d'un mot cible, à partir de textes alignés. L'idée sous-jacente étant que si deux mots apparaissent dans les mêmes paires de phrases plus souvent que par le seul fait du hasard, alors ces deux mots sont probablement des équivalents. La comparaison de la distribution des mots est effectuée à l'aide de différentes mesures : l'information mutuelle (Church et Hanks 1990), le *t-score* (Fung et Church 1994); le rapport de vraisemblance (*log-likelihood ratio*) (Dunning 1993) et la probabilité de l'hypothèse nulle (Kraif 2000), etc.

En général, à ces calculs statistiques simples, des indices linguistiques sont ajoutés pour améliorer l'alignement au niveau des mots, par exemple, la cognation et la position des mots dans la phrase.

Comme nous l'avons vu à la section 2.3.3, les cognats sont des mots de langues différentes ayant une ressemblance graphique et dont les sens sont identiques, par exemple, *atmospheric* (anglais) et *atmosphérique* (français). Les cognats ne doivent pas être confondus avec les faux amis qui sont des mots de langues différentes ayant une ressemblance graphique, mais dont les sens sont différents, par exemple, le mot anglais *opportunity* signifie 'circonstance à propos, occasion' et non *opportunité* 'caractère de ce qui convient de faire' (cf. 3.3.1). Dans les langues apparentées comme le français et l'anglais les cognats et les faux amis sont particulièrement nombreux. Cependant, on peut présumer que, dans une même paire de phrases, les mots présentant une ressemblance graphique seront, sauf exception, des cognats.

La position des mots dans la phrase est un autre indice qui permet d'améliorer l'alignement. Pour certaines langues, comme l'anglais et le français, la position des mots

---

<sup>44</sup> C'est ce type d'outils que nous utilisons dans cette thèse.

n'est pas libre dans la phrase. Par ailleurs, leur position dans la phrase n'est pas toujours la même d'une langue à l'autre. À partir de règles qui régissent la position des mots, il est possible d'affiner les alignements (Ahrenberg *et al.* 1998).

L'extraction de lexiques bilingues se fait relativement facilement et donne de bons résultats. Par exemple, un outil comme Alinea (Kraif 2001) peut proposer plusieurs équivalents<sup>45</sup>. D'après des analyses effectuées sur des listes produites par Alinea (Le Serrec 2008; Le Serrec *et al.* 2010), le premier équivalent est pratiquement toujours valide<sup>46</sup> et les suivants le sont souvent. Seules les propositions dont la fréquence est très basse peuvent ne pas être un équivalent (en général autour de 5 occurrences).

#### 4.2.2.2 Extraction de termes en corpus parallèle

Les extracteurs de termes bilingues doivent surmonter deux difficultés. Premièrement, il faut qu'ils identifient les termes spécifiques au corpus étudié. Deuxièmement, ils doivent appairer des termes complexes. Comme dans le cas de l'extraction de lexiques bilingues, la performance de l'extraction bilingue de termes dépend en premier lieu de la qualité de l'alignement des textes (Chiao 2004 : 33). Diverses stratégies utilisées en extraction de termes sur corpus monolingues et en extraction bilingue de lexiques sont également utilisées en extraction bilingue de termes (cooccurrence parallèle, cognation et position des mots dans la phrase, patron morphosyntaxique, etc.).

Il existe principalement trois approches en extraction bilingue de termes (Chiao 2004; Gaussier 2001; Ozdowska et Bourigault 2004). La première approche, la plus classique, repère dans un premier temps les termes dans le texte source et dans le texte cible. Puis, dans un deuxième temps, elle effectue l'appariement des termes (Daille *et al.* 1994; Gurrutxaga *et al.* 2006; Van der Eijk 1993). La deuxième approche préconise l'extraction des termes dans une seule langue suivie de l'identification des termes équivalents dans l'autre langue (Dagan et Church 1994; Gaussier 1998; Hull 2001). Enfin, dans la troisième approche, on commence par effectuer une analyse syntaxique parallèle des textes sources et cibles, puis on procède à ce que les auteurs nomment *identification*

---

<sup>45</sup> Si un mot source possède plusieurs équivalents, Alinea les présente par ordre décroissant de pertinence.

<sup>46</sup> Sauf dans les cas où le terme source et un mot simple et l'équivalent un syntagme. Par exemple, comme il ne peut pas aligner l'unité lexicale anglaise *potato* à la locution française *pomme de terre*, il choisira soit *pomme*, soit *terre* – le déterminant *de* faisant partie d'une liste d'exclusion.

*simultanée des termes* (Névéol et Ozdowska 2005; Ozdowska et Bourigault 2004; Wu 2000).

Le plus souvent, les candidats termes sont repérés au moyen de patrons morphosyntaxiques du type *Nom + adjectif* (Dagan et Church 1994; Daille *et al.* 1994; Gaussier 1998; Gurrutxaga *et al.* 2006;). Plus rarement, ils sont extraits grâce à la méthode de la reconnaissance des segments répétés (Hull 1998). L'alignement des candidats termes aux équivalents se fait à l'aide de calculs statistiques simples mentionnés dans la section 4.2.2.1. Pour affiner les résultats, plusieurs filtres linguistiques peuvent intervenir (cognation, position des mots dans la phrase, liste d'exclusion<sup>47</sup>, etc.).

Étant donné que notre étude ne porte que sur les termes simples, nous n'entrerons pas plus avant dans la description des extracteurs de termes sur corpus parallèles.

### 4.3 Extraction des équivalents en corpus comparable

Malgré des résultats encourageants, l'extraction des équivalents en corpus comparable reste encore du domaine de la recherche (Chiao 2004 : 36; Prochasson 2009 : 105). L'identification des équivalents dans ce type de corpus pose des difficultés beaucoup plus grandes que celles rencontrées dans les corpus parallèles :

L'espace de recherche de la traduction d'un mot ne peut pas être réduit dans un corpus comparable, contrairement à un corpus parallèle où l'espace de recherche est réduit aux segments alignés. (Chiao 2004 : 40)

En effet, comme nous l'avons vu à la section 4.2, dans un corpus parallèle, les textes sont généralement alignés phrase à phrase. En outre, on peut s'attendre à ce que le mot du texte source dont on cherche l'équivalent se trouve dans la phrase correspondante du texte cible. Ces deux conditions permettent d'identifier plus facilement les équivalents en s'appuyant sur les méthodes esquissées à la section 4.2.2. Dans les corpus comparables, la situation est complètement différente. Tout d'abord, l'espace de recherche des équivalents d'un mot source<sup>48</sup> correspond à l'entièreté du corpus cible. Ensuite, tous les mots cherchés du texte

---

<sup>47</sup> Les listes d'exclusion contiennent généralement les mots grammaticaux (ex. déterminants, pronoms, etc.)

<sup>48</sup> Comme de nombreux auteurs (Prochasson 2009 : 12; Zweigenbaum et Habert 2006 : 25), nous utilisons les termes *source* et *cible* par commodité. La partie *source* du corpus est celle dans laquelle les termes dont il faut trouver les équivalents sont issus. La partie *cible* du corpus est celle dans laquelle les équivalents sont cherchés.

source n'ont pas forcément un équivalent dans le texte cible. Pour mettre en correspondance les équivalents, il faut par conséquent s'appuyer sur d'autres indices que ceux employés dans les corpus parallèles.

L'approche la plus répandue pour extraire les équivalents repose sur le principe de la sémantique distributionnelle (Harris 1988). Cette approche part de l'hypothèse que des mots de langues différentes qui apparaissent respectivement dans les mêmes environnements lexicaux ont de fortes chances d'être des traductions réciproques (Chiao 2004; Morin 2009). D'un point de vue statistique, le problème peut être présenté de la façon suivante :

If – for example – in a text of one language two words A and B co-occur more often than expected by chance, then in a text of another language those words that are translations of A and B should also co-occur more frequently than expected.  
(Rapp 1999 : 520)

Ainsi, plus souvent deux unités de langues différentes sont employées dans des contextes très voisins (les mots qui entourent le terme source et le terme cible sont très similaires), plus la « liaison sémantique » entre ces deux unités est élevée (Chiao 2004 : 11). Par conséquent, on peut dire que la mise en relation s'établit sur le plan sémantique, « dans la mesure où la distribution des mots est une caractéristique indirecte du sens » (Zweigenbaum et Habert 2006 : 26)<sup>49</sup>.

D'un point de vue général (Figure 4.2), pour exploiter cette approche<sup>50</sup> en corpus comparable, on procède en premier lieu à la construction des vecteurs de contexte<sup>51</sup> de tous les mots dans les deux sous-parties du corpus (1). Par la suite, afin de permettre le passage d'une langue à l'autre, les éléments (les cooccurrents) des vecteurs sources, du moins une partie<sup>52</sup>, sont traduits dans la langue cible à l'aide d'un dictionnaire bilingue (2). Finalement, chaque vecteur de contexte traduit est comparé à tous les vecteurs cibles au

---

<sup>49</sup> Même si certaines techniques vues en extraction lexicale et en extraction de termes sur corpus parallèles sont reprises en extraction sur corpus comparable, cette dernière se distingue par la prise en compte du principe de la sémantique distributionnelle.

<sup>50</sup> Approche que plusieurs nomment : *méthode* ou *approche par traduction directe* (Déjean et Gaussier 2002; Morin 2009; Prochasson 2009).

<sup>51</sup> Un vecteur de contexte caractérise un mot sur la base de ses cooccurrents. Prochasson (2009 : 15) nomme *tête du vecteur* le mot à traduire et *éléments du vecteur*, les cooccurrents du mot à traduire.

<sup>52</sup> Tous les mots ne sont pas présents dans les ressources dictionnaires.

moyen de différentes mesures de similarité ou de distance (3) (Chiao 2004; Chiao et Zweigenbaum 2002; Fung et Yee 1998; Prochasson 2009; Rapp 1999). Le terme cible dont le vecteur est le plus similaire du vecteur source est retenu comme celui le plus susceptible d'être sa traduction.

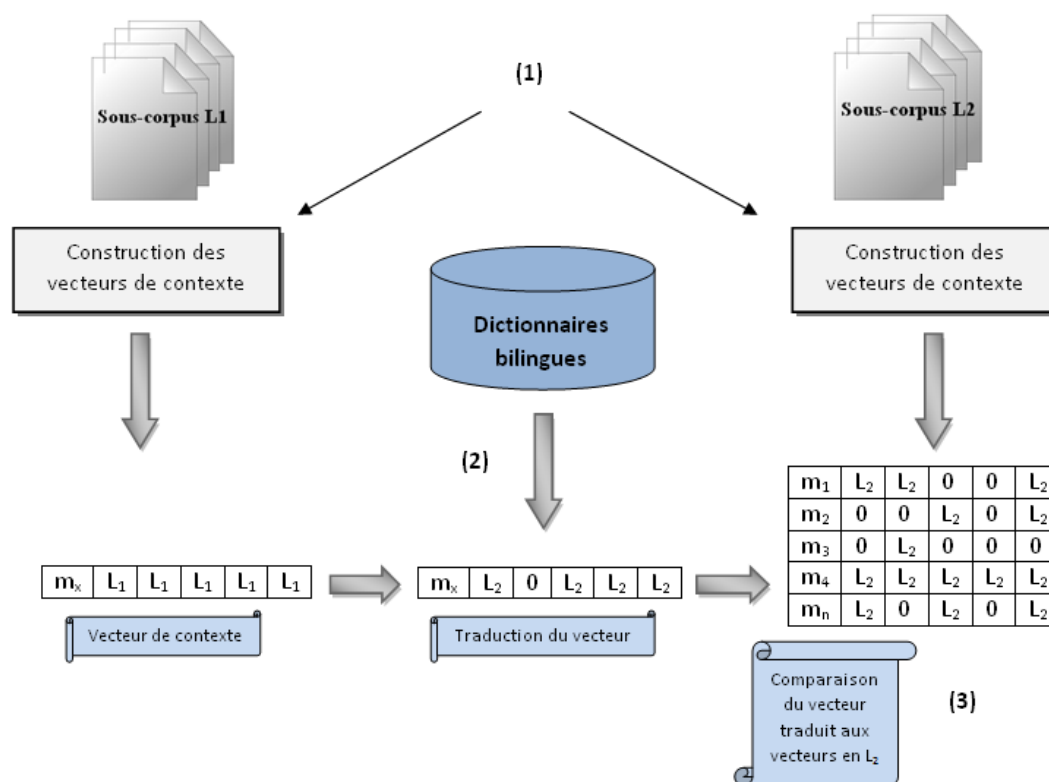


Figure 4.2. Illustration du fonctionnement prototypique de l'extraction sur corpus comparable

Cependant, dans la pratique, l'extraction des équivalents se décompose en de nombreuses étapes, et les choix effectués dans chacune d'elles influencent les résultats de la suivante. Dans les sections qui suivent, nous dressons l'inventaire des différents paramètres utilisés par les chercheurs à chacune de ces étapes. Cet exercice a pour but de montrer la complexité de la tâche et de tenter de découvrir des pistes non explorées.

Toutefois, avant de commencer, nous aimerions souligner qu'en matière d'extraction de lexiques bilingues sur corpus comparable, Fung (1995) et Rapp (1995) font



figure de précurseurs. Dans leurs premiers travaux, ils s'intéressaient à l'environnement des mots sans pour autant utiliser de dictionnaires pivot. Le repérage des équivalents se basait purement sur des calculs statistiques. Malgré des résultats décevants, la recherche en extraction sur corpus comparable était lancée. Par la suite, Fung et McKeown (1997), Fung et Yee (1998) et Rapp (1999) présentent des approches qui utilisent des ressources externes pour faire le lien entre les deux langues, technique que l'on nomme le plus souvent *approche par traduction directe* et que nous décrivons dans les sections qui suivent.

### 4.3.1 Type d'unité identifiée

En analysant les différentes études, on constate que les chercheurs ne s'intéressent pas tous au même type de mot. Certains auteurs cherchent à identifier des unités lexicales simples en vue de constituer un lexique, c'est le cas de Rapp (1999). Fung et McKeown (1997) et Fung et Yee (1998), de leur côté, se penchent sur la traduction des termes simples nouveaux, soit les néologismes, pour enrichir les lexiques existants. D'autres chercheurs, à l'aide d'une liste prédéfinie de termes simples, extraient des équivalents pour construire des terminologies bilingues (Chiao 2004; Chiao et Zweigenbaum 2002; Déjean et Gaussier 2002). D'autres encore, mais plus rarement, s'emploient à extraire à la fois des termes simples et des termes complexes (Laroche et Langlais 2010; Lison 2005; Morin *et al.* 2004), ou seulement des unités complexes (Shahzad *et al.* 1999). Enfin, Léon (2008) s'intéresse uniquement aux unités lexicales complexes ayant pour patron NOM + ADJ et NOM + de + NOM afin de construire une base de données sur les expressions figées.

En résumé, pour la grande majorité, les chercheurs travaillent avec des unités lexicales ou terminologiques simples. En outre, ce sont les unités nominales qui les intéressent plus particulièrement. Il va sans dire que dès que l'on cherche à extraire des unités complexes, les difficultés vont en augmentant. Non seulement en raison de la complexité technique de la chose, mais aussi, comme le souligne Prochasson (2009 : 24), du fait que les termes complexes ont une fréquence moins élevée dans les textes que les mots qui les composent (par exemple, *ozone* et *couche* apparaîtront plus souvent individuellement que *couche d'ozone*). Enfin, que l'on s'intéresse à des mots de la langue générale ou à des termes, la construction du corpus devra tenir compte de ce choix.

### 4.3.2 Type et taille des corpus

Les résultats de l'extraction des équivalents dépendent, bien entendu, du degré de comparabilité des deux parties du corpus. Dans la section sur les corpus comparables (3.2.2), nous avons abordé les points essentiels qui mènent à l'élaboration d'un corpus équilibré. Évidemment, les auteurs construisent des corpus plus ou moins comparables selon la disponibilité des textes ou le type de domaine étudié. Par exemple, le corpus de Déjean et Gaussier (2002) se situe à la limite du corpus parallèle puisque certains textes sont des traductions l'une de l'autre. Les auteurs considèrent leur corpus comme un corpus parallèle bruité, mais ne tiennent pas compte de cette caractéristique au moment de l'extraction lexicale. De leur côté, Haghghi *et al.* (2008) font des expérimentations sur quatre corpus dont le degré de comparabilité varie du plus élevé au moins élevé (corpus parallèle, corpus avec contenu très similaire, corpus constitué de textes appartenant au même domaine, corpus construit avec des textes aux domaines non reliés). Outre la comparabilité, d'autres facteurs influent sur le pourcentage de réussite affiché par les différentes études.

La taille du corpus est un facteur déterminant. Les très gros corpus augmentent, en principe, le pourcentage des traductions correctes, alors que les corpus de petite taille ont tendance à présenter des résultats moins élevés (Prochasson et Morin 2009). Les corpus dans lesquels on extrait des mots de la langue générale sont le plus souvent volumineux. Rapp (1999) construit son lexique à partir d'un corpus de très grande taille, 135 millions de mots pour la partie allemande et 165 millions pour la partie anglaise. Le corpus Web dans lequel Léon (2008) extrait des expressions figées contient environ 4 160 000 pages (Léon 2008 : 149). Pour trouver les néologismes, Fung et McKeown (1997) travaillent aussi sur des corpus de grande taille, mais plus modestes que ceux cités précédemment, 49 millions d'octets<sup>53</sup> pour l'anglais et 127 millions d'octets pour le japonais. Les corpus dans lesquels on extrait des termes sont par nature d'assez petite taille. Cela s'explique par le fait que plus le domaine de spécialité est pointu plus les textes sont rares. Déjean et Gaussier (2002)

---

<sup>53</sup> Évaluer le nombre de mots à partir du nombre d'octets n'est pas facile à effectuer. Toutefois, à titre de comparaison, la partie anglaise de notre corpus parallèle dont la taille est de 6,31 millions d'octets compte environ 1 200 000 mots. Dans le cas de Fung et McKeown (1997), il est encore plus difficile d'évaluer le nombre de mots, puisque les caractères japonais demandent plus d'un octet pour être représentés.

ont utilisé des corpus de très petite taille, 100 000 mots. Le corpus médical de Chiao et Zweigenbaum (2002) contient environ 600 000 mots. Celui de Morin *et al.* (2004), consacré à la forêt et à son industrie, est, quant à lui, relativement plus important, environ 2 500 000 mots.

Dans les travaux cités précédemment, les deux parties du corpus sont de taille plus ou moins égale. À partir de cette constatation et partant de l'idée que les méthodes d'extraction n'imposent pas une taille équivalente pour la partie source et la partie cible, Morin (2009) construit un corpus déséquilibré dont la partie anglaise est 14 fois plus volumineuse que la partie française – dans le contexte de cette étude, le terme *corpus comparable déséquilibré* signifie qu'une des parties du corpus est d'une taille beaucoup plus grande que l'autre.

Le couple de langue traité influence également les résultats. Les langues rapprochées, par exemple l'anglais et le français, sont plus faciles à traiter que les langues éloignées, comme l'anglais et le chinois. Ajoutons que le sens de détection des équivalents peut jouer un rôle important dans le taux de réussite, Chiao a observé de meilleurs résultats dans la direction du français vers l'anglais qu'inversement (Chiao 2004 : 94). À titre d'exemple, voici le sens de traduction pour les couples de langue sur lesquelles quelques chercheurs cités ont travaillé : du japonais vers l'anglais (Fung et McKeown 1997; Sadat *et al.* 2003); du chinois vers l'anglais (Fung et Yee 1998); de l'anglais vers le chinois (Fung 1998); de l'anglais vers l'allemand (Rapp 1999); de l'allemand vers l'anglais (Déjean et Gaussier 2002; Déjean *et al.* 2005); du français vers l'anglais (Chiao 2004; Chiao et Zweigenbaum 2002; Léon 2008; Morin *et al.* 2004). Enfin, sur un corpus trilingue, du français vers le japonais et de l'anglais vers le japonais (Prochasson 2009; Prochasson et Morin 2009).

La plupart des travaux sur corpus comparable ne mettent en jeu qu'un couple de langue. En vue d'obtenir des résultats supérieurs, Prochasson (2009 : 83) construit un corpus multilingue (anglais, français et japonais) afin d'améliorer la qualité des alignements français/japonais et anglais/japonais en s'appuyant sur une liste d'équivalents anglais/français.

### 4.3.3 Traitement du corpus

Une fois les textes collectés, les corpus comparables sont soumis à un nettoyage et à une normalisation standard : élimination des parties non reliées au domaine, conversion de format, etc. Par la suite, une préparation du corpus est effectuée selon les outils dont on dispose<sup>54</sup> ou selon le niveau de dépendance à l'égard des langues que l'on désire garder. Plus on utilise d'outils linguistiques au niveau du traitement du corpus, plus il est difficile de reproduire l'expérience sur d'autres couples de langues où dans un autre sens de traduction.

Il va de soi que tous les corpus doivent être segmentés au niveau des mots. Pour les langues comme l'anglais et le français, la tâche est relativement simple. Cependant, il n'en va pas de même pour les langues agglutinantes comme le japonais et le turc. D'autre part, on procède souvent à une segmentation au niveau de la phrase, cela permet de ne pas inclure dans les vecteurs de contexte des cooccurrents n'appartenant pas à la phrase dans laquelle le mot apparaît.

La plupart du temps, les corpus sont lemmatisés (Chiao 2004; Chiao et Zweigenbaum 2002; Déjean et Gaussier 2002; Fung et McKeown 1997; Rapp 1999; etc.). Cette étape, qui permet de regrouper les formes fléchies est très importante. En plus de simplifier la recherche des équivalents, elle diminue, entre autres, la complexité des calculs sur les très gros corpus (Rapp 1999). La lemmatisation est plus ou moins précise selon les outils utilisés et les langues traitées. Quand il n'est pas possible d'utiliser un lemmatiseur, on peut avoir recours au « stemming »<sup>55</sup> (Lison 2005).

L'étiquetage<sup>56</sup> des corpus est une pratique courante (Déjean et Gaussier 2002; Léon 2008; Morin 2009; Prochasson 2009; Shahzad *et al.* 1999; etc.). Cette technique améliore les résultats, mais rend dépendant de la langue. Par exemple, Chiao (2004 : 54) décide de ne pas étiqueter son corpus afin de conserver la portabilité de son modèle. À l'instar de la

---

<sup>54</sup> Par exemple, les étiqueteurs morphosyntaxiques n'existent pas pour de nombreuses langues.

<sup>55</sup> Le « stemming », parfois nommé *racinisation* (Chiao 2004 : 102; Zweigenbaum et Habert 2006 : 33) permet de réunir par leur racine les dérivés, par exemple, *climat*, *climatique*, *climatisation*, *climatologie* seront regroupés sous *climat*.

<sup>56</sup> Dans le corpus, chaque mot est identifié par sa partie du discours.

lemmatisation, la précision de l'étiquetage dépend de l'outil utilisé et de la langue traitée, d'ailleurs, comme spécifié auparavant, il n'existe pas d'étiqueteur pour toutes les langues.

Les mots grammaticaux<sup>57</sup> génèrent beaucoup de bruit, car ils sont peu discriminants. De plus, ils augmentent la complexité des calculs. Sur de gros corpus, le temps et la mémoire nécessaires sont des facteurs importants à considérer. Pour éliminer les mots grammaticaux, on procède de deux façons. Si le texte n'est pas étiqueté, on utilise une liste de mots grammaticaux (Chiao 2004; Chiao et Zweigenbaum 2002; Rapp 1999). Si le texte est étiqueté, on ne garde que les mots lexicaux (Déjean et Gaussier 2002).

#### 4.3.4 Liste de mots source à traduire

Pour évaluer les performances des systèmes, une liste de mots source doit être constituée. Cette liste, aussi nommée *lexique de référence* (Morin *et al.* 2004), comporte un nombre plus ou moins grand d'unités lexicales, d'une dizaine d'unités (Shahzad *et al.* (1999) à plusieurs centaines<sup>58</sup> (Léon 2008), voire des milliers (Chiao 2004). Le plus souvent, les auteurs évaluent leur système avec une liste de plus ou moins cent mots (Chiao et Zweigenbaum 2002; Morin 2009; Rapp 1999). Le temps d'analyse des résultats entre en jeu dans le choix du nombre de mots à inclure dans la liste.

En règle générale, pour que l'évaluation soit la plus objective possible, on exclut de cette liste les mots qui se trouvent dans le dictionnaire pivot (*cf.* 4.3.7). D'autre part, la plupart du temps, les auteurs s'assurent que les mots du lexique de référence et leur traduction soient présents dans les parties correspondantes du corpus.

La constitution de ces listes fait appel à différentes méthodes, par exemple, celle de Rapp (1999) a été construite par des spécialistes en psycholinguistique (Russell 1970)<sup>59</sup>. Déjean et Gaussier (2002) réutilisent les mots des requêtes ayant servi à repérer les textes du corpus. Morin *et al.* (2007) extraient automatiquement dans les deux sous-parties du corpus des termes simples et des termes complexes pour construire une liste de référence. Lors de l'élaboration de la liste, un nombre minimum d'occurrences dans le corpus peut

---

<sup>57</sup> Les mots grammaticaux (déterminants, prépositions, conjonctions, etc.) sont généralement opposés aux mots lexicaux (nom, verbe, adjectif).

<sup>58</sup> 887 pour être exact.

<sup>59</sup> Liste constituée à l'origine pour étudier des habitudes verbales.

aussi être appliqué aux unités lexicales faisant partie de la liste de référence (Morin 2009; Morin *et al.* 2007).

#### 4.3.5 Construction des vecteurs de contexte

Au début de la section 4.3, il est indiqué que la théorie de la sémantique distributionnelle est à la base de la méthode par traduction directe – approche la plus répandue, rappelons-le. La construction des vecteurs de contexte consiste donc à extraire les cooccurents des mots à traduire afin de représenter le plus fidèlement possible le profil distributionnel d'un mot. Par conséquent, le choix des nombreux paramètres qui entrent en jeu lors de la construction des vecteurs de contexte de tous les mots à traduire de la langue source et de tous les mots de la langue cible est une étape délicate, car les résultats peuvent différer considérablement.

La taille des contextes doit être considérée, car plus la fenêtre est grande plus les calculs sont longs à exécuter. Une fenêtre contextuelle varie de quelques mots (Rapp 1999), à plusieurs phrases (Déjean et Gaussier 2002), voire au paragraphe (Fung et McKeown 1997). Plusieurs auteurs ont fait varier la taille de la fenêtre, afin d'en évaluer les effets sur les résultats (Chiao 2004 : 81; Fung et McKeown 1997 : 196; Laroche et Langlais 2010; Prochasson 2009 : 71–72), par exemple, Prochasson fait varier la fenêtre de 2 à 25 mots. Il arrive à la conclusion que plus les mots sont fréquents dans le texte<sup>60</sup> moins la fenêtre doit être grande pour produire de bons résultats. Par ailleurs, Prochasson (2009 : 72) souligne que les mots peu fréquents sont quand même « moins bien caractérisés que les mots fréquents, même en utilisant une taille de fenêtre optimale ». Enfin, pour éviter de prendre en considération des mots faisant partie d'une autre phrase, il faut initialement découper le corpus au niveau de la phrase (Chiao 2004 : 55).

En principe, seuls les mots lexicaux sont retenus (nom, verbe, adjectif, adverbe) dans les vecteurs de contexte (Déjean et Gaussier 2002). Par ailleurs, partant de l'idée que l'ordre des mots lexicaux est souvent similaire d'une langue à l'autre, Rapp (1999 : 522) tient compte de cette caractéristique dans ses calculs. La fréquence minimale des cooccurents à inclure dans le vecteur de contexte est également un autre facteur qui est pris

---

<sup>60</sup> « Bien que les meilleurs résultats globaux soient obtenus avec une taille de fenêtre de trois mots avant et après le mot vedette (*cf.* section 4.2), nous constatons que la qualité des résultats varie en fonction de la fréquence des mots à traduire » (Prochasson 2009 : 71).

en considération (Morin 2009; Prochasson 2009 : 15–16) – fréquence minimale plus ou moins importante en fonction de la taille des corpus.

L'étiquetage des corpus permet d'éviter de sélectionner les mots d'un homographe catégoriel. Par exemple, si dans un même texte, nous avons le terme *current* ('déplacement continu d'un fluide dans une direction') et l'adjectif *current* ('actuel'), l'étiquetage permet de sélectionner seulement les occurrences du terme. En revanche, les homographes sémantiques ne peuvent être désambiguïsés. Par exemple, si dans un même texte nous avons les termes *organism* ('être vivant') et *organism* ('association, groupement'), tous les cooccurrents de ces homographes seront relevés sans distinction et feront partie des vecteurs de contexte.

#### 4.3.6 Pondération des cooccurrents

Dans un vecteur de contexte, tous les cooccurrents n'ont pas le même statut. En effet, les cooccurrents qui apparaissent avec le terme à traduire plus souvent que le hasard ne le permet sont plus saillantes que les autres. Cependant, le simple calcul de la fréquence des cooccurrents ne constitue pas un indice fiable (Chiao 2004 : 44). Par conséquent, pour détecter les cooccurrents les plus significatifs, il est nécessaire de recourir à des mesures d'association afin de pondérer la valeur des cooccurrents. Ces mesures permettent d'une part de minimiser l'impact des cooccurrents qui s'associent à de nombreux mots. D'autre part, elles harmonisent le poids des cooccurrents dans les deux sous-ensembles du corpus.

Différentes mesures statistiques peuvent être utilisées pour repérer les cooccurrents significatifs : l'*Information mutuelle*; le *Chi 2*; le *tf.idf*; le *Rapport de vraisemblance* (Log Likelihood Ratio); l'*hypothèse nulle*. Toutefois, les mesures le plus souvent employées sont le *Rapport de vraisemblance* (Déjean et Gaussier 2002; Fung et McKeown 1997; Rapp 1999) ou l'*Information mutuelle* (Morin *et al.* 2004; Rapp 1995). Dans une étude plus récente, Laroche et Langlais (2010) ont utilisé avec succès le *log odds ratio*. Dans leur thèse, Chiao (2004) et Prochasson (2009) présentent certaines de ces mesures dans le détail et discutent des avantages des unes par rapport aux autres. Par exemple, Chiao (2004 : 45) explique que l'*Information mutuelle* favorise les mots de faibles occurrences, et que la mesure du *Chi 2* se montre moins efficace pour les mots ayant peu de contextes.

### 4.3.7 Dictionnaire pivot

Pour comparer les vecteurs source aux vecteurs cible, il est nécessaire d'utiliser des ressources linguistiques bilingues afin de permettre la mise en correspondance des cooccurrents source aux cooccurrents cible. Autrement dit, les ressources linguistiques, ou « dictionnaires pivot », font le pont entre les deux langues. Ainsi, les résultats seront directement influencés par le choix des dictionnaires pivot.

Le type de dictionnaire choisi varie selon la nature des équivalents recherchés, les langues étudiées, les ressources disponibles, etc. Certains emploient seulement des lexiques généraux (Morin 2009; Rapp 1999), alors que d'autres utilisent uniquement des lexiques spécialisés (Chiao et Zweigenbaum 2002). Enfin, plusieurs combinent des lexiques spécialisés à des lexiques généraux (Chiao 2004; Laroche et Langlais 2010). Bien sûr, les lexiques bilingues ne sont pas toujours disponibles pour certains domaines de spécialité, sans oublier qu'ils contiennent de nombreux termes complexes, bien souvent inutilisables pour le type d'extraction pratiqué. Comme on le voit, les lexiques utilisés par les chercheurs varient considérablement par la couverture et la pertinence.

Avant d'être employées, les ressources linguistiques doivent être aménagées. Par exemple, Rapp (1999) élimine les entrées complexes et les 100 mots tests (*cf.* 4.3.4). Chiao (2004) s'assure que les entrées figurent dans les deux sous-parties du corpus. Fung et McKeown (1997) sélectionnent à partir de dictionnaires en ligne les entrées à faible taux de polysémie qui ont une fréquence moyenne dans chaque partie du corpus et écartent les entrées grammaticales.

### 4.3.8 Traduction des vecteurs de contexte

Une fois les vecteurs de contexte construits, les éléments des vecteurs sources sont remplacés par leur traduction en langue cible à l'aide du dictionnaire pivot (Chiao 2004 : 57). À ce stade, trois scénarios peuvent se présenter (Prochasson 2009 : 58) :

1) Aucun des éléments du vecteur source ne peut être traduit. Auquel cas, aucun équivalent pour le mot représenté par le vecteur ne pourra être proposé.



2) Tous les éléments ou certains éléments traduits du vecteur source sont polysémiques ou ont des équivalents multiples. Dans ce cas de figure, le problème peut être géré de différentes manières. Certains auteurs ne prennent en considération que le premier équivalent (Chiao et Zweigenbaum 2002). D'autres tiennent compte de tous les équivalents (Déjean et Gaussier 2002; Morin 2009; Shahzad *et al.* 1999).

3) Tous les éléments traduits du vecteur source trouvent une seule traduction. Dans ce dernier cas, la situation est idéale.

### 4.3.9 Comparaison des vecteurs

À cette étape, les vecteurs de contexte traduits sont comparés aux vecteurs cibles. La comparaison des vecteurs s'effectue grâce à des mesures de similarité ou des mesures de distance. L'hypothèse sous-jacente étant que plus deux mots ont des vecteurs similaires plus ils ont de fortes chances d'être des traductions l'un de l'autre. Le premier type de mesure, comme son nom l'indique, regroupe les vecteurs de contexte qui se ressemblent. Parmi ce type de mesures citons l'*indice de Jaccard*, le *coefficient de Dice*, la *mesure cosinus*. Le deuxième type de mesure, soit les mesures de distance, évalue la différence entre les vecteurs de contexte. Ainsi, lorsque deux vecteurs présentent une faible différence, ou distance, on dira qu'ils se ressemblent (Rosset et Jardino 2008). Les mesures de distance utilisées sont, entre autres, la *distance euclidienne*, la *distance de Manhattan (City-block metric)*, la *distance de Minkowski*. Remarquons au passage que les différentes mesures utilisées sont employées dans d'autres domaines, notamment en recherche documentaire (Information Retrieval).

Dans sa thèse, Prochasson (2009 : 60) explique que ces deux types de mesures sont comparables et qu'il est facile de convertir une mesure de similarité par une mesure de distance et inversement. De leur côté, après démonstration, Rosset et Jardino (2008 : 16–18) reconnaissent qu'il existe une relation entre ces mesures, mais que ce n'est pas une relation inverse. Ici, nous n'entrerons pas plus avant dans ces discussions et laisserons cela aux experts en la matière<sup>61</sup>. Par contre, nous désirons attirer l'attention sur le fait que chacune de ces mesures présente des avantages et des inconvénients : comportement

<sup>61</sup> Les textes, parmi d'autres, de Chiao (2004), de Rosset et Jardino (2008) et de Prochasson (2009) décrivent dans le détail les différentes mesures de distance et de similarité.

différent selon le type de données analysées, niveau de complexité des calculs, etc. Par exemple Rapp (1999) rapporte que, malgré la simplicité de la formule, la *distance de Manhattan* donne de meilleurs résultats que l'*indice de Jaccard*, la *mesure cosinus*, la *distance euclidienne* ou le *scalar product*.

#### 4.3.10 Stratégies pour améliorer l'approche par traduction directe

Pour améliorer l'approche par traduction directe, les auteurs ont introduit différentes stratégies.

##### *Calcul de similarité croisé*

Dans sa thèse, Chiao (2004 : 13) émet l'hypothèse suivante : « Si deux mots sont proches dans une direction de traduction ainsi que dans l'autre (langue A  $\Leftrightarrow$  langue B) alors ils ont de plus fortes chances d'être traductions l'un de l'autre que s'ils ne sont proches que pour une seule direction de traduction ». Partant de cette idée, Chiao propose le *calcul de similarité croisé*, une mesure qui opère une mise en relation des candidats équivalents dans les deux directions. L'auteur observe une nette amélioration des résultats, surtout pour les candidats équivalents occupant la première position. Toutefois, Chiao admet que cette stratégie possède ses limites, surtout avec les mots polysémiques. Dans leurs travaux, Sadat *et al.* (2003) suivent le même raisonnement et présentent un modèle d'extraction bilingue nommé *two-stages translation model*.

##### *Cognats*

Jusqu'ici, nous avons vu que pour trouver la cible d'un mot source, les chercheurs utilisent des informations sur les mots qui l'entourent (comparaison de vecteurs). Il est également possible de s'intéresser aux informations internes des mots. La cognation en est un exemple. Cette stratégie, souvent utilisée dans les techniques d'alignement de textes parallèles (*cf.* 4.2.2), est cependant beaucoup moins employée dans les systèmes d'extraction sur corpus comparables. Cela est dû, d'une part, à l'existence de nombreux faux amis dans les langues apparentées (*opportunity* (anglais)  $\neq$  *opportunité* (français)) et,

d'autre part, au peu de mots ayant une ressemblance graphique dans les langues non apparentées. D'ailleurs, Rapp (1999 : 520) juge cette option carrément inutile dans les corpus comparables et Chiao (2004 : 50) ne s'y attarde que très peu, si ce n'est que pour relever le fait que les dictionnaires pivot contiennent un certain nombre de cognats. Néanmoins, malgré ces critiques, cette approche a séduit quelques chercheurs (Haghighi *et al.* 2008; Ismail et Manandhar 2009; Koehn et Knight 2002; Laroche et Langlais 2010; Léon 2008).

Koehn et Knight (2002) commencent par construire un lexique pivot à l'aide de transfuges<sup>62</sup> extraient directement des corpus à l'étude, cela évite l'utilisation de lexiques externes. Puis, dans une deuxième étape, ils considèrent les cognats. Ismail et Manandhar (2009) s'inspirent de Koehn et Knight (2002), mais constituent un lexique avec des cognats<sup>63</sup> prélevés sur des sites Internet spécialisés. Haghighi *et al.* (2008) disent obtenir de très bons résultats avec les langues apparentées en ajoutant la cognation dans leurs systèmes d'extraction. Léon (2008) repère les cognats dans la troisième et dernière phase de son système d'extraction automatique des unités lexicales complexes. Laroche et Langlais (2010) utilisent la cognation sur les équivalents situés dans le *Top<sub>20</sub>*<sup>64</sup> pour améliorer les résultats. Selon ces derniers auteurs, le domaine médical se prête particulièrement bien à la cognation en raison des nombreux mots possédant des racines grecques ou des racines latines.

### *Translittération*

La translittération, comme la cognation, s'intéresse aux informations internes des mots. Malik *et al.* (2008) la définissent de la façon suivante : « Transliteration is mainly used to transcribe a word written in one language in the writing system of the other language, thereby keeping an approximate phonetic equivalence ». En traduction automatique (TA), de nombreux travaux exploitent la translittération (voir le site du projet REFLEX de

---

<sup>62</sup> Comme Kraif (2001) et Chiao (2004), nous établissons une distinction entre les chaînes de caractères parfaitement identiques (*transfuges*) et les chaînes de caractères partiellement identiques (*cognats*) entre deux langues. De nombreux auteurs ne font cependant pas cette distinction et nomment *cognat* les deux types de chaîne, par exemple Léon (2008 : 76–77).

<sup>63</sup> On comprendra ici que les *cognats* sont des chaînes de caractères parfaitement ou partiellement identiques.

<sup>64</sup> *Top<sub>20</sub>* représente une liste d'équivalents potentiels, le *Top<sub>1</sub>* étant le plus probable.

l'University of Illinois<sup>65</sup>). En extraction sur corpus comparable, quelques auteurs se sont intéressés aux translittérations (Knight and Graehl 1998; Prochasson 2009; Prochasson et Morin 2009; Shao et Ng 2004). Par exemple, Prochasson (2009) les utilise pour enrichir les lexiques pivots et comme points d'ancrage dans les vecteurs de contexte.

### *Filtrage morphosyntaxique des résultats*

Sadat *et al.* (2003), ayant observé que l'équivalent d'un mot donné appartient souvent au même groupe syntaxique que ce dernier, proposent de filtrer les résultats de l'approche directe en fonction de la partie du discours des mots. Cette observation, comme le remarquent judicieusement Zweigenbaum et Habert (2006 : 38), se présente assez systématiquement dans les dictionnaires bilingues, mais moins en corpus. Par exemple, on sait que l'adjectif français *climatique* est pratiquement toujours traduit en anglais par le nom *climate*.

Il va sans dire que pour tenir compte de la partie du discours des mots, le corpus doit être étiqueté. Sadat *et al.* (2003) autorisent quatre types d'associations entre un mot et son équivalent : 1) NOM japonais et NOM anglais; 2) VERBE japonais et VERBE anglais; 3) ADJECTIF japonais et ADJECTIF ou ADVERBE anglais; 4) ADVERBE japonais et ADVERBE ou ADJECTIF anglais (en japonais la plupart des adverbes sont formés à partir d'adjectifs).

### *Calcul de similarité avec les ressources pivot*

L'approche de Déjean et Gaussier (2002 : 4) repose sur l'exploitation maximale des ressources pivot : « Deux mots de L<sub>1</sub> et L<sub>2</sub> sont, avec une forte probabilité, traduction l'un de l'autre si leurs similarités avec les entrées des ressources bilingues disponibles sont proches ». Pour ce faire, les auteurs s'appuient sur les propriétés hiérarchiques d'un thésaurus spécialisé ce qui leur permet de construire des vecteurs de contextes pour les

---

<sup>65</sup> Adresse Internet du projet REFLEX : <http://compling.ai.uiuc.edu/webpage/projects/reflex.htm> (page consultée le 20 mars 2012).

entrées du thésaurus. Les vecteurs du corpus source sont alors comparés aux vecteurs du thésaurus. Par la suite, les vecteurs retenus sont comparés aux vecteurs du corpus cible.

#### *Traduction compositionnelle*

Morin et Daille (2006) et Morin *et al.* (2004) introduisent les termes complexes dans leurs vecteurs de contextes. Toutefois, au moment du transfert d'une langue vers l'autre, de nombreux termes complexes ne peuvent être traduits, car ils ne figurent pas dans les dictionnaires pivot. Pour résoudre le problème, les auteurs proposent d'élargir la couverture des dictionnaires en ayant recours à la traduction compositionnelle. C'est-à-dire, en traduisant séparément chacun des éléments de l'unité à traduire.

#### *Analyse syntaxique*

Considérant que l'extraction des cooccurrents figurant dans une fenêtre de contexte ne constitue qu'un sac de mots, certains chercheurs ont pensé améliorer les résultats par l'ajout d'une analyse syntaxique – stratégie permettant de prendre en considération les relations entre les mots. Les approches sont plus ou moins sophistiquées (Gamallo Otero 2007, 2008; Yu et Tsujii 2009). Cependant, cette approche pose certains problèmes en extraction bilingue, car elle restera silencieuse sur certains phénomènes linguistiques. En effet, il est relativement fréquent qu'un terme appartenant à une partie du discours ait pour équivalent un terme dans une autre partie du discours (Chiao 2004 : 43; Déjean et Gaussier 2002 : 5).

#### *Corpus mieux ciblé*

Les résultats en extraction bilingue sur corpus comparables sont en principe meilleurs avec des corpus de très grande taille (*cf.* 4.3.2). Toutefois, en ce qui concerne les langues de spécialité, il n'est pas possible de constituer de très gros corpus comme en langue générale. D'ailleurs, on se souviendra que plus le domaine est pointu, plus les textes sont rares. Dans cette perspective, afin d'améliorer les résultats, Morin *et al.* (2007) préconisent une construction bien pensée des corpus comparables en appliquant des critères de sélection

plus contraignants. Pour en faire la démonstration, les auteurs construisent un corpus comparable appartenant au sous-domaine du diabète et de l'alimentation qu'ils divisent en deux parties inégales, scientifique et vulgarisé – la partie scientifique étant de plus petite taille. Deux expériences sont menées, la première sur l'ensemble du corpus, scientifique et vulgarisé (français 694 000 mots, japonais 808 000 mots) ; la deuxième seulement sur le sous-corpus scientifique (français 426 000 mots, japonais 234 000 mots). Les meilleurs résultats sont obtenus avec le plus petit corpus, soit celui appartenant au registre scientifique.

#### *Sous-parties du corpus déséquilibrées en taille*

Traditionnellement, les chercheurs construisent des corpus comparables dont les deux sous-parties sont de taille plus ou moins identique. Morin (2009) fait remarquer que les ressources n'étant pas toujours disponibles de façon égale dans les deux langues à l'étude forcent les chercheurs à se limiter à un corpus de plus petite taille. À titre d'exemple, dans certains domaines spécialisés, il est plus facile de trouver des documents en anglais que dans d'autres langues. Or, comme l'explique Morin, les méthodes d'extraction bilingue n'imposent pas une taille équivalente pour la partie source et la partie cible. Partant de cette observation, l'auteur effectue plusieurs expériences sur un corpus dont la partie anglaise est 14 fois plus volumineuse que la partie française (*cf.* 4.3.2) et démontre qu'un corpus comparable déséquilibré judicieusement utilisé permet d'améliorer les résultats.

#### **4.3.11 Validation**

Le mode de validation se fait le plus souvent en comparant automatiquement ou manuellement les résultats à la liste de référence (*cf.* 4.3.4). Les mots à traduire et leurs équivalents sont donc déjà identifiés au départ. En somme, il s'agit de vérifier la performance des systèmes.

Une analyse manuelle est parfois effectuée par des terminologues ou des experts (Léon 2008). Cette façon de procéder semble la plus adéquate, mais elle soulève plusieurs problèmes. Elle exige du temps et des efforts, ce qui n'est pas commode lorsqu'on doit faire un grand nombre d'évaluations. En outre, comme chaque expert possède son propre

point de vue sur ce qui est équivalent ou pas, les résultats peuvent diverger d'un évaluateur à l'autre.

#### 4.3.12 Résultats

Les résultats, on s'en doute, sont extrêmement variables et très difficiles, voire impossibles, à comparer, car comme le souligne Gaussier toutes les étapes du processus d'extraction évoquées ci-dessus entrent en ligne de compte :

Une difficulté supplémentaire est introduite toutefois par le fait que ces facteurs ne sont pas forcément indépendants les uns des autres, et une combinaison qui fonctionne bien avec une méthode donnée peut fort bien se révéler désastreuse pour une autre. (Gaussier 2005 : 50)

À l'ordinaire, les auteurs font différentes expérimentations et ne présentent que les plus prometteuses (Chiao et Zweigenbaum 2002; Déjean et Gaussier 2002; Morin et Daille 2006; Morin *et al.* 2004, 2007; Rapp 1999; Sadat *et al.* 2003). Cependant, certains chercheurs, comme Chiao (2004), Gamallo Otero (2008), Laroche et Langlais (2010), Prochasson (2009), se sont aussi donnés pour tâche de faire varier différents paramètres (fenêtre, mesure d'association, mesure de similarité ou dictionnaire pivot) dans le but de découvrir la meilleure formule.

Une fois les calculs terminés, les candidats équivalents sont ordonnés selon les valeurs obtenues. En général, plusieurs résultats sont présentés. Le premier concerne le pourcentage des équivalents occupant la première position ( $Top_1$ ) dans le classement des candidats. Ensuite, les résultats sont présentés par tranche des 10, 20, 30, etc. meilleurs candidats équivalents ( $Top_{10}$ ,  $Top_{20}$ ,  $Top_{30}$ , etc.). Par exemple, le  $Top_{20}$  signifie que l'équivalent se trouve parmi les 20 premiers équivalents proposés. Les seuils sont proposés empiriquement (Chiao 2004 : 12). Selon Fung (1998 : 15), les candidats équivalents situés dans le  $Top_{20}$  sont généralement reliés au mot source. Dans le Tableau 4.2, nous présentons les résultats obtenus par certains chercheurs cités dans ce chapitre.

Tableau 4.2. Quelques résultats d'extraction sur corpus comparables

Tranche de mots évalués	Pourcentage	Auteurs
<i>Top<sub>1</sub></i>	30	Fung et McKeown 1997
	30	Fung 1998
	20	Chiao et Zweigenbaum 2002
	55,2	Laroche et Langlais 2010
	72	Rapp 1999
<i>Top<sub>10</sub></i>	49	Morin <i>et al.</i> 2007
	84	Déjean et Gaussier 2002
	89	Rapp 1999
<i>Top<sub>20</sub></i>	50,9	Fung et McKeown 1997
	60	Chiao et Zweigenbaum 2002
	76	Fung 1998
<i>Top<sub>40</sub></i>	88	Fung 1998

### 4.3.13 Récapitulatif

De l'état de l'art qui vient d'être présenté, il ressort que ce type d'extraction exige la mise en place minutieuse de ressources linguistiques (corpus, lexiques, etc.) et qu'elle nécessite le recours à différentes mesures statistiques. Il a aussi été observé que les chercheurs ont testé un très grand nombre de stratégies différentes pour améliorer l'extraction d'équivalents à partir de corpus comparables. Par ailleurs, il est à noter que dans l'état actuel des techniques informatiques, des choix douloureux doivent être faits soit parce qu'un outil spécifique n'existe pas pour la langue traitée (ex. analyseur syntaxique), soit parce que les calculs sont longs, voire rédhibitoires (Chiao 2004 : 49). À titre d'exemple, Koehn et Knight (2002) disent avoir limité le nombre de mots à analyser pour réduire le temps des calculs. Ainsi, il apparaît évident que les décisions présidant à la recherche en extraction sur corpus comparable exigent la maîtrise approfondie de deux types de connaissances : *linguistiques et informatiques*.

Le tableau 4.3, présente un récapitulatif des différentes étapes ou stratégies intervenant du début à la fin du processus d'extraction. Les étapes ou stratégies qui selon nous posent le plus de problèmes sont précédées d'un crochet (✓) et sont examinées plus avant ci-après.



Tableau 4.3. Tableau récapitulatif des étapes d'extraction sur corpus comparable

Étape ou stratégie	Description	
<b>Unité identifiée</b>	<ul style="list-style-type: none"> <li>• unité lexicale simple</li> <li>• terme simple</li> <li>• néologisme</li> </ul>	<ul style="list-style-type: none"> <li>• terme simple et terme complexe</li> <li>• terme complexe</li> <li>• terme complexe sur un nombre limité de patrons</li> </ul>
<b>Type et taille des corpus</b>	<ul style="list-style-type: none"> <li>• degré de comparabilité</li> <li>• taille du corpus</li> <li>• corpus déséquilibré (taille différente des sous-corpus)</li> </ul>	<ul style="list-style-type: none"> <li>• couple de langue</li> <li>• sens de détection des équivalents (ex. anglais vers le français)</li> <li>• corpus multilingue (plus de deux langues)</li> </ul>
✓ <b>Traitement du corpus</b>	<ul style="list-style-type: none"> <li>• nettoyage standard</li> <li>• segmentation des mots</li> <li>• segmentation des phrases</li> <li>• lemmatisation</li> </ul>	<ul style="list-style-type: none"> <li>• stemming</li> <li>• étiquetage</li> <li>• liste de mots grammaticaux</li> <li>• mots lexicaux (textes étiquetés)</li> </ul>
<b>Liste de mots à traduire</b>	<ul style="list-style-type: none"> <li>• nombre de mots variant de 25 à des milliers</li> <li>• exclusion des mots se trouvant dans le dictionnaire</li> <li>• nombre d'occurrences minimal</li> </ul>	<ul style="list-style-type: none"> <li>• liste préparée manuellement</li> <li>• extraction de termes</li> <li>• vérification de la présence des mots dans les textes source et de l'équivalent dans les textes cible</li> </ul>
✓ <b>Construction des vecteurs de contexte</b>	<ul style="list-style-type: none"> <li>• taille de la fenêtre de contexte <ul style="list-style-type: none"> <li>○ contextes de n mots avant et après le mot</li> <li>○ plusieurs phrases</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>○ paragraphe</li> <li>• ordre des mots</li> <li>• fréquence minimale des cooccurents</li> </ul>
✓ <b>Pondération des cooccurents</b>	<ul style="list-style-type: none"> <li>• mesures d'association <ul style="list-style-type: none"> <li>○ Information mutuelle</li> <li>○ Chi 2; tf.idf</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>○ rapport de vraisemblance (<i>Log Likelihood Ratio</i>)</li> <li>○ hypothèse nulle</li> <li>○ <i>Log Odds Ratio</i></li> </ul>
✓ <b>Dictionnaire pivot</b>	<ul style="list-style-type: none"> <li>• lexiques bilingues généraux</li> <li>• lexiques bilingues spécialisés</li> <li>• combinaison de lexiques bilingues généraux et spécialisés</li> </ul>	<ul style="list-style-type: none"> <li>• élimination dans les dictionnaires des mots à traduire</li> <li>• sélection de termes peu polysémiques</li> <li>• élimination des entrées grammaticales</li> </ul>
✓ <b>Traduction des vecteurs de contexte</b>	<ul style="list-style-type: none"> <li>• prise en compte du premier équivalent</li> </ul>	<ul style="list-style-type: none"> <li>• prise en compte de tous les équivalents</li> </ul>
✓ <b>Comparaison de vecteurs</b>	<ul style="list-style-type: none"> <li>• mesure de similarité (indice de Jaccard; coefficient de Dice; mesure cosinus)</li> </ul>	<ul style="list-style-type: none"> <li>• mesure de distance (distance euclidienne, distance de Manhattan ou <i>City-block Metric</i>, distance de Minkowski)</li> </ul>

✓ <b>Stratégies</b>	<ul style="list-style-type: none"> <li>calcul de similarité croisée <math>\approx</math> <i>two-stages translation model</i></li> <li>prise en compte des transfuges</li> <li>prise en compte des cognats</li> <li>prise en compte des translittérations</li> <li>filtrage morphologique des candidats</li> </ul>	<ul style="list-style-type: none"> <li>calcul de similarité avec les ressources pivot</li> <li>traduction compositionnelle</li> <li>analyse syntaxique</li> <li>corpus mieux ciblé</li> </ul>
✓ <b>Validation</b>	<ul style="list-style-type: none"> <li>automatique</li> <li>manuelle</li> </ul>	<ul style="list-style-type: none"> <li>par des experts du domaine, des terminologues</li> </ul>
<b>Présentation des résultats</b>	<ul style="list-style-type: none"> <li>ordonnés selon les valeurs obtenues</li> <li><i>top1</i></li> </ul>	<ul style="list-style-type: none"> <li><i>top10, top20, top30, etc.</i></li> </ul>

En vue de mieux cerner les étapes ou les stratégies<sup>66</sup> posant des difficultés en extraction bilingue, nous les avons regroupées sous trois catégories : 1) paramètres statistiques; 2) stratégies ou paramètres linguistiques; 3) validation des candidats équivalents. Ce regroupement nous permet de mieux identifier les points sur lesquels notre étude pourrait contribuer à l'avancement de la recherche.

### 1) Paramètres statistiques

Nous avons vu que les paramètres statistiques utilisés peuvent interagir entre eux de façon inattendue : nous pensons aux décisions prises lors de la *construction des vecteurs de contexte*, de la *pondération des cooccurents*, de la *traduction des vecteurs de contexte* et de la *comparaison de vecteurs*, etc. (Gaussier 2005). D'autre part, étant donné que ces paramètres sont nombreux, les combinaisons possibles le sont d'autant plus. Les tester toutes serait pratiquement ingérable et exigerait beaucoup de temps. Sans oublier qu'une combinaison de paramètres efficaces pour un couple de langues ou un genre de corpus donné pourrait ne pas l'être avec un autre. Souvent, pour augmenter la portabilité de leurs systèmes, les concepteurs intègrent le moins de paramètres linguistiques possible. Malgré cela, la portabilité des systèmes n'est toujours pas suffisante et les résultats sont moins bons.

<sup>66</sup> Étapes ou stratégies identifiées par un crochet dans le Tableau 4.3.

Ainsi, la découverte de la meilleure combinaison de paramètres statistiques reste encore un défi de taille. Dans cette étude, il est évident que nous ne sommes pas en mesure d'apporter des pistes de solution à ce problème. Néanmoins, nous ne devons pas le perdre de vue lorsque nous étudions l'équivalence de l'anglais vers le français en corpus comparable afin de ne pas trop généraliser nos conclusions.

## 2) *Stratégies ou paramètres linguistiques*

Certains chercheurs choisissent de faire appel à certains paramètres linguistiques et d'autres de ne pas y recourir. En l'occurrence, à l'étape du *traitement du corpus*, il arrive qu'un concepteur n'étiquète pas le corpus, soit parce qu'il n'existe pas d'étiqueteur pour la langue traitée, soit parce qu'il désire conserver la portabilité de son système. Toutefois, l'étiquetage, lorsqu'il est effectué avec un outil performant, permet la désambiguïsation des homographes catégoriels, que ce soit dans la langue source ou dans la langue cible, par exemple :

- *current*-NOM 'courant marin', *current*-ADJ 'actuel';
- *change*-NOM 'changement', *change*-VER 'changer';
- *récolte*-NOM 'produits recueillis', *récolte* de *récolter*-VER 'faire la récolte';
- *littoral*-NOM 'zone', *littoral*-ADJ 'relatif à cette zone'.

Ainsi, à l'étape de la construction du vecteur de contexte d'un terme (ex. *current*-NOM), l'étiquetage permettrait de réduire le bruit, puisqu'on n'inclurait pas les cooccurrents du mot n'ayant pas la même partie du discours (ex. *current*-ADJ).

À l'inverse de l'exemple précédent, l'usage de la *cognation* risque d'augmenter le silence. Si pour les corpus parallèles, elle est utile (voir section 4.2), elle peut avoir un effet négatif sur les corpus comparables. Selon nous, on devrait recourir à cette stratégie avec beaucoup de prudence et seulement à certaines étapes du traitement. Par exemple, parmi les équivalents des termes anglais *mitigation*, *permafrost*, *latitudinal* on trouve respectivement comme équivalents les cognats *mitigation*, *permafrost*, *latitudinal*, cependant ces termes ne figurent pas au premier rang des équivalents. Par conséquent, leur emploi systématique risque de masquer des équivalents intéressants.

Les *dictionnaires pivot*, quant à eux, posent plusieurs difficultés. Premièrement, il est difficile de disposer de dictionnaires ayant une nomenclature suffisamment exhaustive. Deuxièmement, ils ne proposent que peu d'équivalents par terme. Troisièmement, ils possèdent des entrées polysémiques, comme :

- *wood*-NOM ('matière ligneuse' ou 'ensemble d'arbres');
- *bois*-NOM ('matière ligneuse' ou 'groupe d'arbres');
- *plant*-NOM ('végétal' ou 'usine');
- *réfléchir*-VER ('renvoyer par réflexion' ou 'penser, cogiter').

La polysémie est un problème décrit par de nombreux chercheurs. Dans le but de la limiter, certains se restreignent à des textes appartenant à un registre hautement scientifique, d'autres se limitent à extraire des termes ultra-spécialisés<sup>67</sup>, ou encore éliminent les entrées polysémiques des dictionnaires. Ces différentes stratégies améliorent peut-être les résultats, mais selon nous elles sont trop contraignantes. Dans les deux premiers cas, il n'est pas raisonnable de ne s'intéresser qu'à certains types de corpus ou de termes. Dans le troisième cas, il faudrait savoir à l'avance lesquelles des entrées polysémiques éliminer.

Le filtrage morphologique des équivalents impose également une contrainte qui masque des résultats intéressants. En effet, selon le contexte d'utilisation, certains termes peuvent avoir des équivalents appartenant à différentes parties du discours. Par exemple, le terme anglais *air*-NOM se rend, selon les circonstances, soit par *atmosphère*-NOM soit par *atmosphérique*-ADJ (*air pollution* => *pollution de l'atmosphère* ou *pollution atmosphérique*)

Notre étude comparative de l'équivalence de l'anglais vers le français devrait apporter des pistes de solution pour certains des problèmes qui viennent d'être abordés. Nous pensons notamment montrer l'importance de l'étiquetage, la prise en compte des problèmes de la cognation, de la polysémie des termes et du filtrage morphologique.

---

<sup>67</sup> L'idée derrière cette stratégie est que les termes ultra spécialisés sont le plus souvent monosémiques.

D'autre part, nous pensons que les observations que nous ferons peuvent être étendues aux langues qui partagent les mêmes caractéristiques.

### 3) *Validation des candidats équivalents*

L'étape de la *validation des candidats équivalents* est cruciale, car la présentation des résultats permet de connaître le degré d'efficacité d'un système d'extraction et de le situer par rapport à ses concurrents. Toutefois, il nous semble que cette étape pose de nombreux problèmes. Tout d'abord, évaluer les différents résultats obtenus à l'aide de différentes combinaisons de paramètres linguistiques ou statistiques est complexe. Deuxièmement, le recours à des dictionnaires ou à des listes d'équivalents préparées à l'avance pour valider les candidats impose des résultats qui ne permettent pas de voir le contenu réel des corpus, à plus forte raison, lorsque la validation est effectuée automatiquement. Enfin, nous pensons que l'obstacle principal à une bonne évaluation réside dans le manque de connaissance approfondie de l'équivalence en corpus.

L'analyse comparative de l'équivalence de cette thèse jettera un éclairage plus complet sur l'équivalence, ce qui permettra d'apporter des pistes de solution à l'extraction bilingue en corpus comparable.

## 4.4 Conclusion et contribution à la recherche

Dans ce chapitre, nous avons tout d'abord décrit les principales caractéristiques des extracteurs de termes en corpus monolingue (4.1). Nous avons vu que ces outils ont atteint un degré de maturité suffisant pour être exploitables commercialement. Dans cette thèse, nous utilisons TermoStat pour extraire les termes simples (nom, adjectif, verbe, adverbe) du corpus anglais. Cet outil, défini comme un extracteur hybride, exploite une méthode de mise en opposition de corpus de référence à un corpus spécialisé. Nous le décrivons plus en détail dans le Chapitre 5 (*cf.* 5.2).

Ensuite, nous avons donné les principales caractéristiques des techniques d'alignement et d'extraction lexicale et terminologique en corpus parallèle (4.2). Nous avons dit que les extracteurs de lexiques bilingues sont plus performants que les extracteurs

de terminologie bilingues, mais que ces derniers commencent à faire leur apparition sur le marché. Comme nous nous intéressons aux termes simples, nous avons choisi un extracteur de lexique bilingue, Alinea, pour effectuer l'extraction des candidats équivalents français de nos termes anglais. Ce logiciel est décrit dans le Chapitre 5 (*cf.* 5.15.1).

Pour terminer, nous avons examiné dans le détail les techniques d'extraction bilingue en corpus comparable (4.3). Grâce à cet examen, dans notre méthodologie, nous pourrions nous inspirer de certaines stratégies pour repérer les équivalents dans notre corpus comparable. Parmi toutes les stratégies ou les paramètres analysés, nous en avons identifié un certain nombre qui posent problème et nous les avons divisés en trois catégories : 1) paramètres statistiques; 2) stratégies ou paramètres linguistiques 3) validation des candidats. Nous pensons qu'une étude comparative de l'équivalence en corpus parallèle et en corpus comparable permettra de combler certaines faiblesses ou lacunes constatées dans la deuxième et la troisième catégorie des problèmes discutés à la section 4.3.13.

## **5 MÉTHODOLOGIE COMMUNE AUX DEUX VOLETS DE LA RECHERCHE**

Ce chapitre présente la méthodologie commune aux deux volets de la recherche. En effet, si chacun des deux volets de la recherche adopte une méthodologie propre, ils prennent néanmoins comme point de départ les résultats de la méthodologie commune. Ici, il est utile de rappeler que puisque nous observons l'équivalence de termes simples appartenant à plusieurs parties du discours dans des corpus, nos choix méthodologiques s'inspirent de la terminologie textuelle et de l'approche lexico-sémantique.

Le chapitre est divisé en sept sections. La première décrit en détail les étapes de la mise en forme du corpus parallèle et du corpus comparable. En outre, elle expose le prétraitement des textes sélectionnés et donne les résultats de l'alignement du corpus parallèle (5.1). La deuxième section présente le mode d'extraction automatique des termes anglais et fournit les résultats de cette extraction (5.2). La troisième section explique le nettoyage de la liste d'extraction des termes anglais et leur validation (5.3). La quatrième section décrit la collecte des cooccurrents des termes anglais (5.4). La cinquième section expose la méthode de repérage des équivalents français dans le corpus parallèle (5.5). La sixième section présente la méthode de repérage des équivalents français dans le corpus comparable (5.6). Enfin, la septième section, résume le chapitre (5.7).

### **5.1 Constitution des corpus**

Pour construire les deux types de corpus, nous nous sommes appuyée notamment sur les travaux de Bodson (2005 : 13–33), de Bowker et Pearson (2002 : 12–13, 50–51), de Condamines (2003 : 30–32) et de L'Homme (2004 : 126–127). Par ailleurs, de façon plus spécifique, pour sélectionner les textes du corpus comparable, nous nous sommes basée sur les articles de Maia (2003 : 27–31) et de Morin (2009). De plus, étant donné que notre étude porte sur l'analyse comparative de l'équivalence de termes français en corpus parallèle et en corpus comparable, nous nous sommes tout particulièrement assurée que les

textes français du corpus parallèle soient des traductions et les textes français du corpus comparable, des originaux.

Pour maximiser le degré de similitude entre le corpus parallèle et le corpus comparable, la constitution d'un seul sous-corpus anglais est une des premières stratégies utilisées. De cette manière, le sous-corpus anglais joue le rôle de relais, puisqu'il appartient à la fois au corpus parallèle et au corpus comparable (Figure 5.1). Par conséquent, on peut supposer que le lien qu'il établit entre les deux sous-corpus français augmente leur degré de similitude. Ainsi, les deux parties françaises peuvent être qualifiées de corpus comparable monolingue.

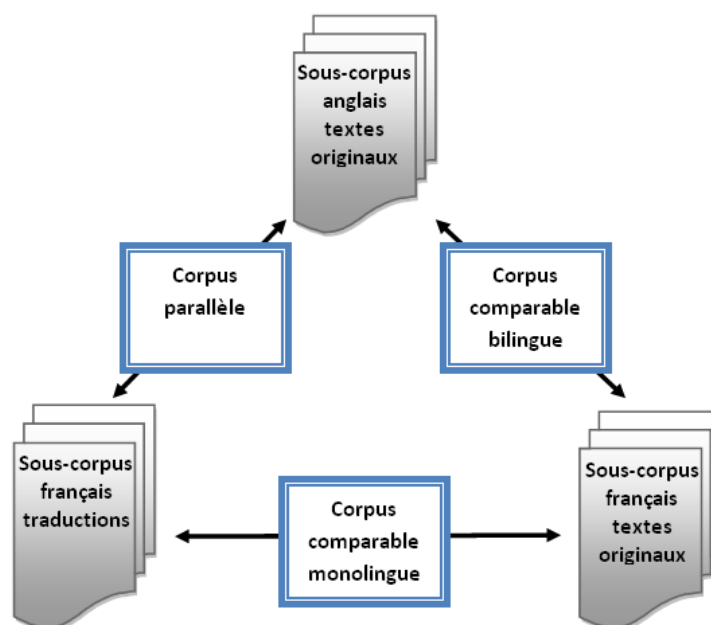


Figure 5.1. Corpus parallèle et corpus comparables

### 5.1.1 Critères de sélection des textes

Nous avons choisi de constituer notre corpus avec des textes relevant du domaine du changement climatique. Deux raisons principales nous ont poussée à choisir ce sujet d'actualité. Premièrement, les textes de qualité qui s'y rapportent sont largement diffusés



sur Internet et librement accessibles. Deuxièmement, nous avons déjà étudié ce sujet dans un précédent travail de recherche (Le Serrec 2008). Les connaissances du domaine que nous avons acquises lors de cette étude nous avantagent, tout particulièrement à l'étape de l'analyse des équivalents.

En nous appuyant notamment sur les travaux cités au point 5.1, nous avons établi six critères de base pour sélectionner les textes du corpus :

1. les auteurs des textes doivent être des spécialistes ou du moins des personnes connaissant très bien le domaine;
2. les textes doivent provenir de sites Internet ou de publications reconnus dans le domaine;
3. les textes doivent appartenir aux mêmes niveaux de spécialisation;
4. l'équilibre entre la taille et le nombre de documents dans le corpus doit être respecté;
5. les textes français du corpus comparable doivent être des textes originaux;
6. les textes français du corpus parallèle doivent être des traductions<sup>68</sup>.

Un des critères que nous aurions pu ajouter est celui de l'origine régionale des textes. Cependant, comme nous avons sélectionné de nombreux rapports internationaux – traductions incontournables du domaine étudié, il n'est pas possible de connaître la nationalité des traducteurs.

Présumant que les documents électroniques parallèles fiables seraient moins nombreux que les données comparables (Pearson 2000; Véronis 2000), nous avons construit le corpus parallèle avant de créer la sous-partie française du corpus comparable. Néanmoins, la sélection des textes pour les deux corpus présente des difficultés qui leur sont propres.

En ce qui a trait à la constitution du corpus parallèle, vu la large couverture médiatique dont jouit le sujet, nous avons trouvé un nombre amplement suffisant de

---

<sup>68</sup> En raison de la nature de notre étude, les deux derniers points sont particulièrement importants.

textes<sup>69</sup>. Par contre, pour certaines paires de documents trouvées (anglais-français), nous avons souvent éprouvé de la difficulté à déterminer lequel des deux textes était l'original ou la traduction. De nombreux sites Web, comme celui de l'Agence européenne pour l'environnement, donnent rarement des indications à cet effet. Lorsque nous n'avons pas pu contacter les responsables des sites concernés pour valider le statut des textes, nous avons écarté ces derniers.

La sélection des textes pour le corpus comparable ne présente pas de difficulté en soi, car les sources de langue française sont nombreuses. Cependant, à l'instar du corpus parallèle, il n'est pas toujours aisé de savoir si un texte est une traduction ou un original. Par ailleurs, afin d'assurer une similarité maximale entre le corpus parallèle et le corpus comparable, en plus des critères de sélection de base, nous avons respecté deux points importants : le genre textuel et la synchronie des textes.

En ce qui concerne le genre textuel, nous avons pris soin d'inclure autant que possible des documents du même genre que ceux du corpus parallèle. Dans le Tableau 5.1, nous remarquons que les rapports sur le changement climatique constituent la majeure partie du corpus. Ce genre de texte, qui s'adresse plus particulièrement aux décideurs et aux chercheurs, se présente généralement sous le même modèle : 1) description du phénomène et les causes; 2) évaluation des impacts; 3) mesures d'atténuation.

Tableau 5.1. Types de document dans les deux corpus

Type de document	Corpus parallèle	Nombre de mots	Corpus comparable	Nombre de mots
Rapport	27	958 295	23	992 515
Ouvrage scientifique	0	0	1	40 317
Étude	2	78 148	3	48 195
Dossier technique	4	109 151	4	73 487
Brochure	2	19 247	0	0
Article de revue ou article spécialisé	3	47 663	14	73 754
<b>Total</b>	<b>38</b>	<b>1 212 504</b>	<b>45</b>	<b>1 228 268</b>

<sup>69</sup> Pour certains domaines moins traduits, il serait plus difficile de rassembler suffisamment de textes pour construire un corpus parallèle de taille importante.

Pour ce qui est de la synchronie dans les deux corpus, les textes ont été écrits dans une fenêtre de temps relativement réduite, en l'occurrence sur une période de 13 ans, de 1997 à 2009 (Tableau 5.2). Afin de construire un corpus couvrant suffisamment le sujet et de taille convenable, nous considérons cette fenêtre de temps adéquate. Bien sûr, compte tenu de la conjoncture actuelle, on peut supposer que le vocabulaire du domaine du changement climatique évolue très vite. Cela est certainement vrai, mais en partie seulement. En effet, nous avons constaté que la hiérarchie domaine et sous-domaine ne convient pas à ce domaine<sup>70</sup>. En fait, les phénomènes du changement climatique ne peuvent être expliqués que grâce à de nombreux domaines considérés autonomes (biologie, botanique, climatologie, développement durable, écologie, environnement, foresterie, géochimie, géologie, hydrologie, météorologie, océanographie, vulcanologie, zoologie, etc.)<sup>71</sup>. Dans ces conditions, nous pouvons nous attendre à ce que la terminologie d'un de ces domaines donnés n'évolue pas au même rythme que celle des autres.

Tableau 5.2. Analyse comparative des années de parution des documents des deux corpus

Année de parution	Corpus parallèle	Corpus comparable
1997	4	3
1998	1	1
1999	2	1
2000	4	1
2001	4	2
2002	3	5
2003	1	2
2004	1	1
2005	4	4
2006	1	6
2007	3	6
2008	4	5
2009	6	8
<b>Total</b>	<b>38</b>	<b>45</b>

<sup>70</sup> De façon classique, les domaines chapeautent des sous-domaines, par exemple : domaine (médecine), sous-domaine (cardiologie, cancérologie, pédologie, etc.).

<sup>71</sup> Le changement climatique n'est pas le seul à faire exception à la règle : domaine, sous-domaine. On peut aussi penser au domaine du développement durable ou à celui des nanotechnologies.

Les textes sélectionnés des deux corpus proviennent principalement d'organismes mondiaux (ex. Intergovernmental Panel on Climate Change (IPCC), UNESCO), européens (ex. Commission européenne (CE), Commission de coopération environnementale (CCE)) et gouvernementaux (ex. France, Canada). De façon générale, les textes sont spécialisés, car ils s'adressent aux décideurs, aux chercheurs et au public averti.

De plus, pour être certaine de bien sélectionner des textes appartenant au domaine du changement climatique et non pas à un autre domaine, au cours de la recherche des documents, nous avons utilisé quelques mots clés très généraux, comme *changement climatique*, *émissions*, *gaz à effet de serre*, *carbone*. Bien sûr, *changement climatique* est le mot clé auquel nous avons eu le plus souvent recours. Toutefois, nous avons évité de nous appuyer trop fortement sur ces indices linguistiques et d'employer des termes plus spécifiques afin de ne pas uniformiser la terminologie des deux corpus, puisque le but de notre recherche est de comparer l'équivalence en corpus parallèle à celle de l'équivalence en corpus comparable et d'accéder à toute la diversité que le phénomène peut présenter. À l'instar de Condamines (2003 : 31–32), nous pensons qu'il n'est pas toujours souhaitable de s'appuyer sur des critères linguistiques « intuitifs » lors de l'élaboration d'un corpus, d'autant plus lorsqu'on veut mettre en évidence des régularités ou des irrégularités linguistiques. En fait, pour assurer la comparabilité des corpus, nous avons surtout misé sur le grand nombre de rapports qui composent chacun des corpus (27 pour le parallèle et 23 pour le comparable). Ces textes, rédigés selon un modèle standard, abordent un grand nombre de sujets communs et tentent de répondre aux mêmes préoccupations.

Enfin, que ce soit pour un corpus ou pour l'autre, nous avons remarqué que, dès que l'on trouve un document de première importance, d'autres sont souvent proposés dans sa bibliographie. En cherchant les textes cités, on parvient de fil en aiguille à constituer un corpus dans lequel les textes entretiennent des liens et forment un tout cohérent.

### **5.1.2 Prétraitement des textes**

Pour être exploités par les différents logiciels que nous utilisons (aligneur, extracteur de termes, concordancier), les documents des corpus sont prétraités, nettoyés et convertis au

format texte. Cette étape est cruciale, car comme le souligne Kilgarriff (2007 : 148–149) « all further layers of linguistic processing depend on the cleanliness of the data ». Les textes sélectionnés se présentent pour une très grande part sous format PDF, quelques-uns sous format HTML, un seul provient d'un ouvrage papier. Malheureusement, comme nous le verrons plus loin, les textes fondamentaux sur le changement climatique, diffusés par les organismes internationaux et gouvernementaux, se présentent la plupart du temps sous forme de PDF.

Pour l'ensemble du corpus (parallèle et comparable), nous avons éliminé des périclives<sup>72</sup> les parties qui présentent peu d'intérêt pour le domaine étudié et qui gonflent inutilement le nombre de mots du corpus. Par exemple, les bibliographies ou les préfaces consacrées aux remerciements des chercheurs scientifiques ayant contribué à l'élaboration du document sont retirées.

La conversion des documents PDF au format texte et la reconnaissance optique de caractères des textes papier engendrent des coquilles qui doivent être corrigées. Parfois, le nettoyage de ces textes convertis comporte tellement d'erreurs qu'il est préférable de les mettre de côté, pour éviter de trop en laisser passer (Figure 5.2). Les PDF ont également la particularité de terminer chacune de leurs lignes par une marque de paragraphe ¶ (Figure 5.2). Étant donné que les marques de paragraphe font partie des délimiteurs de fin de phrase pour les logiciels, tels l'aligneur et le concordancier, il est nécessaire d'éliminer celles qui sont superflues.

L'khelle de temps selon laquelle le climat reagit est de ¶  
 I'ordre de quelques dkennies : aprPs que la concentra- ¶  
 tion de gaz A effet de serre ait et6 stabilisbe, il faudrait ¶  
 attendre plusieurs dizaines d'ankes avant que la tempb- ¶  
 rature ne cesse de croitre. Cette constante de temps s'ad- ¶  
 ditionne A celle necessaire a la stabilisation de la concen- ¶  
 tration pour determiner le rythme auquel le climat @pond ¶  
 aux variations des &missions.¶

Figure 5.2. Exemple de coquilles apparaissant lors de la conversion au format texte

<sup>72</sup> « Ensemble des textes qui complètent le texte principal d'un ouvrage écrit dont ils font partie (préface, notes, glossaire...). » (*Le Nouveau petit Robert* 2002)

Ensuite, pour maximiser la qualité de l'alignement des textes du corpus parallèle plusieurs types de nettoyage sont nécessaires :

- les tableaux sont soit éliminés, soit déplacés<sup>73</sup> (ils sont éliminés s'ils ne contiennent que très peu ou pas de texte et déplacés s'ils comportent suffisamment de texte);
- les glossaires sont enlevés des textes étant donné que leur présentation par ordre alphabétique<sup>74</sup> empêche leur alignement automatique;
- les numéros de pages et les en-têtes sont également supprimés, car lors de la conversion au format texte, ils s'insèrent dans les phrases chevauchant deux pages. Par exemple, au cours de la conversion au format texte, l'en-tête et le numéro de page de la Figure 5.3 se sont insérés dans la phrase qui commence à la page 12 et qui se termine à la page 13, ce qui donne « *Les options de politique pour L'aviation et l'atmosphère planétaire 13 réduire davantage les émissions comprennent...* » (IPCC\_aviation\_1999).

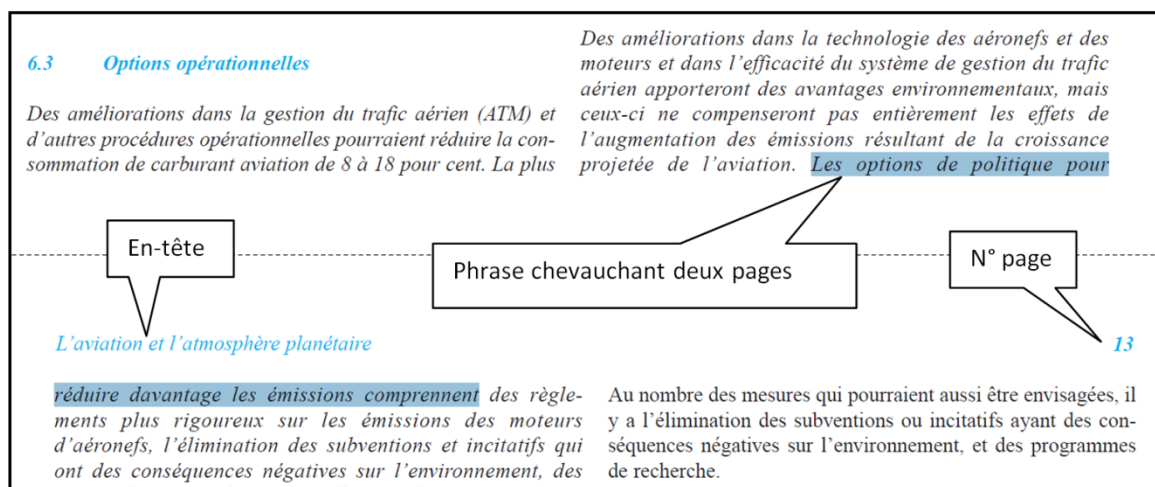


Figure 5.3. Illustration du problème posé par les en-têtes et les numéros de page

- les notes de bas de page sont déplacées en fin de texte, car souvent elles n'apparaissent pas sur les pages qui se correspondent dans les deux langues.

<sup>73</sup> Les tableaux ne sont pas toujours insérés entre les paragraphes qui se correspondent dans les deux langues.

<sup>74</sup> Les termes anglais ne commencent pas toujours par la même lettre que l'équivalent français.

Pour conclure, chaque type de texte ou de format de fichier possède ses particularités qu'il convient de bien analyser avant de procéder au prétraitement des textes. En outre, chaque sorte de corpus demande que l'on tienne compte de ses caractéristiques. Par exemple, les corpus parallèles exigent une attention particulière en ce qui a trait à leur bon alignement.

### 5.1.3 Noms des fichiers

Tous les fichiers sont nommés selon la même procédure : domaine, organisme ou auteur, mot clé du titre, année de parution, langue (ex. CC\_IPCC\_aviation\_1999.fr). Cette façon de désigner les fichiers joue le rôle d'aide mémoire et facilite le travail de classification et d'analyse des textes. En outre, il respecte la convention de nommage de l'aligneur Alinea (cf. 5.1.5.1). Le Tableau 5.3 montre un aperçu des textes des organismes figurant dans les corpus. Le corpus comparable est constitué pour une très large part de textes provenant d'Europe, alors que le corpus parallèle se compose majoritairement de textes internationaux.

Tableau 5.3. Abréviations et noms des organismes apparaissant dans les deux corpus

Organismes du corpus parallèle		Organismes corpus comparable	
Abréviation	Nom au long	Abréviation	Nom au long
AEE	Agence européenne pour l'environnement	AEE	Agence européenne pour l'environnement
CCE	Commission de coopération environnementale	CES	Conseil économique et social (Europe)
CE	Commission européenne	CNRS	Centre national de la recherche scientifique
CES	Conseil économique et social (Europe)	Doc française	La documentation française
ECA	Environnement Canada	ESCRIME	Étude sur les simulations climatiques
IPCC	Intergovernmental Panel on Climate Change	INRA	Institut national de la recherche agronomique
NRTEE	National Round Table on the Environment and the Economy (Canada)	INSU	Institut national des sciences de l'univers
ONU	Organisation des Nations Unies	Lettre	La lettre de l'Académie des sciences

PNUE	Programme des Nations Unies pour l'environnement	ONERC	Observatoire national des effets du réchauffement climatique
RNCA	Ressources naturelles Canada	ONU	Organisation des Nations Unies
UNESCO	Organisation des Nations Unies pour l'éducation, la science et la culture	QUEBEC	Commission des transports et de l'environnement
UNFCCC	Convention-cadre sur les changements climatiques (Nations Unies)	RNCA	Ressources naturelles Canada
UNFPA	United Nations Population Fund	Senat	Sénat (France)

### 5.1.4 Taille des corpus

Le sous-corpus français du corpus comparable compte 1 228 268 occurrences<sup>75</sup> et le sous-corpus français du corpus parallèle, 1 212 504 occurrences. Le sous-corpus anglais, quant à lui, contient 992 488 occurrences (Figure 5.4). Le corpus comparable se compose de 45 textes et le corpus parallèle de 38. On peut consulter la bibliographie complète des deux corpus à l'Annexe A.

Une fois les corpus construits, ils n'ont plus été modifiés par la suite. C'est pourquoi la mise en forme d'un corpus, en particulier un corpus d'analyse, doit faire l'objet d'une démarche précise et rigoureuse.

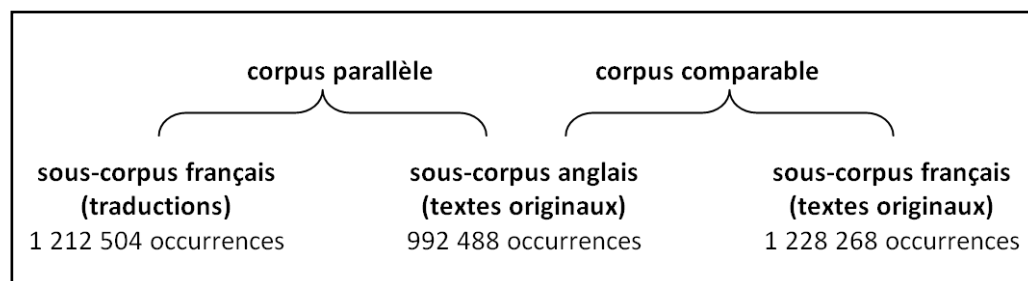


Figure 5.4. Taille des sous-corpus

### 5.1.5 Alignement du corpus parallèle au niveau des phrases

Pour aligner les textes du corpus parallèle, nous utilisons le gratuit Alinea<sup>76</sup> (Kraif 2008). Ce logiciel dédié à la constitution et à l'édition de corpus parallèles cumule plusieurs fonctions : alignement automatique au niveau des phrases et au niveau des mots, extraction

<sup>75</sup> Les occurrences sont calculées à l'aide du logiciel de traitement de texte Word.

<sup>76</sup> Adresse Internet du logiciel Alinea : <http://w3.u-grenoble3.fr/kraif/>



de lexiques bilingues, comparaison d'alignements, édition manuelle des alignements et enfin recherche complexe et concordance avec des critères bilingues. Dans cette étude, nous utiliserons entre autres l'aligneur de phrases, le concordancier et l'extracteur de lexiques bilingues.

Dans les sections qui suivent, nous décrivons brièvement les techniques utilisées par Alinea pour aligner au niveau des phrases (5.1.5.1) et nous donnons les résultats de l'alignement de notre corpus parallèle (5.1.5.2).

### 5.1.5.1 Description de l'aligneur d'Alinea

Avec Alinea, l'alignement au niveau des phrases repose sur deux étapes principales (*cf.* 4.2.1) : *l'extraction des points d'ancrage* et *le calcul du meilleur chemin d'alignement* (Kraif 2008 : 73).

Dans la première étape, Alinea procède à l'identification des chaînes de caractères identiques (transfuges)<sup>77</sup>. Et, lorsque c'est nécessaire<sup>78</sup>, il utilise la ressemblance superficielle de certaines chaînes de caractères (cognats)<sup>79</sup>. Le logiciel commence par délimiter de vastes zones de texte en isolant les indices les plus fiables (les nombres ou les marques de chapitres). Puis, à l'intérieur de ces zones, il extrait les points d'ancrage les moins équivoques (noms propres, abréviations, etc.). De cette façon, l'extraction successive des points d'ancrage permet d'aligner des zones textuelles de granularité de plus en plus fine.

Dans la deuxième étape, l'alignement phrastique proprement dit, Alinea détermine quel est le meilleur chemin d'alignement en s'appuyant principalement sur le rapport de longueur des phrases (calculé sur la base du nombre de caractères ou d'occurrences<sup>80</sup>). D'autres indices peuvent intervenir, mais ils doivent être fixés au choix (densité de cognats, score des correspondances lexicales, nombre de mots équivalents à partir d'un lexique

---

<sup>77</sup> L'auteur nomme *transfuge* toute chaîne de caractères qui ne subit pas de transformation lors du passage du texte source au texte cible, par exemple, les dates, les noms propres, les numéros de chapitre, etc.

<sup>78</sup> Cette étape n'est pas incluse dans les paramètres par défaut, car les cognats sont des points d'ancrage moins fiables que les transfuges. Cependant, ils peuvent être utiles pour certains couples de langues.

<sup>79</sup> Dans Alinea, les cognats sont calculés sur la base de *la plus longue sous-chaîne commune*, par exemple, les chaînes *élémentaire* et *elementary* partagent une même sous-chaîne de 7 caractères : l-m-e-n-t-a-r.

<sup>80</sup> Par défaut, le calcul est paramétré sur le nombre de caractères.

bilingue) (Kraif 2010). Grâce à ces indices, l'algorithme d'alignement se charge alors de calculer le meilleur chemin d'alignement en déterminant la meilleure suite de regroupements de phrases<sup>81</sup>.

Ainsi, en cumulant les informations, des plus fiables au plus aléatoires, et en faisant appel à un algorithme récursif (itératif), le programme converge vers un alignement de plus en plus précis.

Alinea est un produit de laboratoire en constante évolution qui s'adresse surtout aux chercheurs. Le logiciel a été évalué au cours de la campagne d'évaluation Arcade 2 (Chiao *et al.* 2006) et a obtenu, pour des langues apparentées, des résultats très satisfaisants, environ 98 % (à 3 dixièmes du meilleur système (Kraif 2008 : 76)). Dans cette étude, nous utilisons la version 3.55. Le logiciel accepte plusieurs types de formats en entrée (txt, ces, xml, ttg, etc.) et il exporte les fichiers d'alignement sous différents formats (html, cesalign, tmx, txt, etc.).

### 5.1.5.2 Résultats de l'alignement du corpus parallèle au niveau des phrases

Pour ce projet, nous avons commencé par étiqueter les textes du corpus parallèle avec TreeTagger (Schmid 1994). L'étiquetage permet à Alinea de passer à la phase de l'extraction du lexique bilingue, cette étape est expliquée à la section (*cf.* 4.2.2.1). Nous n'avons pas fusionné les textes, car Alinea peut aligner une série de textes si l'on a pris soin d'identifier les fichiers en respectant la convention de nommage. Cette procédure permet d'identifier chacun des documents dans le résultat de l'alignement, il est donc possible de connaître immédiatement à quel texte appartient un contexte (voir *adresse du contexte* dans la Figure 5.5).

Pour aligner notre corpus, Alinea a relevé 26 845 points d'ancrage et produit 39 954 couples de phrases. Nous n'avons pas utilisé l'éditeur d'alignements, car nous considérons le résultat satisfaisant<sup>82</sup>. Naturellement, le logiciel effectue quelques mauvais alignements, mais pas suffisamment pour justifier une vérification phrase à phrase.

<sup>81</sup> Les regroupements (ou transitions) prédéfinis par Alinea sont 1-1, 1-0, 0-1, 1-2, 2-1, 2-2, 3-1 et 1-3.

<sup>82</sup> Dans Le Serrec (2008), nous avons estimé que les alignements étaient bons à environ 98 %.

<p>[CC_ECA_extremes_1998-s5187 CC_ECA_extremes_1998-s5188] Precipitation is one half of the hydrological ( or water ) cycle . ✕ Evaporation ( and transpiration from plants ) is the other .</p>	<p>[CC_ECA_extremes_1998-s5089] Les précipitations constituent une moitié du cycle de l' eau ( cycle hydrologique ) ; l' autre moitié est l' évaporation ( et la transpiration des végétaux ) .</p>
<p>[CC_ECA_extremes_1998-s5189] A virtually certain outcome of a rise in global temperatures is a widespread increase in the amount of water that is moved through the cycle .</p>	<p>[CC_ECA_extremes_1998-s5090] Un des résultats les plus probables d'une hausse des températures planétaires est un accroissement généralisé de la quantité d'eau circulant dans le cycle .</p>
<p>[CC_ECA_extremes_1998-s5190] That is because higher temperatures not only increase evaporation and transpiration , but also raise the air 's capacity to hold moisture .</p>	<p>[CC_ECA_extremes_1998-s5091] En effet , le réchauffement fera augmenter non seulement l' évaporation et la transpiration , mais aussi la capacité de l' air de retenir l' humidité .</p>
<p>[CC_ECA_extremes_1998-s5191] Consequently , more moisture will be available in the atmosphere to fall as rain and snow .</p>	<p>[CC_ECA_extremes_1998-s5092] Il y aura donc dans l' atmosphère plus d'eau susceptible de tomber sous forme de pluie et de neige .</p>
<p>[CC_ECA_extremes_1998-s5192] Add to this a more unstable atmosphere due to increased convection over warmer land and sea surfaces , and the result is an increased potential for major precipitation events in many parts of the world .</p>	<p>[CC_ECA_extremes_1998-s5093] En outre , les terres et les eaux étant plus chaudes , il y aura davantage de convection et l' atmosphère sera plus instable , ce qui accroît le risque d'importants épisodes de précipitations dans bien des régions du monde .</p>

Figure 5.5. Échantillon au format HTML de l'alignement au niveau phrastique

## 5.2 Extraction des candidats termes

L'extraction, dans le sous-corpus anglais, des termes simples (nom, verbe, adjectif, adverbe) est confiée au logiciel d'acquisition automatique de termes TermoStat (Drouin 2003, Drouin et Doll 2008). L'utilisation d'un extracteur de termes permet à l'auteur de ne pas directement prendre part au choix des termes, afin de ne pas influencer les résultats finaux. Cette méthode possède également l'avantage, du moins dans une première étape, d'extraire rapidement des données et de les quantifier.

Dans la section 5.2.1, nous esquissons le fonctionnement de TermoStat. Dans la section 5.2.2, nous donnons les résultats de l'extraction des candidats termes (CT) anglais.

### 5.2.1 Fonctionnement de TermoStat

TermoStat est un logiciel accessible gratuitement en ligne<sup>83</sup>. Il extrait des CT simples et complexes<sup>84</sup>. TermoStat s'appuie sur une approche contrastive, ce qui le distingue de la plupart des outils. En effet, TermoStat met en opposition un corpus non spécialisé (corpus de référence) à un corpus spécialisé<sup>85</sup>. La méthode part du principe que la comparaison d'un corpus de référence de grande taille (textes hétérogènes de langue générale) à un corpus spécialisé (textes homogènes en langue de spécialité) permet de faire ressortir les spécificités lexicales de ce dernier. La notion de « spécificité » renvoie à l'idée que dans un corpus spécialisé certains termes lui sont particuliers et apparaissent relativement plus souvent que dans un corpus de référence.

De par son fonctionnement, TermoStat peut être classé parmi les outils d'extraction hybrides, car il combine de façon équilibrée des méthodes linguistiques et statistiques. TermoStat effectue l'extraction des CT en quatre étapes.

Dans la première étape, TermoStat confie la segmentation, l'étiquetage et la lemmatisation des corpus à TreeTagger (Schmid 1994).

Dans la deuxième étape, TermoStat prend le relais. Il sélectionne des candidats termes en se basant sur les patrons morphosyntaxiques typiques des termes (NOM + NOM, ADJ + NOM, NOM + PRÉP + NOM, etc.).

La troisième étape, qui met en œuvre l'approche contrastive, évalue le caractère terminogène<sup>86</sup> des CT sélectionnés à l'étape précédente. Pour ce faire, grâce à des calculs statistiques, TermoStat compare la fréquence des CT sélectionnés aux mêmes unités du corpus de référence et attribue à chacun d'eux un indice de spécificité (score). Dans la version 3.0 de TermoStat, plusieurs tests statistiques sont offerts : *Spécificité*, *Weirdness*, *Log Likelihood*, *Log Odds Ratio*, *Chi 2*.

---

<sup>83</sup> Adresse Internet du logiciel TermoStat : [http://olst.ling.umontreal.ca/~drouinp/termostat\\_web/](http://olst.ling.umontreal.ca/~drouinp/termostat_web/)

<sup>84</sup> TermoStat offre également la possibilité d'extraire seulement des termes simples ou seulement des syntagmes nominaux.

<sup>85</sup> Notamment, le corpus anglais construit pour cette étude.

<sup>86</sup> Terminologie utilisée par Kocourek (1971 : 105) pour désigner les unités lexicales que l'on peut considérer comme des termes.

La quatrième étape consiste à éliminer (filtrer) les CT ne possédant pas un score suffisamment élevé par rapport à celui observé dans le corpus de référence. Enfin, TermoStat propose une liste de CT triée par ordre décroissant du score attribué. Les CT recevant un score supérieur à un seuil donné sont les plus susceptibles d’être étroitement liés à la terminologie du corpus analysé.

### 5.2.2 Résultats de l’extraction des CT anglais

Pour l’extraction des termes simples anglais, nous avons utilisé la version 3.0 de TermoStat. Les textes soumis à l’extracteur doivent être convertis au format texte brut. Avant de lancer l’extraction, nous avons coché la case “terme simple” et sélectionné les quatre parties du discours : nom, adjectif, verbe, adverbe. Le résultat de l’extraction est présenté sur une page html en cinq colonnes : candidat de regroupement (forme lemmatisée du CT), fréquence, score, variantes orthographiques (formes fléchies) et partie du discours (Figure 5.6). Plusieurs onglets permettent d’accéder à des informations complémentaires, par exemple, les résultats statistiques de l’extraction (Figure 5.7).

<b>change</b>	8912	199.19	<i>change changes</i>	Nom
<b>have</b>	6516	191.34	<i>have had having</i>	Verbe
<b>emission</b>	4483	156.36	<i>emission emissions</i>	Nom
<b>increase</b>	3369	137.55	<i>increase increases increasing</i>	Verbe
<b>impact</b>	3276	125.24	<i>impact impacts</i>	Nom
<b>include</b>	2517	118.88	<i>include includes including</i>	Verbe
<b>water</b>	3757	112	<i>water waters</i>	Nom
<b>adaptation</b>	2238	110.6	<i>adaptation adaptations</i>	Nom
<b>be</b>	1930	104.09	<i>been</i>	PPAdj
<b>reduce</b>	1807	100.71	<i>reduce reduces reducing</i>	Verbe
<b>temperature</b>	2185	100.48	<i>temperature temperatures</i>	Nom

Figure 5.6. Échantillon de page html montrant des résultats de l’extraction avec TermoStat

Candidat de regroupement	Fréquence	Score (Spécificité)	Vari
be	26988	389.91	be am was
climate	9090	223.63	climate climates

Figure 5.7. Image agrandie des onglets de la page des résultats html de TernoStat

Le corpus de référence anglais de la version 3.0 de TernoStat compte environ 8 millions d’occurrences. Il est composé pour moitié de textes provenant d’articles de journaux de *The Gazette* et pour moitié de textes du corpus BNC. Les articles de journaux tirés du quotidien montréalais *The Gazette* ont été publiés en 1989 et traitent de divers sujets d’actualité. Le corpus *BNC* (British National Corpus) compte environ 100 millions d’occurrences. Il est constitué de textes de l’anglais écrit ou parlé qui proviennent de diverses sources.

À l’issue de l’extraction des termes, nous avons retenu le test de spécificité (cf. 5.2.1), car il a fait l’objet d’une analyse par le concepteur de TernoStat dans lequel il s’est révélé utile du point de vue de l’extraction de la terminologie :

One more conclusion can be drawn from our work: although Labbé and Labbé (2001) proved that the specificity test is unstable when looking at lower frequency items, we can still use its results for term extraction and obtain good quality CTs. The test might lead to results that are not interesting from a statistical point of view, but are highly useful from a terminological standpoint. (Drouin 2003 : 12–13)

Les CT pris en considération affichent un score égal ou supérieur à +3,09, seuil adopté par le concepteur du logiciel pour le test de spécificité. Selon Drouin (2003 : 3), les CT qui reçoivent un score supérieur à ce seuil sont susceptibles d’être les formes les plus intéressantes et significatives du corpus. L’extraction a généré 4 677 CT simples, dont

1 928 noms, 1 305 adjectifs, 1 263 verbes et 181 adverbes. Le Tableau 5.4 offre le pourcentage de CT par partie du discours.

Tableau 5.4. Nombre et pourcentage de CT par parties du discours

Partie du discours	Nombre	Pourcentage
nom	1 928	41,2
adjectif	1 305	27,9
verbe	1 263	27,0
adverbe	181	3,9
<b>Total</b>	<b>4 677</b>	<b>100</b>

### 5.3 Nettoyage de la liste d'extraction et validation des CT anglais

La présente section rend compte de la sélection des termes anglais simples à partir de la liste d'extraction des CT simples. Cette sélection s'effectue en deux grandes étapes : 1) nettoyage de la liste d'extraction des 4 677 CT; 2) validation des CT restants grâce aux critères d'identification des termes (*cf.* 5.3.2.1).

Dans les paragraphes qui suivent, nous présentons en premier le nettoyage de la liste de candidats termes (5.3.1). Puis, nous décrivons le processus de validation des CT restants (5.3.2). Enfin, nous donnons les résultats de la validation (5.3.3).

#### 5.3.1 Première étape : nettoyage de la liste de candidats termes

Tout d'abord, afin d'éviter de sélectionner des CT provenant d'un trop petit nombre d'auteurs ou même d'un seul, nous avons décidé d'éliminer les CT présents dans moins de 8 textes, soit 21 % du nombre de textes du corpus. Ce qui a eu pour conséquence immédiate d'éliminer 1 507 CT dont la fréquence était inférieure à 8. Ensuite, sur les 3 170 CT restants, nous avons écarté :

- les CT dont la répartition est inférieure à 8<sup>87</sup> textes, (ex. *deglaciation* ayant une fréquence de 10 n'est présent que dans 6 textes);
- les formes au pluriel ayant déjà une forme au singulier dans la liste d'extraction (ex. *streamflows* est éliminé, car *streamflow* est aussi présent);
- les CT mal étiquetés (ex. *Ganges*, nom propre);
- les CT comportant des erreurs (ex. *conserva* pour *conservation*);
- les CT appartenant à la langue générale (ex. *chapter*, *good*, *very*, *propose*);
- les mesures et les abréviations (ex. *gt*, *yr*);
- les CT faisant uniquement partie d'un terme complexe (ex. le CT *greenhouse* dans *greenhouse effect*, le CT *dioxide* dans *carbon dioxide*);
- les CT n'étant pas reliés de façon évidente au domaine du changement climatique (ex. *policymaking*, *portfolio*, *table*, *figure*).

Au terme de l'étape du nettoyage de la liste de CT anglais, il restait 527 CT<sup>88</sup>.

### 5.3.2 Deuxième étape : validation des 527 CT

Dans la deuxième étape, les 527 CT sont validés à l'aide de critères d'identification des termes. Les CT remplissant les conditions sont retenus, les autres sont écartés.

Nous présentons en premier les critères servant à l'identification des termes (5.3.2.1). Puis nous décrivons brièvement le processus de validation (5.3.2.2).

#### 5.3.2.1 Critères d'identification des termes

Dans cette étude, les termes sélectionnés touchent le domaine du changement climatique. Pour attester les termes, nous nous basons sur les quatre critères d'identification des termes de L'Homme (2004 : 64–66).

<sup>87</sup> La répartition est calculée à l'aide du concordancier AntConc (Anthony 2007). L'onglet *Concordance Plot* permet de rapidement repérer les textes contenant le terme cherché.

<sup>88</sup> L'élimination des CT dont la répartition est inférieure à 8 à probablement écarté des termes valides, mais il était important pour nous de ne pas choisir des termes trop peu fréquents ou trop spécifiques à certains auteurs.



Le critère a) s'appuie sur les connaissances du domaine de spécialité étudié et les critères b), c) et d) sur des indices lexico-sémantiques. Ces trois derniers critères nous permettent de valider des termes que les dictionnaires spécialisés et les banques de terminologie ne contiennent pas toujours, notamment les verbes, les adjectifs et les adverbes ainsi que certains noms dérivés de verbes et d'adjectifs.

- a) L'unité lexicale a un sens qui est lié à un domaine de spécialité; ce dernier est délimité au préalable pour un projet terminologique donné (*ibid.* : 64). Par exemple, tous les experts en terminologie s'accorderont pour associer les lexies anglaises *ozone, gas, aerosol, atmosphere, emission* et *albedo* au domaine du changement climatique.

Pour le critère a), nous nous reposons sur des banques de terminologie (GDT, IATE, Termium, UNTERM), des lexiques et des glossaires (Brodhag et Breuil, IPCC 1995, IPCC 2007, Jacques et Le Treut 2004). Les ouvrages retenus pour la validation des termes anglais couvrent une période allant de 1995 à 2010 (voir Annexe C à la section « Ressources terminologiques utilisées pour la validation des termes anglais »). De plus, en raison de l'organisation peu conventionnelle du domaine (*cf.* 5.1.1), ces sources se rapportent au domaine étudié, mais aussi à des domaines intimement liés à ce dernier – d'autant plus qu'aucun ouvrage sur le changement climatique ne dresse une nomenclature suffisamment exhaustive des termes qui lui sont reliés.

- b) La nature des actants sémantiques peut servir d'indice pour confirmer le sens spécialisé d'une unité lexicale à sens prédicatif. Si les actants sont déjà admis comme des termes en fonction du critère a), l'unité de sens prédicatif risque fort d'être spécialisée elle-même (*ibid.* : 64–65). Par exemple dans la phrase suivante le verbe *warm* admet deux actants appartenant au domaine du changement climatique : *forcing* et *atmosphere* :

This results in a positive radiative forcing that tends to WARM the lower atmosphere and surface. (IPCC\_synthese\_2001)

- c) La parenté morphologique – nécessairement accompagnée d’une parenté sémantique – est un autre indice permettant de confirmer un sens spécialisé. Si des termes ont été retenus en vertu des critères a) et b), leurs dérivés apparentés sémantiquement sont forcément spécialisés (*ibid.* : 65–66). Par exemple, si on a déjà accepté que le terme *atmosphere* fasse partie du domaine du changement climatique, nous devons également inclure *atmospheric* :

ATMOSPHERIC concentration of long-lived, and hence well-mixed, greenhouse gases have increased substantially in all areas of the world. (ECA\_projections\_2000)

...and that very substantial cuts in emissions would be required for stabilization of greenhouse gas concentrations in the ATMOSPHERE. (IPCC\_synthese\_2001)

- d) Toute autre relation paradigmatique, c’est-à-dire autre que morphologique, partagée par une unité lexicale avec un terme déjà admis en fonction des trois premiers critères peut révéler un sens spécialisé (*ibid.* : 66). Par exemple, si le verbe *warm* (to) est retenu parmi les termes du domaine du changement climatique, on devra admettre le verbe *cool* (to), qui en est l’antonyme.

Increases in CO2 concentration are expected TO COOL the stratosphere... (IPCC\_ozone\_2005)

### 5.3.2.2 Étapes de la validation

Une simple liste de CT n’est pas suffisante pour valider les CT, pour cela il faut consulter le contexte. La consultation du corpus, se fait avec le concordancier AntConc (Anthony 2007). Le processus de validation des CT suit l’ordre suivant :

1. Premièrement, pour valider les CT selon le critère a), nous consultons à l’Annexe C les sources de la section A. Par exemple, le nom *desertification* est validé avec ce critère puisqu’il est attesté dans Brodhag et Breuil (2009) et qu’il est associé au domaine.

2. Ensuite, pour les CT absents de ces sources, notamment les verbes, les adjectifs et les adverbes, nous passons aux critères b), c) ou d). Dans l'exemple suivant, le verbe *change*, qui ne figure dans aucune des sources consultées, peut être validé par le critère b), car ses actants (*greenhouse gas* et *climate*) sont déjà admis comme des termes en fonction du critère a).

Greenhouse gases ARE CHANGING our climate.

Pour chaque CT validé, sont relevés dans un fichier Excel :

- sa partie du discours;
- son score donné par TermoStat;
- sa fréquence donnée par TermoStat;
- sa répartition dans le corpus (calculée à l'aide du concordancier AntConc);
- le critère d'identification (a, b, c ou d);
- une définition ou un contexte (pour les critères b), c) et d) notamment, il n'y a pas de définition. Dans ces cas-là, nous prenons un contexte du corpus.);
- la source de la définition ou du contexte;
- l'équivalent français (lorsque présent dans les sources terminologiques et autres dictionnaires de l'Annexe C à « Ressources terminologiques utilisées pour la validation des termes anglais »).

Tableau 5.5. La fiche du terme *permafrost* accompagnée de ses informations

Terme	Partie du discours	Score	Fréquence	Répartition	Critère d'identification	Source
<i>permafrost</i>	NOM	44,39	352	15/38	a)	IATE
<b>Définition ou contexte</b>	Permanently frozen ground that occurs wherever ground temperatures remain continuously below 0 °C for two or more years.					
<b>Équivalents proposés</b>	<i>pergélisol</i> <i>permagel</i>					

Dans le Tableau 5.5, les équivalents apparaissant dans la colonne *Équivalents proposés* sont relevés dans les sources terminologiques consultées au moment de la validation des termes anglais, ce qui permet d'accélérer le processus de recherche des équivalents dans le corpus comparable.

Les CT ne pouvant être validés selon un des critères d'identification sont rejetés. Au terme de cette étape, 343 CT sur 527 remplissent au moins un des critères de L'Homme (2004 : 64–66).

### 5.3.3 Résultats de la validation des termes anglais

Des 343 termes retenus, 271 sont validés avec le critère a) et 72 avec les critères b), c) ou d)<sup>89</sup>. Le Tableau 5.6 donne en détail les données chiffrées de la validation des termes par critère d'identification, par partie du discours et par pourcentage des parties du discours.

Il est intéressant de noter que 29 verbes sur 31 ont été validés par les critères b), c) ou d). Les adjectifs, quant à eux, ont été validés un peu plus souvent par le critère a) que par les trois autres réunis. De leur côté, les noms ont été dans une très large proportion validés par le critère a). Pour la validation des verbes, des adjectifs et des adverbes, les critères b), c) ou d) se montrent très utiles.

Dans le tableau, on remarquera, entre autres, que les noms représentent à eux seuls plus de 73 % des termes alors que les adverbes n'en constituent que 2 %. La comparaison des pourcentages du Tableau 5.4 au Tableau 5.6 montre que, proportionnellement, les noms sont plus susceptibles d'être validés que les autres parties du discours. De 41,2 % dans l'extraction brute, les noms sont passés à 73,8 % au cours de la validation des termes, alors que le pourcentage de toutes les autres parties du discours a baissé.

---

<sup>89</sup> Les critères b), c) et d) sont pris en bloc, car certains termes peuvent être validés à l'aide des trois critères, par exemple, le verbe *cool (to)*.

Tableau 5.6. Données chiffrées de la validation des 343 termes anglais

Critère	Nom	Adjectif	Verbe	Adverbe	Total
a)	236	30	2	3	271
b), c) ou d)	17	22	29	4	72
<b>Total</b>	<b>253</b>	<b>52</b>	<b>31</b>	<b>7</b>	<b>343</b>
<b>Pourcentage</b>	73,8	15,2	9	2	

À l'Annexe B, on trouvera un tableau contenant la liste complète des termes anglais validés ainsi que les informations qui les accompagnent. En outre, à la section 6.1, nous présentons une analyse détaillée des termes anglais quant à leur répartition dans le corpus, leur score et leur fréquence.

#### 5.4 Collecte des cooccurrents des termes anglais

Selon l'hypothèse de la sémantique distributionnelle (*cf.* 4.3), deux mots sont sémantiquement reliés si leurs contextes sont similaires. Par conséquent, plus souvent deux termes de langues différentes partagent des cooccurrents comparables plus ils ont de fortes chances d'être des traductions l'un de l'autre. Au moment du repérage et de la validation des équivalents dans le corpus comparable, les cooccurrents des termes anglais pourront être comparés aux cooccurrents des équivalents français<sup>90</sup>. La comparaison des cooccurrents peut mener à des équivalents jusque-là ignorés.

En extraction bilingue sur corpus comparable, dans une fenêtre plus ou moins large (*cf.* 4.3.5), les mots (cooccurrents) qui se trouvent dans le voisinage du terme pour lequel un équivalent doit être trouvé sont collectés automatiquement et enregistrés sous forme de vecteurs de contextes (Prochasson 2009; Zweigenbaum et Habert 2006).

À l'instar des travaux sur corpus comparable, nous relevons les mots qui entourent chacun des termes anglais validés, mais nous le faisons semi-automatiquement à l'aide du

<sup>90</sup> Les stratégies servant à repérer les équivalents dans le corpus comparable sont énumérées et expliquées à la section 5.6.

concordancier AntConc (Anthony 2007), plus précisément avec la fonction d'extraction des cooccurrents (*Collocates*).

Seuls les mots lexicaux (nom, verbe, adjectif, adverbe) et les noms propres sont retenus. Ces derniers présentent également de l'intérêt : par exemple, *El Niño*, *ENSO*, *Atlantic*, *Pacific* et *Arctic* sont des unités qui caractérisent le terme *oscillation* (Figure 5.8). Pour faciliter la collecte des cooccurrents, nous avons préalablement enlevé les mots grammaticaux (déterminants, prépositions, pronoms, etc.) du corpus, comme on peut le constater dans le résultat de la Figure 5.8. De plus, la fenêtre de recherche est fixée par défaut à deux mots lexicaux à gauche et deux mots lexicaux à droite du terme pour lequel nous collectons les cooccurrents. Cette largeur de fenêtre, souvent utilisée en extraction automatique sur corpus comparable, donne suffisamment de cooccurrents significatifs. Enfin, la fréquence minimale des cooccurrents relevées est fixée à deux (Figure 5.8). Paramétrer la fréquence à une seule occurrence serait beaucoup trop bruyant et peu productif, alors qu'une fréquence plus élevée ne conviendrait pas aux termes peu fréquents.

AntConc 3.2.1w (Windows) 2007

Global Settings Tool Preferences About

Corpus Files  
all.en.txt

Concordance Concordance Plot File View Clusters Collocates Word List Keyword List

Total No. of Collocate Types: 40 Total No. of Collocate Tokens: 341

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	70	1	1	8.052	Oscillation
2	31	31	0	11.49	Southern
3	25	22	3	8.358	North
4	25	25	0	10.70	Niño
5	25	23	2	9.133	Atlantic
6	24	1	23	10.90	ENSO
7	17	11	6	9.562	Pacific
8	17	0	0	-1	oscillations
9	10	9	1	13.29	Decadal
10	9	0	9	10.69	NAO
11	8	0	0	-1	oscillation
12	6	0	6	10.29	phenomenon
13	6	0	6	10.81	PDO
14	6	5	1	6.410	Arctic
15	5	4	1	9.353	decadal
16	4	3	1	2.164	climate
17	3	0	3	4.626	well
18	3	0	3	7.555	linked
19	3	0	3	4.785	events
20	3	2	1	5.550	atmospheric
21	3	0	3	5.127	associated

Search Term  Words  Case  Regexp  
 Window Span  Same  
 From... 2L To... 2R  
 Min. Collocate Frequency 2  
 Sort by Sort by Freq  Invert Order  
 Total No. 1  
 Files Processed  
 Reset

Figure 5.8. Extraction des cooccurrents du terme anglais *oscillation* dans l'onglet *Collocates* d'AntConc

D'autre part, nous évitons de relever les cooccurrents des homographes catégoriels et des homographes sémantiques<sup>91</sup>. Étant donné que nous n'avons pas étiqueté notre corpus pour qu'AntConc puisse distinguer les homographes catégoriels, nous sélectionnons les cooccurrents d'un équivalent en consultant son concordancier (Figure 5.9). Par exemple, pour l'homographe catégoriel anglais *current* ayant pour sens (déplacement continu d'un fluide dans une direction), nous ne retenons que les cooccurrents appartenant au terme déjà validé avec les critères de L'Homme (2004) à la section 5.3.2.2. Ainsi, dans la Figure 5.9,

<sup>91</sup> Les systèmes d'extraction automatiques peuvent distinguer les homographes catégoriels si les textes sont étiquetés, mais ils ne peuvent pas différencier les homographes sémantiques (cf. 4.3.5).

nous observons que le terme *current* a pour cooccurrents *ocean*, *wind*, *surface*. Après vérification de tous les contextes dans lesquels apparaît le terme analysé, la liste des cooccurrents relevés se présente comme suit : *Labrador*, *colder*, *strengthening*, *southward*, *NAO*, *variation*, *carries*, *ice*, *fed*, *cold*, *nutrient*, *transport*, *carry*, *ocean*, *wind*, *surface*, etc. Les cooccurrents de l'adjectif *current* sont quant à elles ignorées (Figure 5.9), car ce mot n'est pas un terme validé. Le même procédé est utilisé pour les homographes sémantiques tels que *source* (origine des émissions) (ouvrage de référence) ou (origine d'un cours d'eau), *wood* (matière ligneuse) ou (groupe d'arbres), *flood* (débordement des eaux d'une rivière) ou (élévation périodique ou anormale d'un cours d'eau).

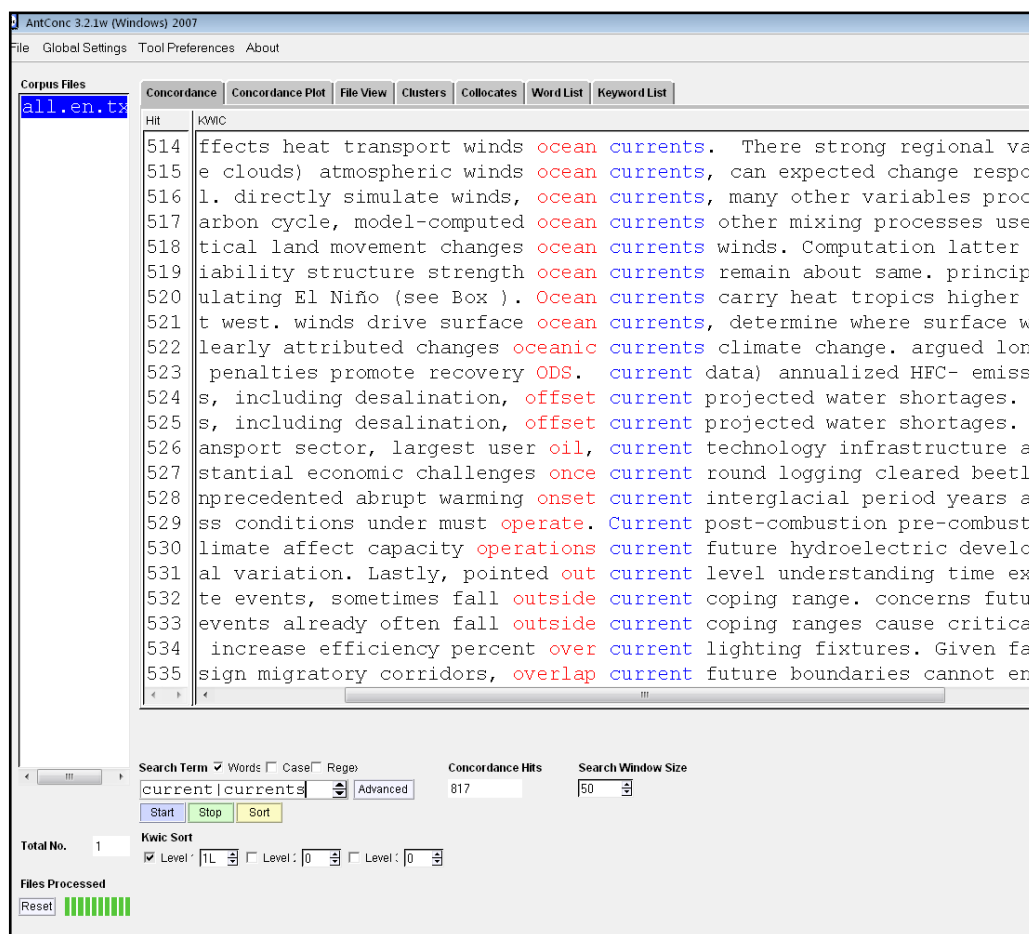


Figure 5.9. Recherche des cooccurrents de l'homographe sémantique *current* dans le concordancier d'AntConc



Le nombre de cooccurrents prélevées pour chaque terme varie en fonction de plusieurs facteurs, le plus important étant la fréquence du terme. En principe, plus il y a de contextes pour un terme, plus les cooccurrents sont nombreux. Le Tableau 5.7 montre la fiche d'un terme dans laquelle sont ajoutées les cooccurrents retenus.

Tableau 5.7. La fiche du terme permafrost accompagnée de toutes ses informations

Terme	Partie du discours	Score	Fréquence	Répartition	Critère d'identification	Source
<i>permafrost</i>	NOM	44,39	352	15/38	a)	IATE
<b>Définition ou contexte</b>	Permanently frozen ground that occurs wherever ground temperatures remain continuously below 0 °C for two or more years.					
<b>Équivalents proposés</b>	<i>pergélisol</i> <i>permagel</i>					
<b>Cooccurrents</b>	<i>ice, thawing, melting, area, ground, change, thaw, warming, sea, region, erosion, infrastructure, northern, coastal, snow, increased, zone, lake, melt, temperature</i>					

## 5.5 Repérage et validation des équivalents dans le corpus parallèle

Le repérage et la validation des équivalents dans le corpus parallèle se font en deux étapes bien distinctes, la première est automatique et la seconde manuelle. La section 5.5.1 présente le repérage des candidats équivalents (CE). La section 5.5.2 décrit la validation des CE à l'aide de plusieurs critères.

### 5.5.1 Repérage des équivalents dans le corpus parallèle

Les corpus parallèles ont la particularité de pouvoir être alignés phrase à phrase (Figure 5.10). L'équivalent d'un terme d'une langue source est en principe présent dans la phrase correspondante de la langue cible. Cette caractéristique permet de repérer automatiquement de façon suffisamment fiable les candidats équivalents simples grâce à l'analyse distributionnelle des chaînes de caractères (*cf.* 4.2).

What are scenarios and what is their purpose?	Que sont les scénarios et quel en est le but ?
Future greenhouse gas (GHG) emissions are the product of very complex dynamic systems, determined by driving forces such as demographic development, socio-economic development, and technological change.	Les émissions futures de gaz à effet de serre sont le produit de systèmes dynamiques très complexes déterminés par des forces motrices telles que la croissance démographique, le développement socio-économique et l'évolution technologique.
Their future evolution is highly uncertain.	Leur évolution future est hautement incertaine.
Scenarios are alternative images of how the future might unfold and are an appropriate tool with which to analyse how driving forces may influence future emission outcomes and to assess the associated uncertainties.	Les scénarios sont des images diverses du déroulement possible du futur et ils constituent un outil approprié pour analyser comment des forces motrices peuvent influencer sur les émissions futures et pour évaluer les incertitudes connexes.
They assist in climate change analysis, including climate modeling and the assessment of impacts, adaptation, and mitigation.	Ils aident à analyser l'évolution du climat, notamment sa modélisation et l'évaluation des impacts, l'adaptation et l'atténuation.

Figure 5.10. Repérage automatique des équivalents en corpus parallèle

Le repérage des équivalents dans le corpus parallèle s'effectue avec le module d'extraction des correspondances d'Alinea (Kraif 2008). Ce module, qui a déjà fait l'objet d'une évaluation dans Le Serrec (2008), extrait les candidats équivalents des termes simples qui lui sont soumis. Les résultats de la mise en correspondance sont présentés sur une page HTML (Figure 5.11). Dans cette figure, il est utile de noter que lorsque plusieurs candidats équivalents sont trouvés pour un même terme source, ils sont présentés par ordre décroissant de fréquence.

measure-VWG	mesurer-VER (14) : <a href="#">s1328</a> <a href="#">s1928</a> <a href="#">s4157</a> <a href="#">s12878</a> <a href="#">s12897</a> <a href="#">s128</a> mesure-NOM (10) : <a href="#">s2071</a> <a href="#">s2540</a> <a href="#">s6901</a> <a href="#">s6952</a> <a href="#">s8228</a> <a href="#">s16454</a>
measure-VVN	mesurer-VER (51) : <a href="#">s352</a> <a href="#">s1617</a> <a href="#">s2307</a> <a href="#">s3640</a> <a href="#">s4192</a> <a href="#">s4215</a> <a href="#">s4</a> <a href="#">s11698</a> <a href="#">s12638</a> <a href="#">s12882</a> <a href="#">s13466</a> <a href="#">s13498</a> <a href="#">s13975</a> <a href="#">s14607</a> <a href="#">s146</a> <a href="#">s24889</a> <a href="#">s25041</a> <a href="#">s25198</a> <a href="#">s25199</a> <a href="#">s25402</a> <a href="#">s25543</a> <a href="#">s25628</a> <a href="#">s256</a> mesure-NOM (9) : <a href="#">s1074</a> <a href="#">s3858</a> <a href="#">s6738</a> <a href="#">s10920</a> <a href="#">s16461</a> <a href="#">s1843</a>
meteorological-	météorologique-ADJ (14) : <a href="#">s3371</a> <a href="#">s3647</a> <a href="#">s3673</a> <a href="#">s3675</a> <a href="#">s13970</a> météorologie-NOM (4) : <a href="#">s20045</a> <a href="#">s28381</a> <a href="#">s28698</a> <a href="#">s37801</a>
methane-NN	méthane-NOM (163) : <a href="#">s77</a> <a href="#">s100</a> <a href="#">s2465</a> <a href="#">s2469</a> <a href="#">s3279</a> <a href="#">s3777</a> <a href="#">s3</a> <a href="#">s4220</a> <a href="#">s4225</a> <a href="#">s4227</a> <a href="#">s4228</a> <a href="#">s4229</a> <a href="#">s4230</a> <a href="#">s4231</a> <a href="#">s4302</a> <a href="#">s4306</a> <a href="#">s6177</a> <a href="#">s6210</a> <a href="#">s6353</a> <a href="#">s6686</a> <a href="#">s6690</a> <a href="#">s7031</a> <a href="#">s7132</a> <a href="#">s7513</a> <a href="#">s8268</a> <a href="#">s8540</a> <a href="#">s8541</a> <a href="#">s8542</a> <a href="#">s8568</a> <a href="#">s8577</a> <a href="#">s8582</a> <a href="#">s8583</a> <a href="#">s8695</a> <a href="#">s8710</a> <a href="#">s13661</a> <a href="#">s13668</a> <a href="#">s16087</a> <a href="#">s16447</a> <a href="#">s16448</a> <a href="#">s16448</a> <a href="#">s16449</a> <a href="#">s164</a> <a href="#">s17937</a> <a href="#">s18005</a> <a href="#">s18074</a> <a href="#">s18153</a> <a href="#">s18271</a> <a href="#">s18272</a> <a href="#">s18907</a> <a href="#">s189</a> <a href="#">s22175</a> <a href="#">s22176</a> <a href="#">s22176</a> <a href="#">s22178</a> <a href="#">s22183</a> <a href="#">s22183</a> <a href="#">s22259</a> <a href="#">s222</a> <a href="#">s23539</a> <a href="#">s23621</a> <a href="#">s24245</a> <a href="#">s24333</a> <a href="#">s25099</a> <a href="#">s25296</a> <a href="#">s25331</a> <a href="#">s253</a> <a href="#">s27078</a> <a href="#">s27923</a> <a href="#">s29209</a> <a href="#">s29222</a> <a href="#">s29222</a> <a href="#">s29231</a> <a href="#">s29232</a> <a href="#">s292</a> <a href="#">s33612</a> CH4-NOM (4) : <a href="#">s22172</a> <a href="#">s22265</a> <a href="#">s22268</a> <a href="#">s22856</a>
mitigate-VV	atténuer-VER (24) : <a href="#">s2115</a> <a href="#">s2137</a> <a href="#">s2173</a> <a href="#">s2875</a> <a href="#">s2964</a> <a href="#">s3028</a> <a href="#">s</a> <a href="#">s17839</a> <a href="#">s18650</a> <a href="#">s18792</a> <a href="#">s28535</a> <a href="#">s29786</a> <a href="#">s32416</a> <a href="#">s41969</a> atténuation-NOM (10) : <a href="#">s1781</a> <a href="#">s3429</a> <a href="#">s4802</a> <a href="#">s6988</a> <a href="#">s9003</a> <a href="#">s10</a>
mitigate-VWG	atténuer-VER (8) : <a href="#">s3069</a> <a href="#">s3453</a> <a href="#">s3463</a> <a href="#">s16270</a> <a href="#">s28541</a> <a href="#">s2876</a> atténuation-NOM (5) : <a href="#">s3052</a> <a href="#">s3419</a> <a href="#">s12258</a> <a href="#">s18542</a> <a href="#">s18554</a>
mitigation-NN	atténuation-NOM (684) : <a href="#">s1</a> <a href="#">s188</a> <a href="#">s193</a> <a href="#">s198</a> <a href="#">s199</a> <a href="#">s419</a> <a href="#">s1652</a> <a href="#">s3034</a> <a href="#">s3068</a> <a href="#">s3333</a> <a href="#">s3342</a> <a href="#">s3344</a> <a href="#">s3349</a> <a href="#">s3352</a> <a href="#">s3367</a> <a href="#">s3377</a> <a href="#">s6668</a> <a href="#">s6702</a> <a href="#">s6704</a> <a href="#">s6711</a> <a href="#">s6712</a> <a href="#">s6714</a> <a href="#">s6715</a> <a href="#">s6719</a> <a href="#">s6722</a> <a href="#">s6830</a> <a href="#">s6830</a> <a href="#">s6831</a> <a href="#">s6846</a> <a href="#">s6847</a> <a href="#">s6847</a> <a href="#">s6854</a> <a href="#">s6874</a> <a href="#">s6884</a> <a href="#">s6932</a> <a href="#">s6938</a> <a href="#">s6942</a> <a href="#">s6951</a> <a href="#">s6953</a> <a href="#">s6956</a> <a href="#">s6958</a> <a href="#">s6960</a> <a href="#">s6964</a>

Figure 5.11. Échantillon de l'extraction des correspondances lexicales avec Alinea

L'utilisation de ce module nous permet d'analyser l'ensemble du corpus beaucoup plus rapidement que si on le faisait à la main et donne de bons résultats. D'autre part, comme le texte est étiqueté, les diverses parties du discours d'une même forme graphique (ex. *current-NOM* et *current-ADJ*) peuvent être distinguées automatiquement. Pour obtenir le maximum d'équivalents par terme, sans toutefois générer trop de bruit, nous avons paramétré<sup>92</sup> les filtres d'Alinea pour qu'ils présentent des équivalents ayant un nombre

<sup>92</sup> Par défaut, Alinea est paramétré de la façon suivante : Nombre minimum de correspondances 2 (calculé en fonction de la fréquence absolue d'une même paire d'équivalents dans le corpus), Proportion minimale de correspondance 0,3 (calculée en fonction de la fréquence relative des toutes les paires d'équivalents dans le corpus. Ce paramètre sert donc à filtrer le nombre de paires d'équivalents afin de ne présenter que les paires les plus fréquentes, ainsi une proportion fixée à 0,5 ne montrerait que la moitié des paires calculées par l'outil). Pour notre type de recherche, la proportion minimale de correspondance de 0,3 présente trop de silences, c'est pourquoi nous l'avons réduite à 0.

minimum de 4 occurrences et nous avons fixé la proportion minimale de correspondances à 0. Dans cette étude, nous estimons que les CE ayant une fréquence inférieure à 4 risqueraient de bruyter nos observations. En effet, les CE dont la fréquence varie entre 1 et 3 occurrences sont souvent des coquilles ou des erreurs accidentelles, ce qui nous en apprendrait peu sur l'équivalence. Le Tableau 5.8 présente quelques erreurs trouvées dans le corpus parallèle. La colonne Erreur indique l'erreur et le nombre d'occurrences entre parenthèses. La colonne Équivalent indique l'équivalent correct.

Tableau 5.8. Exemple de coquilles ou d'erreurs dans le corpus parallèle

Terme anglais	Erreur	Type d'erreur	Équivalent
<i>nutrient</i>	nutrient (1)	faute d'orthographe	<i>nutriant</i>
<i>deposition</i>	déposition (1)	faux sens	<i>dépôt</i>
<i>afforestation</i>	déboisement (2)	faux sens	<i>boisement</i>
<i>precipitation</i>	precipitation (2)	faute d'orthographe	<i>précipitation</i>

### 5.5.2 Validation des équivalents dans le corpus parallèle

La deuxième étape consiste à valider la liste des candidats équivalents présentée par Alinea. Tout d'abord, nous avons départagé les différents sens d'une même forme graphique anglaise, car seuls les équivalents qui correspondent au terme anglais validé sont conservés. Deux cas de figure peuvent se présenter : 1) un terme anglais validé partage la même forme graphique qu'un terme non validé (ex. *plant* 'végétal' et *plant* 'usine'); 2) deux termes validés partagent la même forme (ex. *snowmelt*<sub>1</sub> 'fonte de la neige' et *snowmelt*<sub>2</sub> 'eau de fonte'). Pour trancher entre les différents sens d'une même forme anglaise, nous nous basons sur les définitions ou les contextes des termes anglais relevés au moment de la validation (cf. 5.3.2.2), les cooccurrents significatifs des termes anglais (cf. 5.4) et nos connaissances du domaine.

Une fois le sens du terme anglais bien circonscrit, nous relevons parmi les CE proposés par Alinea celui ou ceux qui lui correspondent. Les hyperliens qui mènent aux

contextes alignés sont également une aide précieuse pour confirmer ou infirmer nos choix. D'ailleurs, tous les équivalents ne sont validés qu'après la vérification des contextes.

Grâce à l'étiquetage, le logiciel est capable de distinguer les différentes parties du discours d'une même forme graphique (ex. *plant*-NOM et *plant*-VER), par conséquent il ne présentera pas les équivalents des deux formes ensemble. Par contre, le logiciel ne peut pas différencier le sens 'végétal' du sens 'usine' et offre tout ce qu'il trouve. Dans le Tableau 5.9, nous présentons les différents CE proposés par Alinea pour l'homographe sémantique *plant*-NOM.

Tableau 5.9. L'homographe sémantique anglais *plant* et ses candidats équivalents français

Terme anglais	Candidat équivalent	Contexte
<i>plant</i> -NOM	<i>plante</i> -NOM (141)	photosynthesis in PLANTS is expected... – la photosynthèse des PLANTES devrait...
	<i>végétal</i> -ADJ (125)	total PLANT biomass – la biomasse VÉGÉTALE totale
	<i>usine</i> -NOM (104)	...a PLANT in Malaysia – ...une USINE en Malaisie
	<i>centrale</i> -NOM (91)	repowering existing PLANTS – la rénovation des CENTRALES existantes
	<i>électrique</i> -ADJ (26)	power PLANT – centrale ÉLECTRIQUE <sup>93</sup>
	<i>installation</i> -NOM (25)	Advanced PLANT designs... – La mise au point d'INSTALLATIONS plus performantes...
	<i>végétal</i> -NOM (16)	non-native PLANTS – des VÉGÉTAUX non-indigènes
	<i>production</i> -NOM (16)	bruit
	<i>central</i> -ADJ (9)	bruit
	<i>flore</i> -NOM (8)	PLANTS and animals – FLORE et faune

En nous basant sur le sens du terme *plant*-NOM ('végétal') retenu lors de la validation des termes et sur les contextes du corpus, 4 équivalents français sont acceptés : *plante*-NOM, *végétal*-ADJ, *végétal*-NOM et *flore*-NOM (Tableau 5.10).

<sup>93</sup> Lors de la mise en équivalence, Alinea a choisi le modificateur *électrique* au lieu de la tête du terme *centrale*.

Tableau 5.10. Équivalents validés pour le terme anglais *plant*-NOM ('végétal')

Terme anglais	Équivalents	ÉqT/ÉqNT
<i>plant</i> -NOM	<i>plante</i> -NOM (141)	ÉqT
	<i>végétal</i> -ADJ (125)	ÉqNT
	<i>végétal</i> -NOM (16)	ÉqNT
	<i>flore</i> -NOM (8)	ÉqNT

Avant de poursuivre, à l'aide des exemples du Tableau 5.10, nous devons introduire une notion qui va nous être utile dans le Chapitre 8 pour classifier les équivalents d'un terme anglais donné. Dans cette thèse, les unités lexicales attestées dans les dictionnaires en tant qu'équivalent français d'un terme anglais sont qualifiées d'*équivalents typiques* (ÉqT). Par exemple, l'unité lexicale *plante*-NOM est l'équivalent typique du terme anglais *plant*-NOM parce qu'il est proposé par les sources terminologiques et autres dictionnaires de l'Annexe C. Par contre, les unités lexicales non attestées dans les dictionnaires en tant qu'équivalent d'un terme anglais sont appelées « équivalent non typique » (ÉqNT), par exemple, les unités lexicales *végétal*-ADJ, *végétal*-NOM et *flore*-NOM ne sont pas des équivalents typiques du terme anglais *plant*-NOM. Pour illustrer plus concrètement un exemple d'équivalent non typique, prenons les contextes suivants :

As a consequence, there would be little need to MONITOR the disposal sites and the associated risks would be very low.

La SURVEILLANCE des sites d'élimination serait donc peu utile, car les risques sont restreints. (IPCC\_dioxyde\_2005)

Dans ces contextes, nous remarquons que le terme anglais (to) *monitor* passe de la forme verbale à la forme nominale dans la traduction française. Les dictionnaires ne tiennent pas compte de ce genre de phénomène et proposent le verbe *surveiller* comme équivalent.

En général, pour un même terme anglais, les équivalents typiques sont plus fréquents que les équivalents non typiques. Par exemple, avec ses 141 occurrences, l'équivalent *plante*-NOM (141) est le plus fréquent des quatre équivalents validés (Tableau 5.10). Par contre, il arrive qu'un ÉqNT soit plus fréquent qu'un ÉqT. Ainsi, dans notre

corpus l'ÉqT *climat*-NOM (2 033) du terme anglais *climate*-NOM est beaucoup moins fréquent que l'ÉqNT *climatique*-ADJ (5 792). La distinction entre ÉqT et ÉqNT nous sera utile notamment pour caractériser l'équivalence en corpus.

## **5.6 Repérage et validation des équivalents dans le corpus comparable**

Contrairement aux corpus parallèles, l'espace de recherche des équivalents ne peut être réduit à la phrase dans les corpus comparables. L'équivalent d'un terme dans une langue A peut se trouver n'importe où dans le texte d'une langue B (Figure 5.12). Tant que des outils d'extraction de terminologie bilingue suffisamment fiables ne sont pas disponibles, le repérage et la validation des équivalents s'effectuent de façon plus classique, c'est-à-dire manuellement, nous pouvons tout au plus disposer de concordanciers. Par conséquent, le repérage et la validation des équivalents ne se font pas en deux opérations aussi distinctes que dans les corpus parallèles. Toutefois, notre validation s'inspire de techniques utilisées en extraction à partir de corpus comparables, sans s'y limiter.

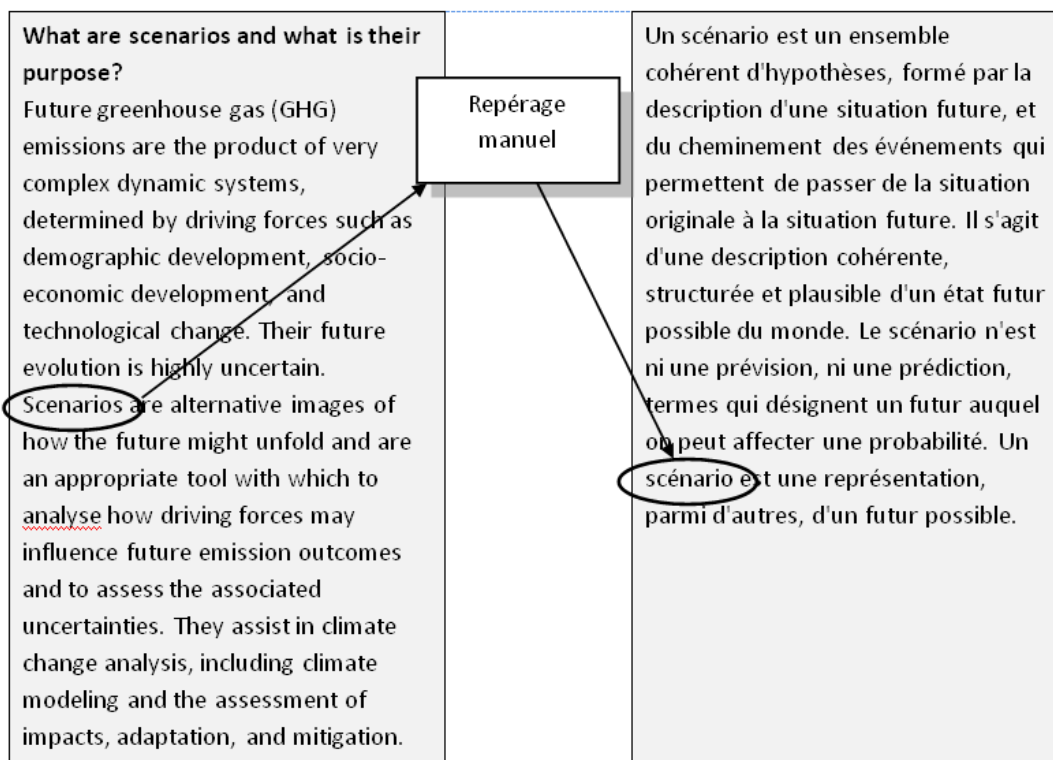


Figure 5.12. Repérage semi-manuel des équivalents en corpus parallèle

Pour les mêmes raisons que dans le corpus parallèle, nous ne retenons dans le corpus comparable que les équivalents dont la fréquence est égale ou supérieure à quatre (4). Une fréquence inférieure à 4 ne nous permet pas de trancher entre une erreur accidentelle ou un équivalent délibérément choisi. Le Tableau 5.11 présente quelques erreurs trouvées dans le corpus comparable. La colonne Erreur indique l'erreur et le nombre d'occurrences entre parenthèses. La colonne Équivalent indique l'équivalent corrigé.

Tableau 5.11. Exemples de coquilles et d'erreurs dans le corpus comparable

Terme anglais	Erreur	Type d'erreur	Équivalent
<i>dioxide</i>	dioxyde (2)	faute d'orthographe	<i>dioxyde</i>
<i>sustainable</i>	soutenable (3)	faux sens	<i>durable</i>
<i>atmospheric</i>	atmosphérique (1)	faute d'orthographe	<i>atmosphérique</i>
<i>decrease</i>	diminution (1)	faute d'orthographe	<i>diminution</i>



### 5.6.1 Méthode de repérage et de validation des équivalents en corpus comparable

Ci-après, nous expliquons les trois grandes étapes du processus de repérage et de validation des équivalents (*analyse du terme anglais, validation de l'équivalent proposé, recherche avancée des équivalents*) que nous illustrons à l'aide du terme *uptake*-NOM.

#### 1) Première étape : analyse du terme anglais

Dans un premier temps, nous prenons connaissance des informations qui figurent dans la fiche du terme anglais validé (Tableau 5.12). Souvenons-nous que deux cas de figure peuvent se présenter : 1) un terme anglais validé partage la même forme graphique qu'un terme non validé; 2) deux termes validés partagent la même forme (*cf.* 5.5.2). Pour trancher entre les différents sens d'une même forme anglaise, nous nous basons donc sur les définitions ou les contextes de cette fiche, les cooccurents significatifs des termes anglais et nos connaissances du domaine.

Tableau 5.12. Fiche du terme anglais *uptake*-NOM

Terme	Partie du discours	Score	Fréquence	Répartition	Critère d'identification	Source
<i>uptake</i>	NOM	24,13	111	16/38	a)	UNTERM
<b>Définition ou contexte</b>	Addition d'une substance (ex. CO <sub>2</sub> , chaleur) à une destination (ex. plantes, océan). « Plants are responsible for the vast majority of UPTAKE of carbon... » (IPCC_biodiversite_2002)					
<b>Équivalents proposés</b>	<i>absorption</i>					
<b>Cooccurents</b>	<i>carbon, ocean, CO<sub>2</sub>, rate, terrestrial, land, photosynthesis, ecosystem, emission, increase, change, respiration, biological, anthropogenic.</i>					

#### 2) Deuxième étape : validation de l'équivalent proposé

Nous cherchons dans la partie française du corpus le ou les équivalents proposés par les ressources terminologiques au moment de la validation des termes anglais (*cf.* 5.3.2). Dans les cas où la forme graphique de l'équivalent est présente, nous nous assurons que le sens qu'il exprime est bien le même que celui du terme anglais. Cette vérification se fait en

consultant les contextes avec le concordancier AntConc (Anthony 2007). Lorsque l'équivalent est trouvé, nous relevons ses cooccurrents selon les mêmes critères que pour l'anglais afin de les comparer aux cooccurrents des termes anglais (*cf.* 5.4)<sup>94</sup>. Pour chacun des équivalents proposés, nous recommençons la deuxième étape.

<b>Équivalent proposé lors de la validation du terme <i>uptake</i>-NOM :</b> <i>absorption</i> -NOM
<b>Contexte d'emploi de l'équivalent <i>absorption</i> dans le corpus comparable :</b> « ... inclure dans les calculs l'ABSORPTION naturelle de CO <sub>2</sub> par les forêts... » (Jacques_Le-Treut_chang_2004)
<b>Cooccurrents du terme anglais <i>uptake</i>-NOM :</b> <i>carbon, ocean, CO<sub>2</sub>, rate, terrestrial, land, photosynthesis, ecosystem, heat, emission, increase, change, respiration, biological, anthropogenic, plant</i>
<b>Cooccurrents de l'équivalent <i>absorption</i>-NOM :</b> <i>rayonnement, gaz, capacité, puits, CO<sub>2</sub>, chaleur, augmentation, énergie, émission, réduire, forêt, infrarouge, dioxyde, océan</i>
<b>Au terme de cette analyse, l'équivalent <i>absorption</i>-NOM est validé.</b>

### 3) Troisième étape : recherche avancée des équivalents

Afin de trouver les équivalents d'un terme anglais, il est nécessaire de recourir à d'autres méthodes, car deux cas de figure peuvent se présenter : 1) tous les équivalents d'un terme ne sont pas toujours proposés par les ressources terminologiques; 2) pour certains termes anglais validés, nous n'avons pas pu relever d'équivalent dans les ressources qui ont servi à leur validation, notamment pour les verbes, les adjectifs et les adverbes. À cette fin nous consultons trois types de documents et avons recours à sept stratégies.

*Les trois types de documents consultés :*

1. Dictionnaires bilingues de langue générale papier (ex. Robert & Collins 2003) ou dictionnaires bilingues de langue générale en ligne (ex. OneLook Dictionary Search).

<sup>94</sup> Dans la plupart des cas, il n'y a pas de parallélisme parfait entre la liste anglaise des cooccurrents et la liste française. Toutefois, l'ensemble de ces cooccurrents nous fournit des indices sur le sens de l'équivalent analysé.

2. Dictionnaires monolingues français spécialisés (ex. Glossaire pour le développement durable – Des mots pour tous les mots de la planète 2008; Le dictionnaire de la foresterie 2003).
3. Dictionnaires des synonymes français (ex. Antidote 2011; CRISCO) ou des dictionnaires analogiques (Petit Robert).

La bibliographie complète de ces sources se trouve à l'Annexe C dans les sections B et C. Ces deux types de source servent à repérer des indices d'équivalence et à les valider le cas échéant. Elles couvrent une plus longue période que celle de la section A, 1977 à 2011. En effet, pour repérer des candidats équivalents et mieux comprendre certaines notions, il est important d'examiner une plus large couverture de sources.

*Les sept stratégies utilisées :*

1. Sont envisagés les hyponymes et hyperonymes des équivalents déjà validés (ex. le terme anglais *energy*-NOM est traduit par *énergie*-NOM ou *électricité*-NOM selon le contexte) :

The use of biomass for ENERGY generation offers a promising alternative...

L'utilisation de la biomasse pour la production d'ÉLECTRICITÉ offre une solution prometteuse... (CCE\_carbone\_2001)

2. Sont envisagés les méronymes et les holonymes des équivalents déjà validés (ex. le terme anglais *tree*-NOM est traduit par *arbre*-NOM ou *forêt*-NOM selon le contexte) :

The long growth period before TREES are harvested...

La longue période de croissance qui précède l'exploitation d'une FORÊT... (RNCA\_impacts\_2008)

3. Sont considérées les ressemblances graphiques (cognation) du terme anglais analysé (ex. *offshore*-ADJ);

4. Sont considérées les unités lexicales morphologiquement reliées à un équivalent déjà validé (ex. si l'équivalent *variation*-NOM a été validé, *varier*-VER et *variabilité*-NOM sont envisagés) :

...climate VARIATION

... VARIATION du climat, ... le climat VARIE ou ...la VARIABILITÉ du climat

5. Sont considérés les équivalents des cooccurrents du terme anglais<sup>95</sup>, l'observation des cooccurrents peut mener à des équivalents jusque-là ignorés (ex. le cooccurrent *recul*-NOM permet de mettre en relation *trait de côte*-NOM et *côte*-NOM) :

recul du TRAIT DE CÔTE ou recul de la CÔTE

6. Sont cherchés les contextes définitoires, par exemple le contexte suivant permet de mettre en relation les équivalents *adventice*-NOM et *mauvaise herbe*-NOM;

La prolifération des ADVENTICES ("mauvaises herbes"), dont la croissance sera favorisée par l'augmentation de la température... » (ONERC\_viticulture\_2006)

7. Finalement sont cherchés les équivalents validés dans le corpus parallèle. Cette dernière stratégie est doublement importante. Tout d'abord, elle constitue une stratégie supplémentaire à celles qui viennent d'être décrites. Deuxièmement, elle permet de s'assurer que les équivalents trouvés dans le corpus parallèle soient également pris en compte dans le corpus comparable, permettant ainsi une meilleure comparabilité des équivalents dans les deux types de corpus.

Le Tableau 5.13 illustre la méthode de recherche des équivalents. Nous reprenons le terme anglais *uptake*-NOM et son équivalent *absorption*-NOM, puis nous y ajoutons les équivalents validés grâce aux méthodes énumérées ci-dessus. À côté de chaque équivalent validé apparaissent la méthode et le contexte qui justifient la validation.

---

<sup>95</sup> La section 5.4 décrit la collecte des cooccurrents anglais.

Tableau 5.13. Équivalents validés pour le terme anglais *uptake*-NOM

Équivalent	Explication et contexte
<i>absorption</i> -NOM	<ul style="list-style-type: none"> <li>• <b>Méthode</b> : équivalent proposé par les sources terminologiques lors de la validation du terme anglais <i>uptake</i>-NOM.</li> <li>• <b>Contexte</b> : « ... inclure dans les calculs l'ABSORPTION naturelle de CO<sub>2</sub> par les forêts... » (Jacques_Le-Treut_chang_2004)</li> </ul>
<i>absorber</i> -VER	<ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 4, équivalent morphologiquement relié au terme <i>absorption</i>-NOM.</li> <li>• <b>Contexte</b> : « ... les végétaux, et notamment les arbres, ABSORBENT du gaz carbonique... » (Doc_Francaise_changement_2009)</li> </ul>
<i>fixation</i> -NOM	<ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 7, terme validé dans le corpus parallèle.</li> <li>• <b>Contexte</b> : « ... la FIXATION du carbone par les forêts joue un rôle essentiel... » (Senat_ampleur_2002)</li> </ul>
<i>fixer</i> -VER	<ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 4, équivalent morphologiquement relié au terme validé <i>fixation</i>-NOM.</li> <li>• <b>Contexte</b> : « ... le bois est un matériau qui FIXE le carbone... » (Senat_emission_1999)</li> </ul>
<i>assimilation</i> -NOM	<ul style="list-style-type: none"> <li>• <b>Méthode</b> : 3<sup>e</sup> type de document consulté, synonyme d'<i>absorption</i>-NOM trouvé dans Antidote.</li> <li>• <b>Contexte</b> : « ... Il y a deux principales voies métaboliques d'ASSIMILATION du gaz carbonique atmosphérique par les végétaux... » (ONERC_partie3_2009)</li> </ul>

Le repérage des équivalents est un processus itératif. Toute nouvelle recherche peut apporter des solutions aux analyses précédentes. Une fois toutes les méthodes épuisées, il peut arriver qu'aucun équivalent ne soit trouvé dans la partie française du corpus comparable. Dans ce dernier cas, nous jugeons qu'il est fortement probable que la notion véhiculée par le terme anglais n'est pas présente dans le corpus comparable.

### 5.6.2 Exemples supplémentaires de recherches avancées des équivalents

Afin de mieux illustrer notre méthodologie, nous présentons dans le Tableau 5.14 plusieurs exemples d'équivalents trouvés avec les méthodes décrites au point 5.6.1. La première ligne (*Terme anglais*) présente le terme anglais et un contexte d'emploi. La deuxième ligne

(*Équivalent typique*) donne un équivalent typique, celui proposé par les ressources terminologiques ou dictionnairiques, et un contexte d'emploi. La troisième ligne (*Autre équivalent*) montre un équivalent trouvé à l'aide d'une des méthodes de repérage des termes en corpus comparable (ouvrages ou stratégies) et un contexte d'emploi. Il est à noter que *autre équivalent* désigne soit un ÉqT soit un ÉqNT. En effet, comme nous n'avons pas eu accès à tous les équivalents typiques lors de la validation des termes anglais, nous en avons trouvé d'autres dans les documents consultés à cette étape, par exemple, *fonte nivale*-NOM est aussi un ÉqT (voir fiche *snowmelt*<sub>2</sub>-NOM du Tableau 5.14).

Tableau 5.14. Exemples d'équivalents validés dans le corpus comparable

<b>Terme anglais</b>	<i>reduce</i> -VER <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « ... it includes strategies to REDUCE greenhouse gas sources and emissions... » (NRTEE_true_north_2009)</li> </ul>
<b>Équivalent typique</b>	<i>réduire</i> -VER <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Les marges de manoeuvres économiques, sociales et technologiques afin de RÉDUIRE les émissions de gaz à effet de serre » (Dron_homme_2002)</li> </ul>
<b>Autre équivalent</b>	<i>diminuer</i> -VER <ul style="list-style-type: none"> <li>• <b>Méthode</b> : 1<sup>er</sup> type de document consulté, synonyme trouvé dans le Robert &amp; Collins (2003).</li> <li>• <b>Contexte</b> : « La technique suffit-elle à DIMINUER les émissions de gaz à effet de serre... » (Dron_homme_2002)</li> </ul>

<b>Terme anglais</b>	<i>resilience</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « The RESILIENCE of these ecosystems will depend upon the rate and magnitude of climate change » (UNESCO_patrimoine_2009)</li> </ul>
<b>Équivalent typique</b>	<i>résilience</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « ... 1) le renforcement de la RÉSILIENCE " naturelle " permettant à l'écosystème de résister... » (ONERC_partie2_2009)</li> </ul>
<b>Autre équivalent</b>	<i>résistance</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : 2<sup>e</sup> type de document consulté, quasi-synonyme trouvé dans le Dictionnaire de la protection de la nature (2009).</li> <li>• <b>Contexte</b> : « ...au développement physiologique et à la RÉSISTANCE au froid chez des espèces qui modifient leur aire de répartition. » (ONERC_partie2_2009)</li> </ul>

<b>Terme anglais</b>	<i>livestock</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Impacts on LIVESTOCK » (RNCA_adaptation_2004)</li> </ul>
<b>Équivalent typique</b>	<i>bétail</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Impacts sur le BÉTAIL » (RNCA_adaptation_2004)</li> </ul>
<b>Autre équivalent</b>	<i>cheptel</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : 3<sup>e</sup> type de document consulté, quasi-synonyme trouvé dans Antidote.</li> <li>• <b>Contexte</b> : « ... Anticipation des impacts de l'évolution du climat sur le CHEPTEL, ... » (ONERC_partie3_2009)</li> </ul>

<b>Terme anglais</b>	<i>precipitation</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Small increases in the variability in PRECIPITATION regimes can significantly affect wetland and animals... » (IPCC_eau_2008)</li> </ul>
<b>Équivalent typique</b>	<i>précipitation</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Tout changement du régime des PRÉCIPITATIONS en montagne aura des conséquences sur les systèmes hydrologiques... » (ONERC_impact_2008)</li> </ul>
<b>Autre équivalent</b>	<i>pluie</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 1, hyponyme de <i>précipitation</i>-NOM.</li> <li>• <b>Contexte</b> : « Ainsi, la modification des écoulements dans les sols liée aux modifications du régime des PLUIES... » (ONERC_impact_2008)</li> </ul>

<b>Terme anglais</b>	<i>animal</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Missing snow cover exposes plants and ANIMALS to frost, ... » (IPCC_eau_2008)</li> </ul>
<b>Équivalent typique</b>	<i>animal</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « ... en sont autant de conséquences dévastatrices, aussi bien pour les hommes que pour les ANIMAUX et les plantes. » (ONERC_viticulture_2006)</li> </ul>
<b>Autre équivalent</b>	<i>faune</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 2, holonyme d'<i>animal</i>-NOM.</li> <li>• <b>Contexte</b> : « Les inquiétudes portent sur les risques de fuites et leurs impacts sur la FAUNE et la flore et sur la santé humaine;... » (Campos_captage-et-stockage_2007)</li> </ul>

<b>Terme anglais</b>	<i>mitigation</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Can contribute to the MITIGATION of carbon dioxide emissions. » (IPCC_methodologie_2000)</li> </ul>
<b>Équivalent typique</b>	<i>atténuation</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « peuvent contribuer à l'ATTÉNUATION des émissions de dioxyde de carbone. » (IPCC_methodologie_2000)</li> </ul>
<b>Autre équivalent</b>	<i>mitigation</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 3, équivalent basé sur la ressemblance graphique (cognat).</li> <li>• <b>Contexte</b> : « ... les solutions de MITIGATION des émissions de CO<sub>2</sub>... » (Campos_captage-et-stockage_2007)</li> </ul>

<b>Terme anglais</b>	<i>delta</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Especially at risk are large DELTA regions of Bangladesh » (IPCC_region_1997)</li> </ul>
<b>Équivalent typique</b>	<i>delta</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Les zones les plus menacées sont les grandes régions des DELTAS du Bangladesh » (IPCC_region-1997)</li> </ul>
<b>Autre équivalent</b>	<i>deltaïque</i> -ADJ <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 4, équivalent morphologiquement relié à <i>delta</i>.</li> <li>• <b>Contexte</b> : « ... sont également menacés les espaces DELTAÏQUES... » (Senat_ampleur_2002)</li> </ul>

<b>Terme anglais</b>	<i>coastline</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « accelerated erosion and degradation of the dunes and COATLINES... » (RNCA_impacts_2008)</li> </ul>
<b>Équivalent typique</b>	<i>littoral</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « érosion et dégradation accélérées des dunes et du LITTORAL... » (RNCA_impacts_2008)</li> </ul>
<b>Autre équivalent</b>	<i>trait de côte</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 5, équivalent repéré à l'aide des cooccurrents <i>érosion</i> et <i>recul</i>.</li> <li>• <b>Contexte</b> : « ... l'impact à attendre du fait du recul du TRAIT DE CÔTE par érosion est faible en moyenne... » (ONERC_partie2_2009)</li> </ul>



<b>Terme anglais</b>	<i>snowmelt<sub>2</sub></i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « SNOWMELT is projected to become earlier... » (IPCC_eau_2008)</li> </ul>
<b>Équivalent typique</b>	<i>fonte de la neige</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « La FONTE DE LA NEIGE devrait se produire plus tôt... » (IPCC_eau_2008)</li> </ul>
<b>Autre équivalent</b>	<i>fonte nivale</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 5, autre ÉqT repéré à l'aide de la tête de terme <i>fonte</i> sur le même principe que pour les cooccurrents.</li> <li>• <b>Contexte</b> : « ... l'augmentation des températures induit une date de FONTE NIVALE plus précoce... » (ONERC_Alpes_2008)</li> </ul>

<b>Terme anglais</b>	<i>mitigate</i> -VER <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « Nations considering how best TO MITIGATE climate change need to consider the cost-effectiveness of the options available to them. » (PNUE_bioseq_2009)</li> </ul>
<b>Équivalent typique</b>	<i>atténuer</i> -VER <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « ... contenant des mesures destinées à ATTÉNUER les changements climatiques... » (ONERC_impact_2008)</li> </ul>
<b>Autre équivalent</b>	<i>mitiger</i> -VER <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 6, contexte définitoire.</li> <li>• <b>Contexte</b> : « ... qu'il " incombe aux parties de prendre des mesures de précaution pour prévoir, prévenir ou atténuer les causes du changement climatique ou en limiter les effets néfastes " (c'est-à-dire " MITIGER " le changement climatique). » (Senat_emissions_1999)</li> </ul>

<b>Terme anglais</b>	<i>ocean</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « The ecological consequences of OCEAN acidification caused by increased uptake of inorganic carbon are largely unknown » (PNUE_bioseq_2009)</li> </ul>
<b>Équivalent typique</b>	<i>océan</i> -NOM <ul style="list-style-type: none"> <li>• <b>Contexte</b> : « ... notamment par l'acidification des Océans due à l'augmentation des concentrations en CO2... » (ONERC_partie1_2009)</li> </ul>
<b>Autre équivalent</b>	<i>mer</i> -NOM <ul style="list-style-type: none"> <li>• <b>Méthode</b> : stratégie 7, terme validé dans le corpus parallèle.</li> <li>• <b>Contexte</b> : « Acidification des MERS : Impacts potentiels » (ONERC_impact_2008)</li> </ul>

## 5.7 Conclusion du chapitre

Dans cette section, nous apportons quelques observations et réflexions sur la méthodologie utilisée dans cette thèse.

### *Contextes et listes*

Premièrement, une des grandes leçons à retenir de la validation des termes anglais et de la validation des équivalents français est qu'il faut toujours revenir aux contextes. Aucune conclusion fiable et précise ne peut être tirée à partir de listes d'extraction, que ce soit de listes constituées de données textuelles ou de données chiffrées, d'où l'importance de travailler avec un corpus de taille adéquate – c'est-à-dire de taille suffisamment importante afin d'observer les phénomènes linguistiques avec un maximum de précision. Ce n'est que dans le texte que l'on peut vérifier si *bois*-NOM s'applique à ('matière ligneuse') ou à ('groupe d'arbres'), et ce n'est que dans le texte que l'on peut déterminer la fréquence exacte d'un homographe sémantique. Néanmoins, il reste que les listes sont des auxiliaires précieux dans toute recherche.

### *La validation des équivalents dans les deux corpus*

Que ce soit dans le corpus anglais ou dans le sous-corpus français comparable, il n'est pas toujours évident, même en contexte, de distinguer les sens de certains termes. Par exemple,

dans le corpus anglais pour le terme *plant*-NOM, il est parfois difficile de déterminer en contexte si l'auteur a voulu inclure les arbres ou non. Pour le savoir, il est nécessaire de lire les passages qui précèdent le contexte ou le suivent. Parfois ce n'est pas suffisant et c'est au lecteur d'interpréter le sens. De même, il n'est pas toujours aisé de décider si l'emploi d'un terme est intentionnel ou non. Par exemple, dans le sous-corpus comparable français, dans la phrase suivante, l'auteur a utilisé le verbe *prédire* au lieu du verbe idiomatique *prévoir*.

CC : ... tous les modèles, et cette affirmation reste quasiment vérifiée maintenant que le nombre de simulations a été multiplié par près d'un facteur 10, PRÉDISENT un réchauffement. (CNRS\_lettre14\_2002.fr.txt)

CC :... À la fin de ce siècle, les simulations de dizaines de modèles PRÉVOIENT un réchauffement compris entre 1,4 et 5,8 °C... (Lettre\_changement\_2003)

Les exemples que nous venons de présenter ne sont pas des exceptions. Enfin, dans une traduction, il est plus facile de repérer les erreurs d'interprétation et de les critiquer, puisque nous avons accès au texte source.

#### *Les ressources terminologiques et les dictionnaires généraux*

Bien souvent, les équivalents répertoriés dans les dictionnaires spécialisés bilingues et les dictionnaires de langue générale bilingues ne concordent pas avec ceux observés dans les textes. Par exemple, pour le terme anglais *precursor*-NOM, le GDT propose *polluant précurseur*-NOM, dans les deux sous-corpus français seul l'équivalent *précurseur*-NOM est présent. Dans les textes, par économie, on passe souvent d'un terme complexe (NOM+ADJ) à un terme simple (NOM).

D'autre part, les ressources spécialisées ne s'entendent pas toujours sur le nombre de sens à attribuer à une même forme graphique. Par exemple, certaines ressources distinguent deux sens à *régime*-NOM ('régime d'un cours d'eau' et 'régime des pluies') (GDT) alors que d'autres ne font pas cette distinction ('variation d'un paramètre (température, pluie, débit d'un fleuve, etc.) au cours d'un temps donné (année par exemple)') (Dutuit et Gorenflot 2008 : 158).

Parfois, les ressources spécialisées ne s'accordent pas sur la définition d'un terme. Par exemple, certains considèrent la *biosphère-NOM* comme l'«ensemble des parties de la planète où se développe le vivant» (Dutuit et Gorenflot 2008), alors que d'autres la définissent comme l'«ensemble des vivants de la planète» (Villeneuve et Richard 2005).

Certaines ressources spécialisées tiennent pour synonymes deux termes; alors que d'autres ne le font pas. Par exemple, *boisement-NOM* «création d'une forêt sur un terrain qui n'a jamais été boisé ou ne l'est pas depuis longtemps» et *reboisement-NOM* «création d'une forêt sur un terrain précédemment boisé» sont des termes distincts si l'on tient compte du paramètre « durée » (en souligné), et sont synonymes si l'on n'en tient pas compte.

Ces observations peuvent être en partie expliquées par le fait que les notions sont envisagées sous différents angles selon les domaines ou selon le niveau de spécialisation (scientifique, technique, général). L'idéal serait de ne tenir compte que d'une ressource terminologique, préférablement sur le changement climatique. Cependant, aucune ressource spécialisée ne contient une entrée pour tous les termes analysés dans cette étude, ce qui nous force à consulter de nombreuses ressources terminologiques. Dans ces conditions, nous devons nous appuyer sur plusieurs sources, même si pour une forme graphique donnée, elles ne s'entendent pas toujours sur le nombre de ses significations, sa définition, ses synonymes ou ses équivalents.

#### *Consensus sur la validation des termes et des équivalents*

Pour les raisons évoquées ci-dessus, même une analyse manuelle ne garantit pas un consensus absolu sur la validation des termes anglais, sur leur sens et sur leurs équivalents.

#### *Une réflexion sur les outils informatiques*

Pour notre étude, l'extraction automatique des termes s'est montrée très pratique, car elle nous a permis de garder une certaine distance par rapport au choix des termes anglais. De plus, le repérage manuel de ces termes aurait été beaucoup trop chronophage. Les indications statistiques ont été utiles en ce qui concerne la fréquence, du moins pour les formes graphiques dont la partie du discours est différente. En ce qui concerne les silences, nous pensons qu'il est faible, mais nous n'avons pas fait de recherche en ce sens.

Le calcul de la répartition avec AntConc est utile, mais chaque fois que l'on a affaire à une forme graphique ayant plusieurs sens, il faut vérifier les contextes pour compter le nombre de textes dans lesquels la forme choisie se trouve (*cf.* 6.1.1).

L'aligneur de textes est indispensable, sans lui, le travail aurait été très long et fastidieux. L'extraction lexicale a donné des résultats utiles. Sur un échantillon de 70 termes anglais (environ 20 % des 343 termes), le module a repéré 280 (96,9 %) équivalents sur les 292 que nous avons validés manuellement. Les équivalents ignorés ont des basses fréquences se situant entre 4 et 8. Toutefois, pour ce type d'étude, nous aurions pu nous en passer puisque nous avons tout vérifié et revérifié manuellement. Néanmoins, il était intéressant d'avoir une liste d'équivalents sous la main avec des statistiques (fréquence des équivalents).

## **6. RÉSULTATS DE LA MÉTHODOLOGIE COMMUNE AUX DEUX VOILETS DE LA RECHERCHE**

Ce chapitre présente les résultats de la méthodologie commune aux deux volets de la recherche. Ces résultats constituent le point de départ des analyses des deux volets. Comme annoncé au Chapitre 5 (5.3.3), la section 6.1 donne une description détaillée des termes anglais. La section 6.2 explique les résultats de la validation des équivalents dans le corpus parallèle et la section 6.3 ceux de la validation des équivalents dans le corpus comparable. Enfin, la section 6.4 résume les points abordés dans les parties précédentes et apporte quelques réflexions sur l'extraction automatique et sur la validation des équivalents.

### **6.1 Description détaillée de la liste de termes anglais**

La section 5.3.3 a présenté les résultats liés à la validation des termes (critères d'identification, nombre et pourcentage des parties du discours). Dans la présente section, nous nous penchons de plus près sur les 343 termes validés et analysons leur répartition (6.1.1), leur score (6.1.2) et leur fréquence (6.1.3) à partir des données générées par l'extracteur de termes et le concordancier que nous avons utilisés, à savoir TermoStat (Drouin 2002) et AntConc (Anthony 2007). Ces analyses ne portent que sur la liste des termes validés afin de mieux interpréter leurs équivalents. Enfin, nous terminons par une conclusion à la section 6.1.4.

#### **6.1.1 La répartition**

La répartition correspond au nombre de textes du corpus dans lequel un terme donné apparaît. Pour éviter les termes propres à un trop petit nombre d'auteurs ou même à un seul, nous avons établi empiriquement de conserver les termes présents dans au moins 8 textes (*cf.* 5.3.1). La répartition est calculée sous l'onglet *Concordance Plot* du concordancier AntConc (Anthony 2007). La Figure 6.1 montre la répartition du terme anglais *biota*-NOM. La case *Concordance Hits* indique le nombre d'occurrences du terme, soit 28. Les barres

horizontales situées dans la fenêtre principale représentent les textes dans lesquels apparaît le terme, en l'occurrence 8 pour le terme *biota*-NOM.

The screenshot displays the AnctConc software interface. On the left, a list of corpus files is shown under the heading "Corpus Files". The main area displays concordance results for the search term "biota|biotas". Each result includes the number of hits and the file length in characters. The search options at the bottom are set to "Words" and "Case" is unchecked. The search term is "biota|biotas" and the number of concordance hits is 28.

File Name	HIT FILE	No. of Hits	File Length (in chars)
CC_CES_ozone_2007.en.txt	7	1	51158
CC_IPCC_bilan_2001.en.txt	17	1	280923
CC_IPCC_biodiversité_2002.en.txt	18	14	226233
CC_IPCC_eau_2008.en.txt	20	6	543159
CC_IPCC_modele_1997.en.txt	23	1	128385
CC_IPCC_synthese_2001.en.txt	27	1	228908
CC_RNCA_impacts_2008.en.txt	33	2	1298536
CC_UNESCO_patrimoine_2009.en.txt	35	2	165858

Search Term:  Words  Case  Regexp  
 Search Term: biota|biotas  
 Concordance Hits: 28  
 Start Stop

Figure 6.1. Calcul de la répartition du terme *biota* dans AnctConc

Pour les termes monosémiques, comme *tornado*-NOM, cet outil donne des résultats fiables<sup>96</sup>. Cependant, pour les termes polysémiques, comme AntConc ne distingue pas les différents sens, il ne peut donner qu'une répartition approximative. Par conséquent, pour obtenir la répartition exacte d'un terme, comme *plant*-NOM ('végétal'), il faut vérifier manuellement tous les contextes dans lesquels ce terme apparaît afin de les départager des contextes du terme *plant*-NOM ('usine'). En l'occurrence, une fois les contextes vérifiés, le sens cherché de ce terme n'est présent que dans 25 textes au lieu des 30 annoncés par le logiciel.

Le graphique (Figure 6.2) montre la répartition des 343 termes anglais validés en fonction du nombre de textes, chaque colonne représentant un nombre de textes. Nous avons tenu compte des sens différents en vérifiant manuellement les termes polysémiques. Malgré une distribution en dent de scie, on y observe sans surprise que de manière générale plus le nombre de textes est élevé, plus le nombre de termes diminue. Par exemple, dans les colonnes situées à chaque extrémité du graphique, on observe que 32 termes figurent dans 8 textes, alors que seulement 5 termes sont présents dans 38 textes. Toutefois, comme nous l'avons dit, cette observation ne se vérifie pas tout au long du graphique, par exemple, dans la colonne des 18 textes, nous avons 17 termes, alors que dans la colonne des 13 textes, il n'y en a que 7.

---

<sup>96</sup> À condition de ne pas oublier de mettre toutes les formes fléchies du terme, *tornado*, *tornados*, *tornadoes*.



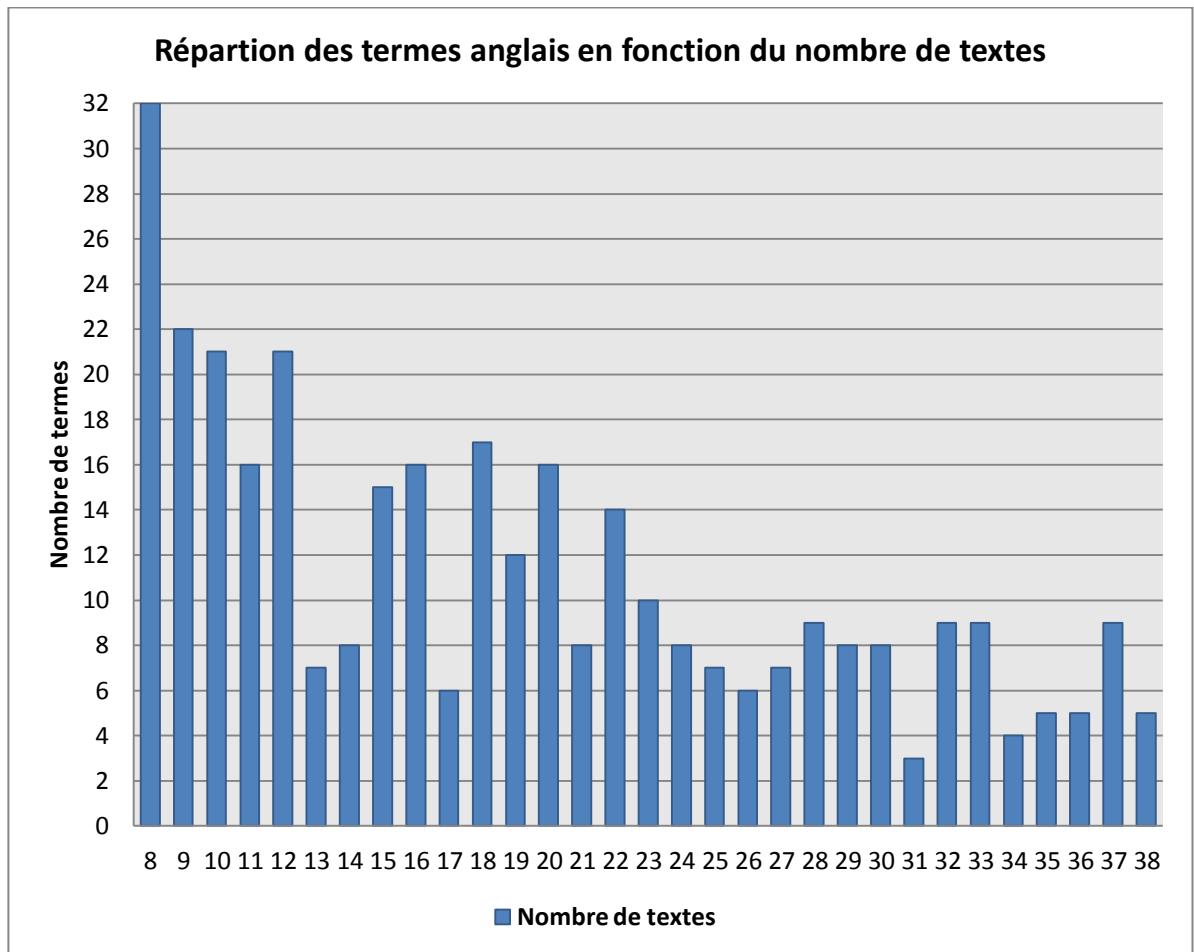


Figure 6.2. Répartition des termes anglais en fonction du nombre de textes

L'analyse des termes présents dans les 38 textes (l'ensemble du corpus) est intéressante sur plusieurs points (Tableau 6.1). Premièrement, ces termes sont très fréquents. Deuxièmement, 3 termes sur 5 ne figuraient pas parmi nos mots clés<sup>97</sup>, *increase-VER*, *impact-NOM* et *energy-NOM*. *A priori*, nous n'aurions pas choisi ces termes parce qu'ils nous auraient semblé beaucoup moins représentatifs du domaine du changement climatique que des termes comme *carbon-NOM* ou *emission-NOM*. Cependant, à bien y réfléchir, ces derniers termes ne sont pas des termes exclusifs à ce domaine. Deuxièmement, nous avons été surprise de voir qu'un verbe, en l'occurrence *increase-*

<sup>97</sup> Les mots clés tels *climate change*, *emission*, *greenhouse gases*, et *carbon* ont servi à chercher des textes du changement climatique sur Internet (cf. 5.1.1).

VER, occupe une place si importante dans le corpus. Le contexte suivant montre un exemple d'usage de ce terme dans le domaine du changement climatique :

At the same time, concentrations of carbon dioxide (CO<sub>2</sub>) [...] have INCREASED dramatically. (CE\_changclim\_2002)

La présence des termes *impact*-NOM et *energy*-NOM, quant à eux, surprend moins quand on sait qu'un nombre important des textes du corpus sont des rapports rédigés à l'intention des décideurs. Ces textes présentent la consommation d'énergie comme la principale cause anthropique des changements climatiques et informent des impacts potentiels de ces changements sur la population et les infrastructures.

Tableau 6.1. Les 5 termes présents dans les 38 textes

Terme anglais	Fréquence
<i>climate</i> -NOM	9 090
<i>change</i> -NOM	8 912
<i>increase</i> -VER	3 369
<i>impact</i> -NOM	3 276
<i>energy</i> -NOM	1 446

L'examen de la colonne des 8 textes (Tableau 6.2), celle dans laquelle nous trouvons le plus de termes validés, montre que les termes de cette catégorie sont généralement liés à des sujets moins abordés dans notre corpus (*biota*-NOM, *albedo*-NOM ou *peat*-NOM). Rien n'empêcherait qu'ils soient plus fréquents dans d'autres types de textes sur le changement climatique. Bien entendu, tous ces termes sont généralement peu fréquents, moins de 100 occurrences, sauf *capture*-NOM (182), le moins fréquent étant *latitudinal*-ADJ (15).

Tableau 6.2. Les 32 termes de la colonne 8 de la Figure 6.2

Terme anglais	Fréquence	Terme anglais	Fréquence
<i>capture-NOM</i>	182	<i>biota-NOM</i>	28
<i>flood-VER</i>	83	<i>diversion-NOM</i>	28
<i>grain-NOM</i>	73	<i>snowmelt<sub>1</sub>-NOM</i>	28
<i>variable-ADJ</i>	72	<i>snowmelt<sub>2</sub>-NOM</i>	28
<i>human-NOM</i>	71	<i>albedo-NOM</i>	27
<i>abatement-NOM</i>	69	<i>circumpolar-ADJ</i>	26
<i>streamflow-NOM</i>	67	<i>hail-NOM</i>	22
<i>halocarbon-NOM</i>	62	<i>mountainous-ADJ</i>	21
<i>sequester-VER</i>	50	<i>oxidation-NOM</i>	19
<i>troposphere-NOM</i>	48	<i>mapping-NOM</i>	18
<i>thunderstorm-NOM</i>	38	<i>peat-NOM</i>	18
<i>offshore-ADJ</i>	34	<i>dieback-NOM</i>	17
<i>infestation-NOM</i>	33	<i>phytoplankton-NOM</i>	17
<i>seasonality-NOM</i>	33	<i>lowland-NOM</i>	16
<i>carbonate-NOM</i>	31	<i>retention-NOM</i>	16
<i>tornado-NOM</i>	31	<i>latitudinal-ADJ</i>	15

Les termes présents dans 22 textes (colonne 22 de la Figure 6.2) sont au nombre de 14 (Tableau 6.3). La fréquence de ces termes varie de quelques dizaines à plus d'un millier. À part cela, ces termes ne se distinguent pas de ceux des textes de la colonne 8.

Tableau 6.3. Les 14 termes de la colonne 22 de la Figure 6.2

Terme anglais	Fréquence	Terme anglais	Fréquence
<i>ice-NOM</i>	1 018	<i>tree-NOM</i>	290
<i>river-NOM</i>	452	<i>coast-NOM</i>	241
<i>stress-NOM</i>	440	<i>feedback-NOM</i>	204
<i>ozone-NOM</i>	436	<i>rain-NOM</i>	164
<i>disaster-NOM</i>	390	<i>nitrogen-NOM</i>	141
<i>sensitivity-NOM</i>	377	<i>emit-VER</i>	103
<i>damage-NOM</i>	362	<i>mitigate-VER</i>	63

### 6.1.2 Le score

Le score est la valeur donnée par TermoStat (Drouin 2002) à chacun des candidats termes en fonction du test statistique sélectionné (le test de *spécificité* dans le cas de cette extraction). Les CT qui reçoivent un score supérieur à un certain seuil (+3,09 en l'occurrence) sont susceptibles d'être les formes les plus intéressantes et les plus importantes du corpus. Lorsque TermoStat affiche les résultats de l'extraction, le CT ayant reçu le score le plus élevé est placé en tête de liste, c'est-à-dire au premier rang, et le CT ayant reçu le score le moins élevé est mis au dernier rang. Les scores sont à relativiser, car si TermoStat peut assez bien différencier les parties du discours, il ne peut pas distinguer les différents sens d'un terme. Par exemple, *flow*-NOM peut désigner ('débit d'une rivière' 'écoulement d'une rivière'), ('montée de l'eau'), ('déplacement d'une quantité de biens'), etc. Pour obtenir le score exact d'un terme, il faudrait que TermoStat puisse distinguer les différents sens d'une même forme graphique.

Dans la liste des 4 677 CT présentés par TermoStat, *climate*-NOM, le premier terme validé par nos critères occupe le 2<sup>e</sup> rang<sup>98</sup> et le dernier terme validé, *tide*-NOM, se situe au 4 010<sup>e</sup> rang. Le nombre de termes validé par tranche de 200 CT<sup>99</sup> est indiqué dans le Tableau 6.4. Dans la tranche des premiers 200 CT, 88 termes ont été validés, soit 25,6 % de tous les termes validés.

---

<sup>98</sup> Le premier CT est *be*-VER, nous l'avons retranché lors de la validation des CT.

<sup>99</sup> Chaque tranche représente 200 CT, par exemple, dans la tranche 1-199 du Tableau 6.3, les CT vont du rang 1 au rang 199, et ainsi de suite.

Tableau 6.4. Nombre de termes validés en fonction du rang (par tranche de 200)

Tranche de 200 CT	Nombre de CT validés
1-199	88
200-399	51
400-599	41
600-799	39
800-999	29
1000-1 199	24
1200-1 399	21
1400-1 599	11
1600-1 799	7
1800-1 999	12
2000-2 199	4
2200-2 399	4
2400-2600	1
2600-2800	2
2800-3000	4
3000-3200	0
3200-3400	1
3400-3600	2
3600-3800	0
3800-4000	0
4000-4200	2
<b>Total</b>	<b>343</b>

Le graphique (Figure 6.3) permet de visualiser la courbe du nombre de termes validés par tranche de 200 CT. Du rang 1 au rang 1 399, 293 termes sont validés, soit 85,4 % de tous les termes retenus. Jusqu'au rang 1 799, la courbe descend fortement. Puis, à partir de ce point, elle suit une ligne plus ou moins horizontale.

Comme nous nous y attendions, les 5 termes présents dans tous les textes (*cf.* 6.1.1) figurent parmi les termes dont le score est le plus élevé. Si la plupart des termes validés sont situés dans les premières tranches de 200, il n'en reste pas moins que les termes validés dans les derniers rangs peuvent être tout aussi intéressants sur le plan terminologique. Par exemple, le dernier terme validé, *tide-NOM*, occupe le 4 010<sup>e</sup> rang.

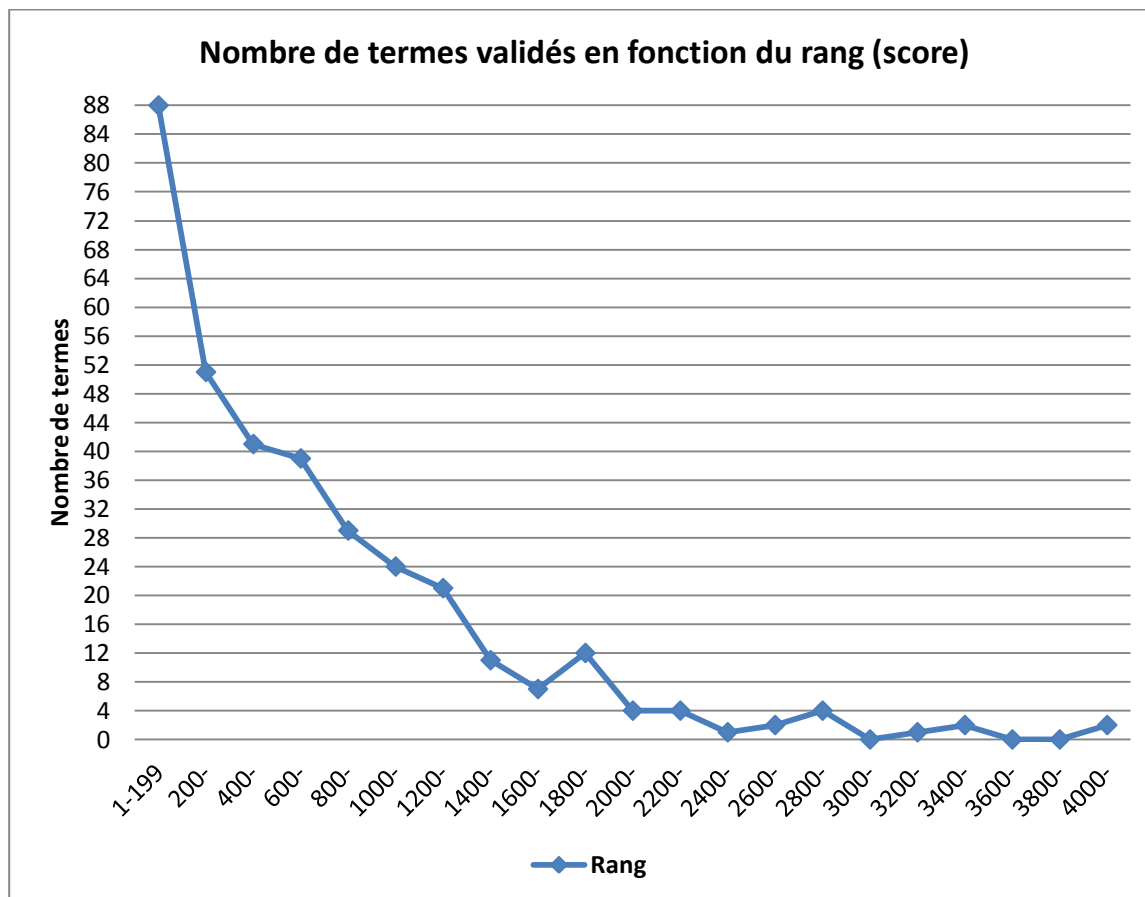


Figure 6.3. Nombre de termes validés en fonction du rang (score) attribué par TermoStat

### 6.1.3 La fréquence

La fréquence est le nombre d'occurrences d'un terme à l'intérieur d'un corpus. La fréquence a été calculée avec TermoStat. TermoStat distingue les parties du discours puisque le texte analysé est étiqueté à l'aide de TreeTagger (Schmid 1994). Cependant, comme TermoStat ne distingue pas les différents sens d'une même forme graphique (ex. *wood*-NOM ('matière ligneuse' ou 'ensemble d'arbres')), il faut relativiser la fréquence des termes polysémiques.

Le graphique de la Figure 6.4 présente la fréquence des termes validés. Les termes dont la fréquence est inférieure à 1 000 sont représentés par tranche de 100 (colonnes en gris). Les termes dont la fréquence est égale ou supérieure à 1 000 sont représentés par

tranche de 1 000 (colonnes en noir). Les termes avec une fréquence de 0 à 99 sont les plus nombreux, soit 136. Les termes dont la fréquence est supérieure à 1 000 sont peu nombreux, soit 32. Nous observons également que 200 termes ont une fréquence inférieure à 200, ce qui représente 58,3 % des termes validés.

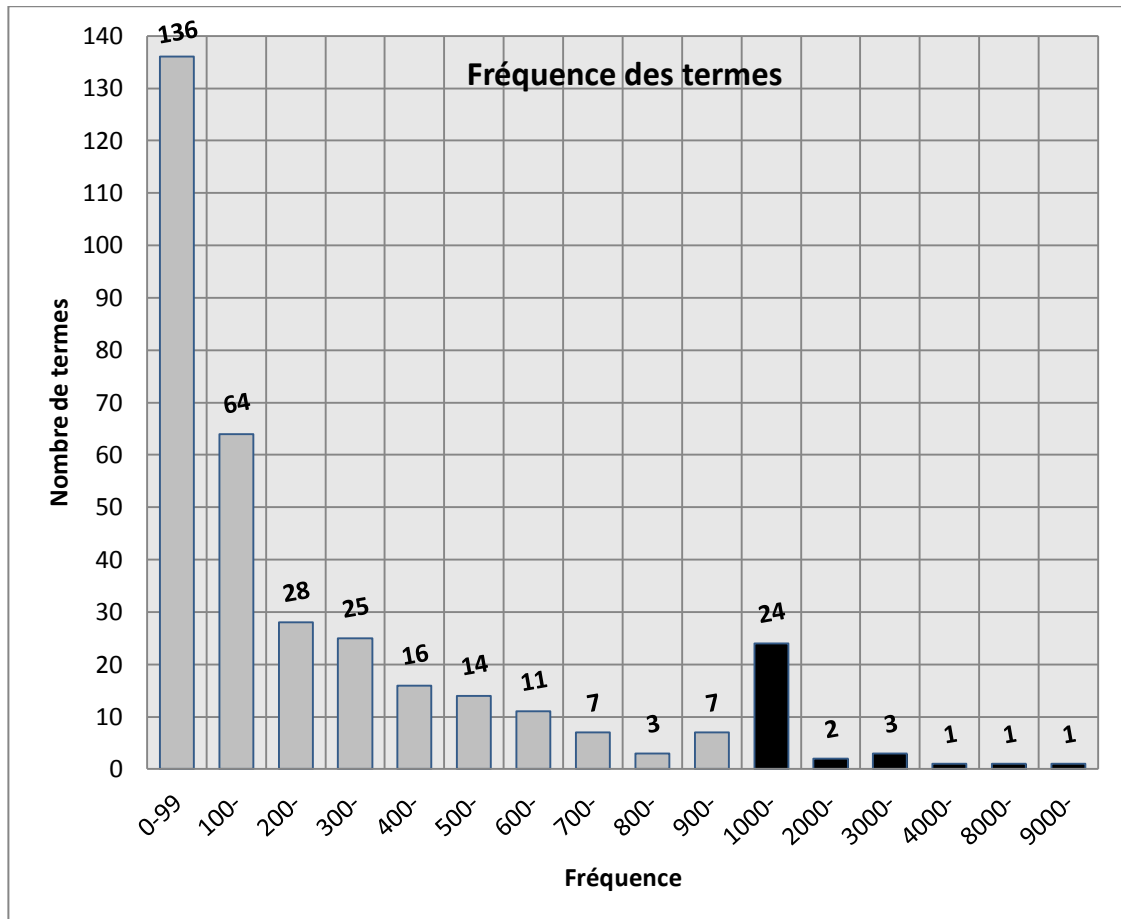


Figure 6.4. Fréquence des termes validés

L'écart entre les termes les plus fréquents et les moins fréquents peut être considérable, par exemple, *climate-NOM* est représenté 9 090 fois, alors que *dieback-NOM* n'est présent que 17 fois. Le Tableau 6.5 montre les 5 termes les plus fréquents et les 5 termes les moins fréquents de la liste des termes validés.

Tableau 6.5. Les 5 termes les plus fréquents et les 5 termes les moins fréquents

Terme anglais	Fréquence	Terme anglais	Fréquence
<i>climate</i> -NOM	9 090	<i>dieback</i> -NOM	17
<i>change</i> -NOM	8 912	<i>phytoplankton</i> -NOM	17
<i>emission</i> -NOM	4 483	<i>lowland</i> -NOM	16
<i>water</i> -NOM	3 757	<i>retention</i> -NOM	16
<i>increase</i> -VER	3 369	<i>latitudinal</i> -ADJ	15

## 6.2 Résultats de la validation des équivalents dans le corpus parallèle

Les résultats de la validation des équivalents dans le corpus parallèle sont présentés dans l'ordre suivant : données globales (6.2.1); nombre d'équivalents par terme anglais (6.2.2); recensement et analyse des équivalents simples et des équivalents complexes (6.2.3); distribution des équivalents par partie du discours (6.2.4). À cette étape, les résultats sont quantitatifs. Les données quantitatives des deux types de corpus examinées séparément dans ce chapitre seront comparées au Chapitre 7. En outre, une analyse qualitative y sera présentée.

### 6.2.1 Données globales

Sur les 343 termes anglais retenus pour notre étude, nous avons validé en tout 811 équivalents. La moyenne d'équivalents par terme anglais est de 2,4. Toutefois, parmi les équivalents, il existe de nombreux doublons, c'est-à-dire qu'un même équivalent français sert d'équivalent à plusieurs termes anglais différents, par exemple, *côte*-NOM est l'équivalent des termes anglais : *coast*-NOM, *coastal*-ADJ, *shoreline*-NOM, *coastline*-NOM et *glaciaire*-ADJ est l'équivalent de : *glacial*-ADJ, *ice*-NOM et *glacier*-NOM.

### 6.2.2 Nombre d'équivalents par termes anglais

Des 343 termes anglais, 113 ont 1 seul équivalent et 230 sont associés à 2 ou à plusieurs équivalents, jusqu'à concurrence de 9. Le Tableau 6.6 présente dans le détail le nombre de termes en fonction du nombre d'équivalents. La première colonne présente le nombre de



termes anglais, la deuxième, le nombre d'équivalents. La troisième donne des exemples de termes anglais.

Tableau 6.6. Nombre de termes par nombre d'équivalents dans le corpus parallèle

Nombre de termes anglais	Nombre d'équivalents	Exemples de termes anglais
113	1	<i>abatement-NOM, activity-NOM, carbon-NOM, dieback-NOM, hurricane-NOM, infrastructure-NOM [...]</i>
114	2	<i>concentration-NOM, forested-ADJ, runoff-NOM, scenario-NOM, water-NOM [...]</i>
54	3	<i>emit-VER, environment-NOM, flood-NOM, mining-NOM, sustainable-ADJ [...]</i>
35	4	<i>compliance-NOM, damage-NOM, extreme-ADJ, flooding-NOM, forest-NOM [...]</i>
10	5	<i>effect-NOM, hazard-NOM, human-ADJ, migration-NOM, outbreak-NOM, release-VER, rise-NOM, river-NOM, shoreline-NOM, wildlife-NOM.</i>
8	6	<i>assess-VER, decrease-VER, human-NOM, monitoring-NOM, reduce-VER, rise-VER, risk-VER, vulnerable-ADJ.</i>
4	7	<i>increase-NOM, global-ADJ, magnitude-NOM, pattern-NOM.</i>
3	8	<i>increase-VER, project-VER, globally-ADV.</i>
2	9	<i>impact-NOM, impact-VER.</i>
<b>Total 343</b>		

Afin de mieux visualiser ces résultats, la Figure 6.5 illustre les données des deux premières colonnes du Tableau 6.6. On remarque dans ce graphique que les termes ayant 1 ou 2 équivalents sont en nombre à peu près égal. Ils sont en outre nettement plus nombreux que les termes qui ont plus de 2 équivalents. Les termes ayant 3 ou 4 équivalents sont, quant à eux, encore relativement fréquents. Enfin, les termes avec 5 équivalents et plus sont beaucoup moins nombreux.

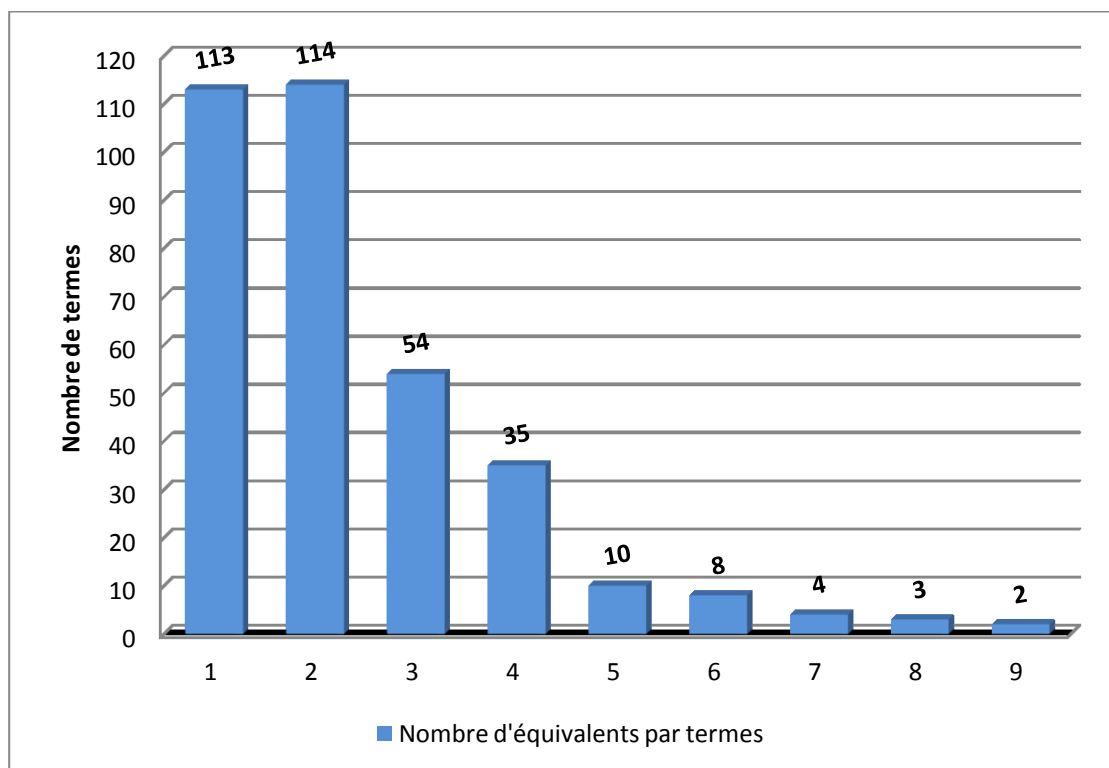


Figure 6.5. Nombre de termes en fonction du nombre d'équivalents (corpus parallèle)

### 6.2.3 Équivalents complexes et équivalents simples

Sur les 811 équivalents validés, 47 sont complexes (5,8 % du total des équivalents) et 764 simples (94,2 % du total des équivalents). Dans cette thèse, un équivalent simple est constitué d'une chaîne de caractères continue (ex. *circumpolaire-ADJ*), un équivalent complexe est constitué de plusieurs chaînes de caractères séparés soit pas des espaces, soit par des traits d'union (ex. *diversité biologique-NOM*, *non-viabilité-NOM*).

Dans le Tableau D.1 de l'Annexe D, on trouvera la liste des équivalents termes complexes. Il est à noter que les équivalents complexes ou simples ne sont pas tous des équivalents typiques (cf. 5.5.2). En effet, sur les 47 termes complexes, nous avons recensé 32 équivalents typiques (ÉqT) et 15 équivalents non typiques (ÉqNT). Par exemple, les équivalents *hydrocarbure halogéné-NOM* ou *halocarbure-NOM* sont des ÉqT du terme anglais *halocarbon-NOM*. L'équivalent *nappe phréatique-NOM* est un ÉqNT du terme

anglais *groundwater*-NOM. Les termes simples, quant à eux, comptent 371 ÉqT et 393 ÉqNT. Dans le Tableau 6.7, nous constatons que les termes complexes sont plus fréquemment des ÉqT que des ÉqNT et que c'est l'inverse pour les termes simples.

Tableau 6.7. Répartition des équivalents par ÉqT et ÉqNT dans le corpus parallèle

Type d'équivalent	Nombre	Pourcentage
Terme complexe ÉqT	32	3,9
Terme complexe ÉqNT	15	1,9
Terme simple ÉqT	371	45,7
Terme simple ÉqNT	393	48,5
<b>Total</b>	<b>811</b>	<b>100</b>

#### 6.2.4 Distribution des équivalents par partie du discours

Les 811 équivalents sont distribués de la façon suivante : 532 noms, 165 adjectifs, 103 verbes, 7 adverbes, 4 locutions ou expressions (ex. *en mer, dans le monde*) (Tableau 6.8). Le nombre de noms est largement supérieur aux autres parties du discours. Les pourcentages par parties du discours des équivalents ne présentent pas des différences très marquées des pourcentages des termes anglais (Tableau 6.8).

Tableau 6.8. Nombre d'équivalents dans le corpus parallèle par partie du discours et en pourcentage

Partie du discours	Nombre	Pourcentage d'équivalents par partie du discours	Pourcentage des termes anglais validés par partie du discours
Nom	532	65,6	73,8
Adjectif	165	20,3	15,2
Verbe	103	12,7	9
Adverbe	7	0,9	2
Autre	4	0,5	0
<b>Total</b>	<b>811</b>	<b>100</b>	<b>100</b>

## **6.3 Résultats de la validation des équivalents du corpus comparable**

Les résultats de la validation des équivalents dans le corpus comparable sont présentés dans le même ordre que les résultats de la validation dans le corpus parallèle. Premièrement, nous présentons les données globales (6.3.1). Deuxièmement, nous donnons le nombre d'équivalents par terme anglais (6.3.2). Troisièmement, nous faisons le recensement et l'analyse des équivalents simples et des équivalents complexes (6.3.3). Enfin, nous terminons avec la distribution des équivalents par partie du discours (6.3.4).

### **6.3.1 Données globales**

Sur les 343 termes, nous avons validé en tout 848 équivalents. La moyenne d'équivalents par terme anglais est de 2,5. Toutefois, parmi les équivalents, il existe de nombreux doublons, c'est-à-dire qu'un même équivalent français sert d'équivalent à plusieurs termes anglais, par exemple, *atmosphère*-NOM est l'équivalent des termes anglais : *atmosphere*-NOM, *atmospheric*-ADJ, *air*-NOM.

### **6.3.2 Nombre d'équivalents par termes anglais**

Sur les 343 termes anglais, 4 ont 0 équivalent, 109 ont 1 seul équivalent et 230 sont associés à 2 ou à plusieurs équivalents, jusqu'à concurrence de 10. Le Tableau 6.9 présente dans le détail le nombre de termes en fonction du nombre d'équivalents. La première colonne présente le nombre de termes anglais, la deuxième, le nombre d'équivalents. La troisième donne des exemples de termes anglais.

Tableau 6.9. Résultats de la validation des équivalents dans le corpus comparable

Nombre de termes anglais	Nombre d'équivalents	Exemples de termes anglais
4	0	<i>biome-NOM, biota-NOM, diversion-NOM, unsustainable-ADJ.</i>
109	1	<i>acidification-NOM, afforestation-NOM, albedo-NOM, aquatic-ADJ, biomass-NOM [...]</i>
108	2	<i>adaptation-NOM, absorb-VER, atmosphere-NOM, climate-NOM, climatic-ADJ [...]</i>
55	3	<i>adaptive-ADJ, glacial-ADJ, intensity-NOM, marine-ADJ, model-VER [...]</i>
34	4	<i>consequence-NOM, population-NOM, weather-NOM, assessment-NOM, change-NOM, [...]</i>
9	5	<i>hazard-NOM, monitoring-NOM, release-VER, rise-NOM, river-NOM [...]</i>
10	6	<i>assess-VER, decrease-VER, human-ADJ, human-NOM, livestock-NOM, magnitude-NOM, reduce-VER, rise-VER, shoreline-NOM, vulnerable-ADJ.</i>
5	7	<i>global-ADJ, increase-NOM, pattern-NOM, risk-NOM, stress-NOM.</i>
3	8	<i>globally-ADV, increase-VER, project-VER.</i>
4	9	<i>capture-NOM, capture-VER, impact-NOM, impact-VER.</i>
2	10	<i>sequester-NOM, sequestration-NOM.</i>
<b>Total 343</b>		

Pour mieux interpréter ces résultats, la Figure 6.6 illustre les données des deux premières colonnes du Tableau 6.9, soit le nombre de termes anglais et le nombre d'équivalents par terme. Le graphique affiche deux colonnes de plus que celui du corpus parallèle (*cf.* Figure 6.5), la colonne des 0 équivalents<sup>100</sup> et la colonne des 10 équivalents. Pour les autres colonnes, les résultats suivent sensiblement les mêmes tendances que dans le corpus parallèle.

<sup>100</sup> Les termes pour lesquels il n'y a pas d'équivalent sont décrits à la section 7.2.1.

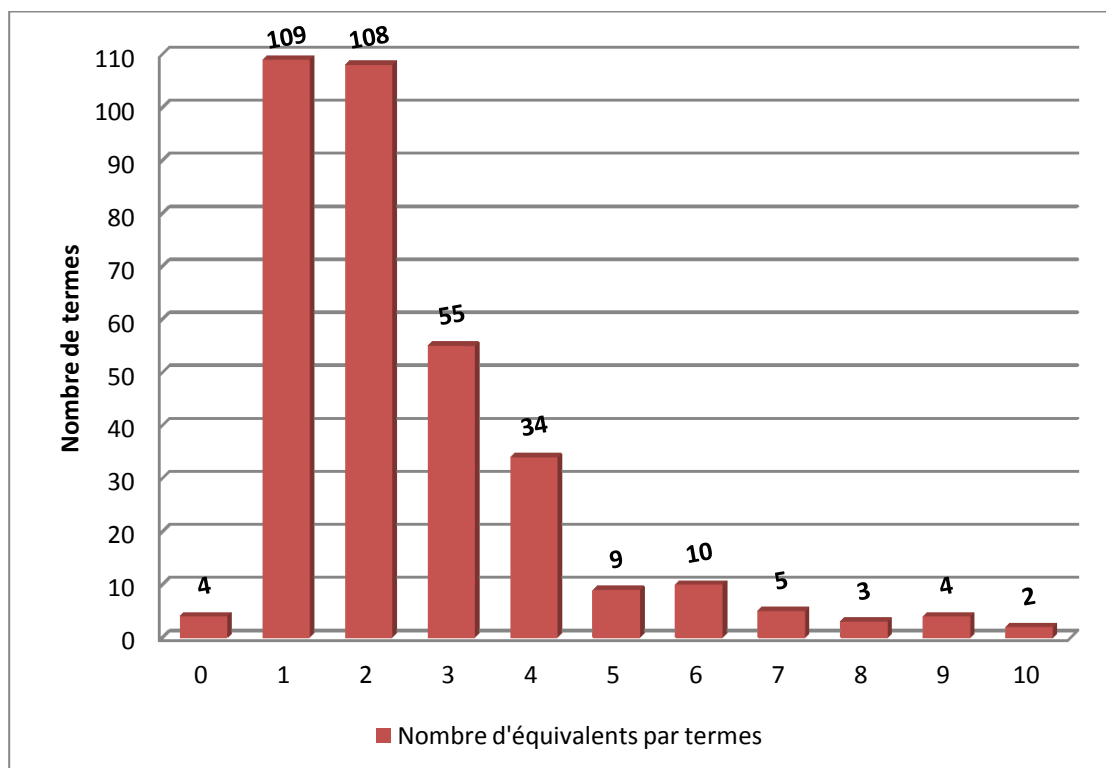


Figure 6.6. Nombre de termes en fonction du nombre d'équivalents (corpus comparable)

### 6.3.3 Équivalents complexes et équivalents simples

Sachant qu'un terme simple peut être rendu par un équivalent terme complexe (*cf.* 2.3.3), nous nous sommes interrogée sur l'importance du phénomène dans notre corpus comparable.

Sur les 848 équivalents validés, 43 sont complexes (5,1 % du total des équivalents) et 805, simples (94,9 % du total des équivalents). Dans le Tableau D.2 de l'Annexe D, on trouvera la liste des équivalents termes complexes du corpus comparable. Comme pour le corpus parallèle, ces équivalents ne sont pas tous des équivalents typiques (*cf.* 5.52). Sur les 43 équivalents complexes, 25 sont des équivalents typiques (ÉqT) et 18 sont des équivalents non typiques (ÉqNT) et sur les 805 équivalents 379 sont des ÉqT et 426, des ÉqNT. Dans le Tableau 6.10, comme dans le corpus parallèle (*cf.* 6.2.3), nous constatons que chez les termes complexes le pourcentage d'ÉqT est supérieur à celui des ÉqNT, alors que c'est l'inverse pour les termes simples.

Tableau 6.10. Répartition des équivalents par ÉqT et ÉqNT dans le corpus comparable

Type d'équivalent	Nombre	Pourcentage
Terme complexe ÉqT	25	3
Terme complexe ÉqNT	18	2,1
Terme simple ÉqT	379	44,7
Terme simple ÉqNT	426	50,2
<b>Total</b>	<b>848</b>	<b>100</b>

### 6.3.4 Distribution des équivalents par partie du discours

Les 848 équivalents se répartissent en 560 noms, 162 adjectifs, 117 verbes, 6 adverbes et 3 locutions ou expressions (ex. *au large*, *à venir*) (Tableau 6.11). Ici encore le nombre de noms est largement supérieur aux autres parties du discours. De plus, comme dans le corpus parallèle, les pourcentages par parties du discours des équivalents suivent les mêmes tendances que les pourcentages des termes anglais (Tableau 6.11).

Tableau 6.11. Pourcentage en parties du discours des équivalents

Partie du discours	Nombre	Pourcentage d'équivalents par partie du discours	Pourcentage des termes anglais validés par partie du discours
Nom	560	66	73,8
Adjectif	162	19,1	15,2
Verbe	117	13,8	9
Adverbe	6	0,7	2
Autre	3	0,4	0
<b>Total</b>	<b>848</b>	<b>100</b>	<b>100</b>

## 6.4 Conclusion du chapitre

Dans ce chapitre, nous avons présenté les résultats touchant l'extraction des termes anglais (6.1), la validation des équivalents dans le corpus parallèle (6.2) et la validation des équivalents dans le corpus comparable (6.3). Dans les sections qui suivent, nous formulons plusieurs observations concernant les résultats présentés. Les résultats touchant les

ressemblances et les différences quantitatives et qualitatives entre les deux corpus sont présentés, quant à eux, dans le Chapitre 7.

#### *Les résultats statistiques des termes anglais*

L'utilisation des outils automatiques en extraction de termes est indispensable dans ce type d'étude, car cela nous permet d'avoir un portrait relativement fiable de la terminologie du corpus, d'avoir rapidement accès aux termes et de rester neutre vis-à-vis de la sélection des CT à valider. Néanmoins, les résultats statistiques obtenus en calculant la répartition, le score ou la fréquence ne sont pas toujours aussi précis qu'on le souhaiterait. La cause de ce problème est largement due à la polysémie des termes. Ce n'est qu'après une vérification manuelle en contexte qu'il est possible de tirer des conclusions plus justes. En d'autres termes, les informations statistiques proposent des indices, des pistes utiles qui permettent de se faire une idée générale du comportement des termes en corpus, mais elles ne sont en aucun cas des informations fiables à 100 %<sup>101</sup>.

En consultant les résultats présentés dans les graphiques (Figures 6.2, 6.3 et 6.4), on pourrait être tenté de tirer des conclusions hâtives sur le comportement général des termes (ex. plus des termes sont présents dans un grand nombre de textes d'un corpus, plus ils sont importants pour ce domaine). Il est important de nuancer ce genre d'observation et d'éviter les généralisations, car dans un domaine donné, les termes ayant une fréquence, une répartition ou un score peu élevé ne signifient pas qu'ils soient moins significatifs que les autres sur le plan terminologique.

Lorsqu'on change le type de tri (répartition, score, fréquence), les termes jouent à la chaise musicale. Par exemple, lors du tri par fréquence, *water-NOM* arrive au 4<sup>e</sup> rang; par score, au 7<sup>e</sup> rang; et par répartition, dans la colonne des 33 textes. Cependant, quel que soit le type de tri, les termes restent en général plus ou moins au même niveau. Ainsi, le terme *tide-NOM*, lorsque trié par fréquence, se trouve au 326<sup>e</sup> rang, au tri par score, au 343<sup>e</sup>, et au tri par répartition, dans la colonne des 8 textes (le minimum de textes dans lequel un terme peut apparaître dans notre analyse).

---

<sup>101</sup> Nous avons déjà signalé (*cf.* 5.7) que même une analyse manuelle ne garantit pas un consensus absolu.



*Au sujet du degré de précision de l'analyse des équivalents*

Il est important de signaler que la fréquence d'un équivalent joue un rôle important dans son analyse (ex. *climatique*-ADJ est présent 12 120 fois dans les deux corpus alors que *tornado*-NOM n'y apparaît que 36 fois). Les équivalents très fréquents permettent une analyse beaucoup plus fine que les équivalents peu fréquents, mais beaucoup plus longue à réaliser.

*Au sujet des équivalents validés*

La validation des équivalents nous montre qu'une très grande majorité des termes anglais sont associés à plusieurs équivalents (67,1 %). Par conséquent, il nous semble important que cette caractéristique soit prise en compte dans des domaines d'application telle l'extraction bilingue de termes.

Le pourcentage des équivalents complexes varie entre 5 et 6 % de l'ensemble des équivalents. De plus, les équivalents complexes comptent en moyenne moins d'ÉqNT que les équivalents termes simples.

La répartition des équivalents par partie du discours confirme sans surprise la prépondérance des noms. D'un autre côté, nous avons remarqué la faible représentativité des adverbes, surtout parmi les équivalents. En effet, le pourcentage des équivalents adverbes est moins élevé dans les deux corpus que celui des termes anglais adverbe. Cette constatation nous portera à analyser les équivalents des adverbes plus en détail dans le deuxième volet de la recherche (*cf.* 8.3).

## **PREMIER VOLET :**

ANALYSE COMPARATIVE DE L'ÉQUIVALENCE  
TERMINOLOGIQUE EN CORPUS PARALLÈLE ET  
EN CORPUS COMPARABLE

## **7 MÉTHODOLOGIE ET RÉSULTATS DE L'ANALYSE COMPARATIVE DE L'ÉQUIVALENCE EN CORPUS PARALLÈLE ET EN CORPUS COMPARABLE**

Dans ce chapitre, nous effectuons l'analyse comparative de l'équivalence dans les deux types de corpus pour vérifier si l'équivalence terminologique observable dans les corpus parallèles se distingue significativement de celle que l'on trouve dans les corpus comparables afin de répondre à la première question de notre projet de recherche. À la section 7.1, nous décrivons la méthodologie élaborée pour procéder à l'analyse comparative. La section 7.2 est consacrée aux résultats de l'analyse comparative quantitative et qualitative des équivalents dans les deux corpus et se divise comme suit : 7.2.1 analyse comparative quantitative générale; 7.2.2 analyse comparative des équivalents identiques dans les deux corpus; 7.2.3 analyse comparative des équivalents différents dans les deux corpus. Enfin, à la section 7.3, nous reprenons les points saillants de l'analyse et les commentons.

### **7.1 Méthodologie de l'analyse comparative des équivalents du corpus parallèle aux équivalents du corpus comparable**

L'analyse comparative des équivalents consiste à comparer les équivalents français identifiés dans le corpus parallèle aux équivalents français validés dans le corpus comparable (Figure 7.1).

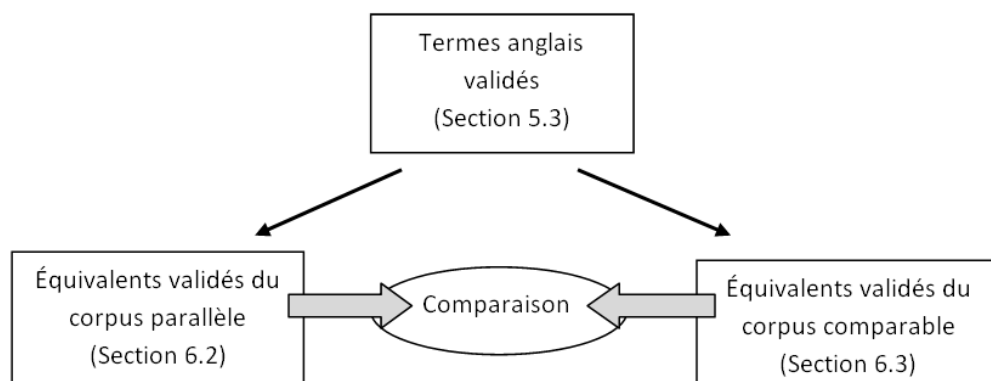


Figure 7.1. Analyse comparative des équivalents identifiés dans les deux corpus

La comparaison des équivalents des deux corpus s'effectue manuellement et à l'aide de concordanciers. À cette étape, nous nous servons des données colligées aux sections 6.2 et 6.3 du Chapitre 6 (*Résultats de la méthodologie commune aux deux volets de la recherche*) afin d'analyser les similitudes et les différences que présente l'équivalence dans les deux types de corpus et d'en donner des résultats quantitatifs et qualitatifs à la section 7.2. L'analyse se répartit de la façon suivante : 1) recensement et analyse des termes anglais présentant des équivalents identiques dans les deux corpus; 2) recensement et analyse des termes anglais comportant des équivalents différents dans les deux corpus.

### 7.1.1 Recensement des termes anglais ayant des équivalents identiques dans les deux corpus

Pour recenser les termes anglais ayant les mêmes équivalents dans les deux corpus, nous avons réuni les résultats recueillis aux sections 6.2 et 6.3 dans une même grille (le Tableau 7.1 présente un échantillon de cette grille). Le recensement des termes anglais ayant des équivalents identiques dans les deux corpus se calcule à partir de la colonne *Nombre*. Pour qu'un terme soit qualifié de terme anglais ayant des équivalents identiques, il faut que les équivalents dans le corpus parallèle (CP) et le corpus comparable (CC) soient semblables par la forme et de nombre égal. Par exemple, les termes anglais *tropospheric-ADJ* et *glacial-ADJ* comportent des équivalents identiques dans les deux corpus (identifiés

par le signe égal (=) dans la colonne *Nombre*). Ces équivalents peuvent être soit des équivalents typiques, soit des équivalents non typiques, par exemple *glaciaire-ADJ* est un ÉqT du terme anglais *glacial-ADJ*, alors que *glacier-NOM* et *glace-NOM* sont des ÉqNT. Le tableau de l'Annexe E contient la liste complète des termes ayant des équivalents identiques dans les deux corpus.

Tableau 7.1. Cinq termes anglais et leurs équivalents dans le CP et le CC

Terme anglais	Équivalent CP	ÉqT/ ÉqNT	Nombre	Équivalent CC	ÉqT/ ÉqNT
<i>tropospheric-ADJ</i>	<i>troposphérique-ADJ</i> <i>troposphère-NOM</i>	ÉqT ÉqNT	2 = 2	<i>troposphérique-ADJ</i> <i>troposphère-NOM</i>	ÉqT ÉqNT
<i>glacial-ADJ</i>	<i>glaciaire-ADJ</i> <i>glacier-NOM</i> <i>glace-NOM</i>	ÉqT ÉqNT ÉqNT	3 = 3	<i>glaciaire-ADJ</i> <i>glacier-NOM</i> <i>glace-NOM</i>	ÉqT ÉqNT ÉqNT
<i>coal-NOM</i>	<i>charbon-NOM</i> <i>houille-NOM</i> <b><i>houiller-ADJ</i></b>	ÉqT ÉqT ÉqNT	3 ≠ 3	<i>charbon-NOM</i> <i>houille-NOM</i> <b><i>charbonnier-ADJ</i></b>	ÉqT ÉqT ÉqNT
<i>ice-NOM</i>	<i>glace-NOM</i> <i>glaciaire-ADJ</i> <b><i>glaciel-ADJ</i></b>	ÉqT ÉqNT ÉqNT	3 ≠ 2	<i>glace-NOM</i> <i>glaciaire-ADJ</i>	ÉqT ÉqNT
<i>sequester-VER</i>	<i>piéger-VER</i> <i>fixer-VER</i> <i>piégeage-NOM</i> <b><i>séquestrer-VER</i></b>	ÉqT ÉqT ÉqNT ÉqNT	4 ≠ 5	<i>piéger-VER</i> <i>fixer-VER</i> <i>piégeage-NOM</i> <b><i>capturer-VER</i></b> <b><i>fixation-NOM</i></b>	ÉqT ÉqT ÉqNT ÉqNT ÉqNT

### 7.1.2 Recensement et analyse des termes anglais ayant des différences d'équivalence dans les deux corpus

Le recensement des termes anglais ayant des différences d'équivalence dans les deux corpus s'effectue à partir de la même grille que décrite à la section 7.1.1. Le signe différent (≠) de la colonne *Nombre* du Tableau 7.1 identifie les termes anglais, dont les équivalents dans le CP et le CC présentent des différences.

Pour affiner l'analyse, en plus du décompte des termes ayant des différences d'équivalence, nous avons réalisé une analyse approfondie des équivalents différents en quatre étapes :

1. collecte d'informations sur les équivalents différents (7.1.2.1);
2. analyse des équivalents différents selon des critères définis à la section 7.1.2.2;
3. notation des termes anglais ayant des équivalents différents (7.1.2.3);
4. classification des termes anglais en fonction du type de différence d'équivalence (7.1.2.4).

### 7.1.2.1 Collecte d'informations sur les équivalents différents

La collecte d'information consiste à relever deux types de renseignements sur les équivalents différents : 1) équivalent typique (ÉqT) ou équivalent non typique (ÉqNT); 2) équivalent fréquent (ÉqF) ou équivalent peu fréquent (ÉqPF).

1) Les équivalents différents peuvent être soit non typiques soit typiques. À la section 5.5.2 du Chapitre 5, nous avons défini l'équivalent typique comme celui qui est attesté par les ressources terminologiques et dictionnaires, alors que l'équivalent non typique ne l'est pas. Les équivalents *région boisée*-NOM et *surface boisée*-NOM du terme anglais *woodland*-NOM (Tableau 7.2) sont des exemples d'ÉqT, alors que l'équivalent *préserver*-VER du terme anglais *preservation*-NOM (Tableau 7.3) est un exemple d'ÉqNT. Ces renseignements sont utiles lors de l'analyse des équivalents différents.

Tableau 7.2. Exemples d'ÉqT (*région boisée*-ADJ et *surface boisée*-NOM)

Terme anglais <i>woodland</i> -NOM	
<b>Équivalents analysés</b> : <i>région boisée</i> -NOM (ÉqT, ÉqPF) et <i>surface boisée</i> -NOM (ÉqT, ÉqPF)	
<b>Équivalent CP</b> <i>région boisée</i> -NOM (5)	<b>Équivalent CC</b> <i>surface boisée</i> -NOM (6)
Ces équivalents du CP et du CC sont peu fréquents. Les deux équivalents du terme anglais diffèrent par la tête du terme ( <i>région</i> => <i>surface</i> ) et ne s'emploieront pas toujours dans les mêmes contextes, mais ils sont synonymes selon IATE. Dans les dictionnaires, le terme <i>woodland</i> -NOM peut être rendu par plusieurs autres équivalents : <i>zone boisée</i> , <i>terre boisée</i> , <i>terrain boisé</i> , <i>zone arborée</i> , etc. (Termium ou GDT).	

Tableau 7.3. Exemple d'ÉqNT (*préserver-VER*)

Terme anglais <i>preservation-NOM</i>	
<b>Équivalent analysé</b> : <i>préserver-VER</i> (ÉqNT)	
<b>Équivalent CP</b> <i>préservation-NOM</i>	<b>Équivalent CC</b> <i>préservation-NOM</i> <i>préserver-VER</i>
L'équivalent <i>préserver-VER</i> est un ÉqNT transcatégoriel du terme anglais.	

2) Le deuxième renseignement que nous relevons touche la fréquence de l'équivalent différent (ÉqPF ou ÉqF). Nous nous attardons particulièrement sur les équivalents dont la fréquence est très basse ou très élevée. Dans notre corpus, de façon empirique, nous considérons les équivalents de 15 occurrences et moins comme peu fréquents (ÉqPF) et les équivalents de 60 occurrences et plus comme des équivalents fréquents (ÉqF). Par exemple, au Tableau 7.4, l'équivalent *tourbeux-ADJ* du CP est un ÉqPF. Par contre, dans le Tableau 7.5, l'équivalent *capture-NOM* est un ÉqF. La prise en considération de la fréquence d'un équivalent nous permet de comparer son importance par rapport aux autres équivalents d'un même terme anglais.

Tableau 7.4. Exemple d'ÉqPF (*tourbeux-ADJ*)

Terme anglais <i>peat-NOM</i>	
<b>Équivalent analysé</b> : <i>tourbeux-ADJ</i> (ÉqNT, ÉqPF)	
<b>Équivalent CP</b> <i>tourbe-NOM</i> <i>tourbeux-ADJ</i> (5)	<b>Équivalent CC</b> <i>tourbe-NOM</i>
L'équivalent <i>tourbeux-ADJ</i> du CP est à la fois un ÉqNT et un ÉqPF. Toutefois, cet équivalent est morphologiquement relié à l'équivalent typique et peut le remplacer dans certains contextes comme l'illustre le deuxième exemple ci-dessous.	
CP : Large amounts of carbon are stored in PEAT soils worldwide...	
CP : De grandes quantités de carbone sont stockées dans les sols TOURBEUX du monde entier... (PNUE_bioseq_2009)	

Tableau 7.5. Exemple d'ÉqF (*capture*-NOM)

<b>Terme anglais</b> <i>capture</i> -NOM	
<b>Équivalent analysé</b> : <i>capture</i> -NOM (ÉqT, ÉqF)	
<b>Équivalent CP</b>	<b>Équivalent CC</b>
<i>piégeage</i> -NOM (145)	<i>capture</i> -NOM (120)
<i>piéger</i> -VER (14)	<i>capturer</i> -VER (8)
	<i>captage</i> -NOM (11)
	<i>capter</i> -VER (6)
	<i>piégeage</i> -NOM (6)
	<i>piéger</i> -VER (4)
<p>Comme en témoignent les chiffres, la fréquence du cognat <i>capture</i>-NOM domine largement dans le CC, les autres équivalents étant très peu fréquents. Dans le CP, cet équivalent est absent. <i>Captage</i>-NOM est le terme recommandé dans le Journal officiel de la République française. Même si tous ces équivalents sont en usage et acceptables (Brodhag et Breuil 2009 : 22) (voir contextes ci-dessous), il n'en demeure pas moins que l'équivalent <i>capture</i>-NOM est un cognat utilisé dans de nombreux textes du CC, alors qu'il ne l'est pas dans le CP.</p> <p>CC : ... la CAPTURE et le stockage du CO<sub>2</sub>... (Campos_captage-et-stockage_2007)</p> <p>CP : ... PIÉGEAGE et stockage du carbone... (IPCC_eau_2008)</p>	

### 7.1.2.2 Analyse des équivalents différents

Au Chapitre 3, nous avons souligné que l'emploi des textes traduits est critiqué par certains auteurs. D'après ces auteurs, les traducteurs sous l'influence de la langue source produiraient des textes non idiomatiques. Au niveau lexical ou terminologique, cette influence se manifesterait dans les traductions notamment sous la forme d'emprunts fautifs, de faux sens ou de cognats dont la fréquence est plus élevée que la normale. Par conséquent, pour analyser les équivalents des termes anglais ayant des équivalents différents, nous allons nous appuyer sur ces trois critères (emprunt fautif, faux sens, cognat) que nous décrivons ci-après. Les emprunts et les faux sens sont attestés comme tels dans les ressources terminologiques et autres dictionnaires de l'Annexe C.

Les emprunts fautifs sont des mots appartenant à une autre langue, en l'occurrence l'anglais. Le Tableau 7.6 montre un exemple d'emprunt fautif<sup>102</sup> trouvé dans le CC

<sup>102</sup> Les emprunts à l'anglais sont aussi appelés *anglicismes*.



(*mitigation*-NOM) et le Tableau 7.7 présente deux exemples trouvés dans le CP (*anthropogène*-ADJ, *anthropogénique*-ADJ).

Tableau 7.6. Exemple d'emprunt fautif trouvé dans le CC (*mitigation*-NOM)

Termes anglais <i>mitigation</i> -NOM	
<b>Équivalent CP</b> <i>atténuer</i> -VER <i>atténuation</i> -NOM	<b>Équivalent CC</b> <i>atténuation</i> -NOM <b><i>mitigation</i>-NOM (35)</b> <i>atténuer</i> -VER <i>mitiger</i> -VER
<p>Dans Termium, le terme <i>mitigation</i>-NOM est un anglicisme. La Commission générale de terminologie et de néologie (France) l'accepte seulement dans le domaine des sciences et des technologies spatiales, domaine qui n'est pas relié aux changements climatiques<sup>103</sup>. L'équivalent typique du terme anglais est <i>atténuation</i>-NOM. Le premier contexte appartient au CC, le deuxième au CP.</p> <p>CC : ...pour débattre autour d'interrogations diverses portant sur le changement climatique et les solutions de MITIGATION des émissions de CO2... (Campos_captage-et-stockage_2007)</p> <p>CP : ...Ainsi, l'ATTÉNUATION des émissions de ces autres gaz est un élément important... (IPCC_stabilisationdesgaz_1997)</p>	

Tableau 7.7. Exemples d'emprunts fautifs trouvés dans le CP (*anthropogène*-ADJ, *anthropogénique*-ADJ)

Terme anglais <i>anthropogenic</i> -ADJ	
<b>Équivalent CP</b> <i>anthropique</i> -ADJ (230) <i>humain</i> -ADJ <i>anthropogène</i> -ADJ (11) <i>anthropogénique</i> -ADJ (4)	<b>Équivalent CC</b> <i>anthropique</i> -ADJ (233) <i>humain</i> -ADJ <i>homme</i> -NOM
<p>Les dictionnaires ne s'entendent pas toujours sur les équivalents du terme anglais <i>anthropogenic</i>-ADJ ('provoqué directement ou indirectement par l'action de l'homme'). IATE propose <i>anthropogénique</i>-ADJ, <i>anthropique</i>-ADJ et <i>anthropogène</i>-ADJ, ceux-là mêmes trouvés dans le CP. Toutefois, si l'on se fie à Termium, le seul équivalent valable est <i>anthropique</i>-ADJ et les deux autres sont des anglicismes à éviter. D'autre part, <i>anthropique</i>-ADJ est de loin le plus fréquent dans le CP et c'est le seul qui soit attesté dans Brodhag et Breuil (2009), Bioret <i>et al.</i> (2009) ou Dutuit et Gorenflot (2008) – tous des sources européennes. Nous considérons donc que les équivalents <i>anthropogène</i>-ADJ et <i>anthropogénique</i>-ADJ sont à éviter.</p>	

<sup>103</sup> Il est à noter que depuis le 13 juillet 2012, FranceTerme accepte le terme *mitigation* dans le domaine de l'environnement (rappelons que le texte le plus récent du corpus date de 2009).

Les faux sens sont des fautes consistant à attribuer à un mot un sens erroné. Le Tableau 7.8 illustre un exemple de faux sens repéré dans le corpus parallèle, soit l'équivalent *refléter-VER*.

Tableau 7.8. Exemple de faux sens (*refléter-VER*)

<b>Terme anglais</b> <i>reflect-VER</i>	
<b>Équivalent CP</b> <i>réfléchir-VER</i> <i>refléter-VER</i> (6)	<b>Équivalent CC</b> <i>réfléchir-VER</i> <i>renvoyer-VER</i> (6)
<p>Le verbe anglais <i>reflect-VER</i> lorsqu'il est employé dans le sens 'renvoyer de la lumière, de la chaleur ou du son' à pour équivalent <i>réfléchir-VER</i> ou <i>renvoyer-VER</i>, mais pas <i>refléter-VER</i> ('réfléchir de façon atténuée et plus ou moins vague') (Van Roey <i>et al.</i> 1991).</p> <p>CP : Once in the atmosphere, thes gases trap and REFLECT heat back toward the Earth's surface...</p> <p>CP : Une fois dans l'atmosphère, ces gaz piègent et REFLÈTENT la chaleur vers la surface de la Terre... (RNCA_adaptation_2004)</p> <p>CP :... clouds (which both REFLECT sunlight and absorb outgoing heat radiation)...</p> <p>CP :... des nuages (qui RÉFLÉCHISSENT la lumière du Soleil et absorbent le rayonnement thermique réfléché)... (RNCA_impacts_2008)</p>	

Le recours aux cognats n'est pas une faute en soi. Il est même inévitable entre l'anglais et le français dans certains cas. Dans cette thèse, les cognats critiquables doivent remplir les deux caractéristiques suivantes : 1) être synonymes d'équivalents ayant une graphie différente de celle du terme anglais; 2) posséder une fréquence qui se démarque dans un corpus par rapport à l'autre. Pour illustrer cet énoncé, prenons l'exemple de *permafrost-NOM* (Tableau 7.9). Selon le premier point, cet équivalent peut être remplacé avantageusement par les équivalents *pergélisol-NOM* ou *permagel-NOM*. En outre, selon le deuxième point, *permafrost-NOM* est non seulement plus fréquent dans le CC (87) que dans le CP (16), mais l'équivalent *pergélisol-NOM* est beaucoup plus fréquent dans le CP (308) que dans le CC (96). Dans le cadre de notre analyse, l'équivalent *permafrost-NOM*<sup>104</sup> est donc considéré comme un cognat critiquable dans le CC.

<sup>104</sup> Nous aurions pu classer l'équivalent *permafrost-NOM* parmi les emprunts fautifs, mais étant donné que cet équivalent est accepté dans certaines parties de la francophonie, nous l'avons classé parmi les cognats.

Tableau 7.9. Exemple de cognat critiquable (*permafrost*-NOM)

Termes anglais <i>permafrost</i> -NOM	
<b>Équivalent CP</b> <i>pergélisol</i> -NOM (308) <b><i>permafrost</i>-NOM (16)</b> <i>pergélisolé</i> -ADJ	<b>Équivalent CC</b> <i>pergélisol</i> -NOM (96) <b><i>permafrost</i>-NOM (87)</b>
Comme l'indiquent les chiffres entre parenthèses, l'équivalent <i>permafrost</i> -NOM est pratiquement aussi fréquent que <i>pergélisol</i> -NOM dans le CC alors que dans le CP la fréquence de <i>pergélisol</i> -NOM est de loin supérieure à celle de <i>permafrost</i> -NOM. Selon Termium, <i>permafrost</i> est un anglicisme à éviter. La Commission générale de terminologie et de néologie (France), quant à elle, admet <i>permafrost</i> , mais recommande <i>permagel</i> ou <i>pergélisol</i> .	

À titre de deuxième exemple (Tableau 7.10), nous voyons que le cognat *capture*-NOM est présent 120 fois dans le CC, mais complètement absent du CP. En ce qui concerne cet exemple, les équivalents *captage*-NOM et *piégeage*-NOM, des synonymes de *capture*-NOM, constituent des choix tout aussi bons que le cognat.

Tableau 7.10. Exemple de cognat critiquable (*capture*-NOM)

Terme anglais <i>capture</i> -NOM	
<b>Équivalent CP</b> <i>piégeage</i> -NOM (145) <i>piéger</i> -VER (14)	<b>Équivalent CC</b> <i>capture</i> -NOM (120) <i>capturer</i> -VER (8) <i>captage</i> -NOM (11) <i>capter</i> -VER (6) <i>piégeage</i> -NOM (6) <i>piéger</i> -VER (4)
Comme en témoignent les chiffres, la fréquence du cognat <i>capture</i> -NOM domine largement dans le CC, les autres équivalents étant très peu fréquents. Dans le CP, cet équivalent est absent. <i>Captage</i> est le terme recommandé dans le Journal officiel de la République française. Même si tous ces équivalents sont en usage et acceptables (Brodhag et Breuil 2009 : 22) (voir contextes ci-dessous), il n'en demeure pas moins que l'équivalent <i>capture</i> -NOM est un cognat utilisé dans de nombreux textes du CC, alors qu'il ne l'est pas dans le CP. CC : ... la CAPTURE et le stockage du CO <sub>2</sub> ... (Campos_captage-et-stockage_2007) CP : ... PIÉGEAGE et stockage du carbone... (IPCC_eau_2008)	

### 7.1.2.3 Notation des termes anglais ayant des équivalents différents

Chacun des équivalents différents est classé dans une des trois catégories suivantes :

1. **Équivalent critiquable dans le CP ( $\neq$ CP)** : l'équivalent différent est un emprunt fautif, un faux sens dans le CP ou un cognat dont la fréquence se démarque dans le CP par rapport au CC (à l'Annexe F, le Tableau F.2 présente l'analyse des termes anglais avec des  $\neq$ CP).
2. **Équivalent critiquable dans le CC ( $\neq$ CC)** : l'équivalent différent est un emprunt fautif, un faux sens dans le CC ou un cognat dont la fréquence se démarque dans le CC par rapport au CP (à l'Annexe F, le Tableau F.3 présente l'analyse des termes ayant des  $\neq$ CC).
3. **Équivalent non critiquable ( $\pm$ )** : l'équivalent différent ne présente aucune des caractéristiques des catégories 1 et 2 (à l'Annexe F, le Tableau F.1 présente l'analyse des termes anglais comportant des  $\pm$ ).

### 7.1.2.4 Classification des termes anglais en fonction du lieu (corpus) où se produit la différence d'équivalence

Comme nous l'avons vu, la différence d'équivalence se manifeste dans le corpus parallèle ( $\neq$ CP), dans le corpus comparable ( $\neq$ CC) ou dans les deux corpus à la fois ( $\pm$ ). Après avoir relevé les informations sur les équivalents différents et effectué leur analyse qualitative, nous ajoutons une information supplémentaire soit la classification des termes anglais en fonction du lieu (corpus) où se produit la ou les différences d'équivalence. Cette classification constitue une information particulièrement utile pour décrire les termes anglais avec équivalent différent non critiquable dans le CP ou le CC ( $\pm$ ). En effet, cela permet de voir où les différences ont lieu afin d'en évaluer l'importance. Les différences d'équivalence se manifestent selon deux scénarios : différence d'équivalence unilatérale (7.1.2.4.1) et différence d'équivalence bilatérale (7.1.2.4.2).

#### 7.1.2.4.1 Différence d'équivalence unilatérale

Il y a différence d'équivalence unilatérale lorsque pour un même terme anglais un des deux corpus propose un ou des équivalents différents par la forme<sup>105</sup> par rapport à l'autre corpus. La Figure 7.2 illustre un cas de différence unilatérale dans lequel l'équivalent est différent par la forme. En effet, dans le CP, en plus des équivalents *glace-NOM* et *glaciaire-NOM*, le terme anglais *ice-NOM* est associé à l'équivalent différent par la forme *glaciel-NOM* (Figure 7.2).

Terme anglais <i>ice-NOM</i>	
<b>CP</b>	<b>CC</b>
<i>glace-NOM</i>	<i>glace-NOM</i>
<i>glaciaire-ADJ</i>	<i>glaciaire-ADJ</i>
<b><i>glaciel-ADJ</i></b>	

Figure 7.2. Différence d'équivalence unilatérale avec un équivalent différent par la forme (*glaciel-ADJ*)

L'exemple de la Figure 7.3 montre un cas de différence d'équivalence unilatérale dans lequel les équivalents différents sont cette fois-ci situés dans le CC, puisque les équivalents *mitigation-NOM* et *mitiger-VER* ne sont pas présents dans le CP.

Terme anglais <i>mitigation-NOM</i>	
<b>CP</b>	<b>CC</b>
<i>atténuation-NOM</i>	<i>atténuation-NOM</i>
<i>atténuer-VER</i>	<i>atténuer-VER</i>
	<b><i>mitigation-NOM</i></b>
	<b><i>mitiger-VER</i></b>

Figure 7.3. Différence d'équivalence unilatérale avec deux cognats critiquables (*mitigation-NOM* et *mitiger-VER*)

<sup>105</sup> Un équivalent dans un corpus se distingue par la forme quand il n'a pas son semblable dans l'autre corpus, par exemple, l'équivalent *glaciel-NOM* du CP n'a pas son pendant dans le CC (Figure 7.2).

#### 7.1.2.4.2 Différence d'équivalence bilatérale

Il y a différence d'équivalence bilatérale lorsqu'un terme anglais est associé à un ou des équivalents différents par la forme et/ou à un ou des cognats critiquables<sup>106</sup> dans les deux corpus. Par exemple, pour le terme anglais *coal-NOM*, les deux corpus proposent trois équivalents chacun, mais seulement deux sont identiques par la forme dans les deux corpus puisque *houiller-ADJ* et *charbonnier-ADJ* sont respectivement présents dans un seul corpus (Figure 7.4).

Terme anglais <i>coal-NOM</i>	
<b>CP</b>	<b>CC</b>
<i>charbon-NOM</i>	<i>charbon-NOM</i>
<i>houille-NOM</i>	<i>houille-NOM</i>
<i>houiller-ADJ</i>	<i>charbonnier-ADJ</i>

Figure 7.4. Différence d'équivalence bilatérale avec des équivalents différents par la forme (*houiller-NOM* et *charbonnier-NOM*)

L'exemple de la Figure 7.5 illustre un cas de différence d'équivalence bilatérale dans lequel on trouve un équivalent différent par la forme (*pergélisolé-ADJ*) dans le CP et un cognat critiquable (*permafrost-NOM*) dans le CC.

Terme anglais <i>permafrost-NOM</i>	
<b>CP</b>	<b>CC</b>
<i>pergélisol-NOM</i> (308)	<i>pergélisol-NOM</i> (96)
<i>permafrost-NOM</i> (16)	<i>permafrost-NOM</i> (87)
<i>pergélisolé-NOM</i> (6)	

Figure 7.5. Différence d'équivalence bilatérale avec équivalent différent par la forme (*pergélisolé-ADJ*) et cognat critiquable (*permafrost-NOM*)

<sup>106</sup> Un cognat critiquable se distingue par sa fréquence plus élevée dans un corpus par rapport à un autre équivalent dans le même corpus ou par rapport à l'équivalent de l'autre corpus. Ainsi, la fréquence de *permafrost-NOM* dans le CC est significativement plus élevée que celle du CP. De plus, l'écart de fréquence entre *pergélisol-NOM* et *permafrost-NOM* est beaucoup plus important dans le CP que dans le CC (Figure 7.5). Le cognat critiquable peut également ne pas avoir son semblable dans l'autre corpus, par exemple, l'équivalent *mitigation-NOM* et *mitiger-VER* ne sont pas présents dans le CP (Figure 7.3)

## **7.2 Résultats de l'analyse comparative des équivalents du corpus parallèle aux équivalents du corpus comparable**

Ce chapitre présente les résultats de l'analyse comparative des équivalents du corpus parallèle aux équivalents du corpus comparable. La section 7.2.1 porte sur l'analyse comparative quantitative globale, la section 7.2.2 traite des termes anglais avec des équivalents identiques dans les deux corpus et la section 7.2.3 décrit les termes anglais avec des différences d'équivalence dans les deux corpus.

### **7.2.1 Analyse comparative quantitative globale**

Dans le corpus parallèle, nous avons répertorié 811 équivalents (*cf.* 6.2) et dans le corpus comparable 848 (*cf.* 6.3), ce qui représente une différence de 37 équivalents. La Figure 7.6 illustre le nombre de termes anglais en fonction du nombre d'équivalents pour un même terme dans le corpus parallèle et dans le corpus comparable. Les colonnes indiquent le nombre d'équivalents et s'échelonnent de 0 à 10 équivalents. Dans le graphique, nous observons tout d'abord que les colonnes affichent des valeurs assez semblables dans les deux corpus. Par exemple, les termes avec 1 équivalent sont au nombre de 113 dans le CP et de 109 dans le CC et ainsi de suite. Par ailleurs, les termes anglais sont majoritairement associés à 1 ou 2 équivalents dans les deux corpus. D'autre part, nous remarquons que le CC est le seul à avoir 0 ou 10 équivalents par terme. D'ores et déjà, nous constatons que les deux corpus présentent de grandes similitudes sur le plan quantitatif global.

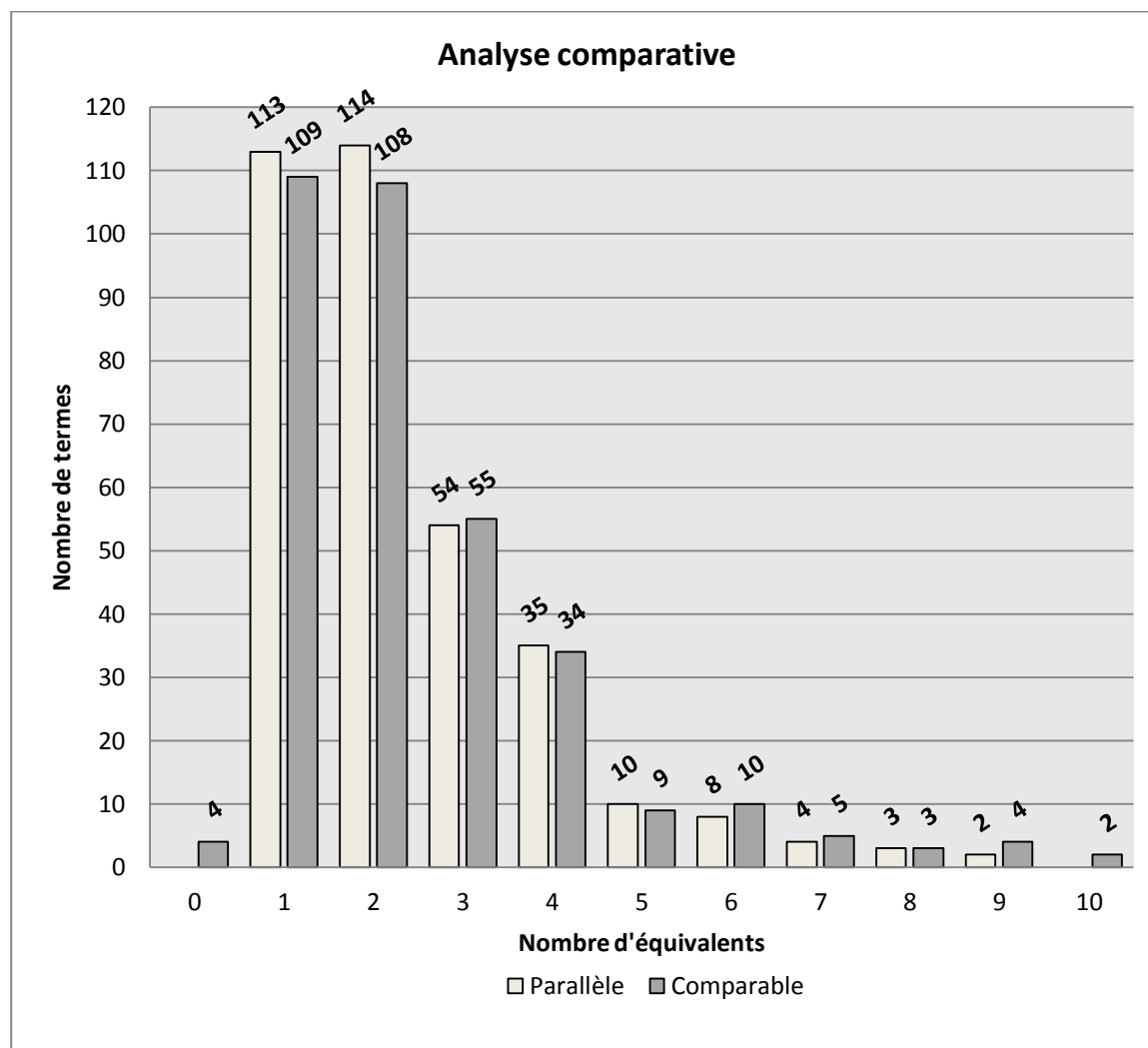


Figure 7.6. Nombre de termes en fonction du nombre d'équivalents (corpus parallèle et corpus comparable)

Sur les 343 termes anglais, 272 termes ont des équivalents identiques dans les deux corpus, 67 termes anglais présentent des différences d'équivalence dans les deux corpus et 4 termes n'ont pas d'équivalent dans le corpus comparable (Tableau 7.11).



Tableau 7.11. Résultat de l'analyse comparative quantitative

Description	Total	Pourcentage
Équivalents identiques dans le CP et le CC	272	79,3
Différence d'équivalence dans le CP et le CC	67	19,5
Zéro équivalent dans le CC	4	1,2
<b>Total</b>	<b>343</b>	<b>100</b>

Les quatre termes anglais, 3 noms et 1 adjectif, pour lesquels il n'y a pas d'équivalent dans le corpus comparable sont *biome*-NOM, *biota*-NOM, *diversion*-NOM et *unsustainable*-ADJ (Tableau 7.12 et Tableau 6.9). Dans le corpus parallèle, le terme *diversion*-NOM (action de changer le cours, la direction d'un cours d'eau) est utilisé surtout dans les textes traitant des problèmes liés aux Grands Lacs. Ce sujet est très peu abordé dans le corpus comparable. Pour les termes *biome*-NOM (milieu de la planète couvrant de vastes surfaces), *biota*-NOM (ensemble des êtres vivants d'un lieu donné) et *unsustainable*-ADJ (non durable), nous ne sommes pas en mesure d'expliquer la cause de l'absence de ces notions dans le corpus comparable.

Tableau 7.12. Équivalents absents du corpus comparable

Terme anglais	Corpus parallèle	Corpus comparable
<i>biome</i> -NOM	<i>biome</i> (51)	∅
<i>biota</i> -NOM	<i>biote</i> (24)	∅
<i>diversion</i> -NOM	<i>dérivation</i> (11) <i>détournement</i> (9)	∅ ∅
<i>unsustainable</i> -ADJ	<i>non durable</i> (17)	∅

À partir de cette section, nous baserons nos calculs sur 339 termes anglais au lieu des 343 termes, car 4 termes ne sont pas représentés dans le CC.

Le Tableau 7.13 montre la répartition des équivalents complexes et des équivalents simples par ÉqT et ÉqNT. Nous observons que chez les équivalents complexes, les ÉqT sont relativement plus nombreux que les ÉqNT dans les deux corpus. Par contre, chez les termes simples, les ÉqT sont moins nombreux que les ÉqNT dans les deux corpus.

Tableau 7.13. Répartition des équivalents par ÉqT et ÉqNT dans les deux corpus

Type d'équivalent	Nombre CP	Nombre CC
Terme complexe ÉqT	32	25
Terme complexe ÉqNT	15	18
Terme simple ÉqT	371	379
Terme simple ÉqNT	393	426
<b>Total</b>	<b>811</b>	<b>848</b>

Les 272 termes anglais avec équivalents identiques se répartissent en 195 noms, 45 adjectifs, 26 verbes et 6 adverbes et les 67 termes anglais avec des différences d'équivalence se répartissent en 55 noms, 6 adjectifs, 5 verbes et 1 adverbe (Tableau 7.14). Les données du Tableau 7.14 indiquent que les noms sont relativement plus nombreux dans la catégorie « équivalents différents » (82,1 %) que dans celle des « équivalents identiques » (71,7 %). Nous constatons également que les adjectifs, les verbes et les adjectifs sont relativement plus fréquents dans la catégorie « équivalents identiques » que dans la catégorie « équivalents différents ». Cependant, en ce qui concerne les adverbes, ils sont trop peu nombreux dans la liste des termes anglais (7) pour que l'on puisse tirer des conclusions significatives. De cette analyse, nous en arrivons à la conclusion que les noms semblent plus sujets à présenter des différences d'équivalence dans les deux corpus que les autres parties du discours.

Tableau 7.14. Répartition des termes anglais par parties du discours

Partie du discours	Équivalent identique		Équivalent différent	
	Nombre	%	Nombre	%
Nom	195	71,7	55	82,1
Adjectif	45	16,5	6	8,9
Verbe	26	9,6	5	7,5
Adverbe	6	2,2	1	1,5
<b>Total</b>	<b>272</b>	<b>100</b>	<b>67</b>	<b>100</b>

Sur les 272 termes anglais ayant des équivalents identiques, 206 ont été validés selon le critère a)<sup>107</sup> et 66 en fonction du critère b), c) ou d)<sup>108</sup> (cf. 5.3.2.1). Des 67 termes anglais ayant des différences d'équivalence, 61 ont été sélectionnés avec le critère a) et 6 avec le critère b), c) ou d). Les chiffres du Tableau 7.15 montrent clairement que les termes sélectionnés sur la base du critère b), c) ou d) sont relativement inférieurs dans la catégorie « équivalent différent » (9 %) par rapport à la catégorie « équivalent identique » (24 %) et que c'est l'inverse pour les termes sélectionnés selon le critère a) : « équivalent différent » 91 %, « équivalent identique » (76 %). À ce stade, il est important de rappeler que les termes validés avec le critère a) sont majoritairement des noms. Ce résultat confirme donc notre analyse précédente, indiquant que les parties du discours, autres que le nom, n'influent pas significativement sur le nombre d'équivalents différents associé à un terme anglais.

Tableau 7.15. Répartition des termes anglais par critère de sélection a), b), c) et d)

Critère	Équivalent identique		Équivalent différent	
	Nombre	%	Nombre	%
critère a)	206	76	61	91
Critère b), c) ou d)	66	24	6	9
<b>Total</b>	<b>272</b>	<b>100</b>	<b>67</b>	<b>100</b>

Au terme de cette première analyse, nous avons constaté que la majorité des termes anglais (272) sont rendus par des équivalents identiques dans le CP et le CC. En revanche, 67 termes anglais présentent des différences d'équivalence. Nous avons par ailleurs observé que les termes appartenant à la partie du discours du nom sont relativement plus fréquents que les autres parties du discours dans la catégorie « équivalent différent ». Nous avons aussi remarqué que les termes sélectionnés avec le critère a) sont relativement plus fréquents dans la catégorie « équivalent différent ».

<sup>107</sup> Le critère a) s'appuie sur les connaissances du domaine de spécialité étudié et ont été validés par des ressources terminologiques.

<sup>108</sup> Les critères b), c) et d) font appel à des indices lexico-sémantiques. Les critères b), c) et d) sont pris en bloc, car certains termes peuvent être validés à l'aide des trois critères, par exemple, le verbe *cool (to)*.

Afin de nous faire une idée plus juste des termes avec des équivalents identiques et des termes avec des équivalents différents, nous les avons analysés plus en détail dans les sections 7.2.2 et 7.2.3.

### **7.2.2 Résultats de l'analyse des termes anglais avec des équivalents identiques dans les deux corpus**

Sur les 339 termes anglais analysés, 272 comportent des équivalents identiques dans les deux corpus (semblables par la forme et de nombre égal), ce qui représente 80,2 % de l'ensemble des équivalents observés. Dans le graphique de la Figure 7.7, nous avons reproduit les données de la Figure 7.6 et ajouté le nombre de termes anglais ayant des équivalents identiques dans les deux corpus en fonction du nombre d'équivalents par terme (tracé noir de la Figure 7.7). Dans ce nouveau graphique, nous remarquons que des 272 termes, 99 ont un seul et même équivalent dans les deux corpus, 89 en ont 2 et ainsi de suite. Le tracé noir permet de montrer que le nombre d'équivalents associé à un terme anglais n'influence pas fortement le nombre de termes ayant des équivalents identiques. En effet, les tranches ayant 3 équivalents et plus par terme présentent quand même des équivalents identiques dans plus ou moins les mêmes proportions que les termes ayant 1 ou 2 équivalents identiques.

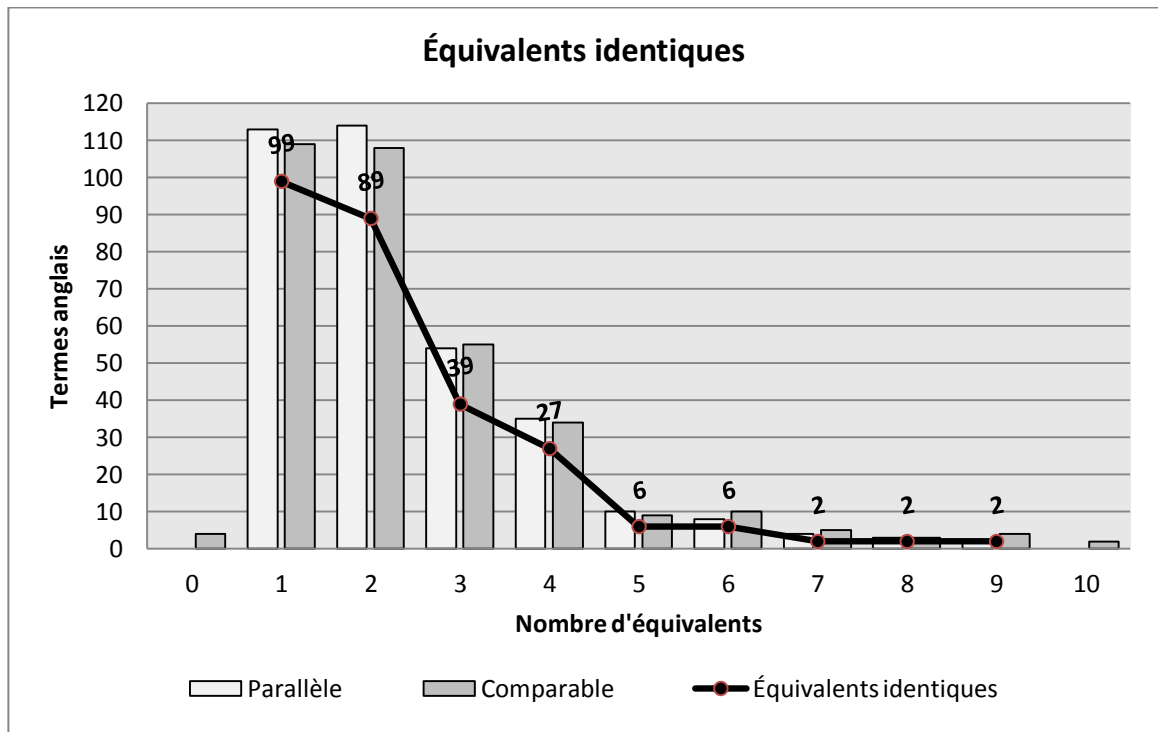


Figure 7.7. Termes avec équivalents identiques dans le CP et le CC

Le Tableau 7.16 montre la distribution des termes avec équivalents identiques en fonction de leur partie du discours et de leur nombre d'équivalents. Par exemple, dans la tranche des termes ayant 1 équivalent, on retrouve 81 noms, 16 adjectifs, 2 verbes et aucun adverbe et dans la tranche des termes ayant 9 équivalents, 1 nom, 0 adjectif, 1 verbe et aucun adverbe.

Tableau 7.16. Termes anglais et distribution par parties du discours et par nombre d'équivalents

Partie du discours	1 équivalent	2	3	4	5	6	7	8	9	Total
Nom	81	60	25	20	5	1	2	0	1	<b>195</b>
Adjectif	16	17	6	5	0	1	0	0	0	<b>45</b>
Verbe	2	10	5	1	1	4	0	2	1	<b>26</b>
Adverbe	0	2	3	1	0	0	0	0	0	<b>6</b>
<b>Total</b>										<b>272</b>

Pour illustrer plus concrètement les données du Tableau 7.16, nous les avons reproduites sous la forme d'un graphique à colonnes à la Figure 7.8. Dans ce graphique, nous constatons que les verbes sont très peu fréquents dans le groupe des termes anglais

ayant 1 équivalent et que les adverbes y sont absents. Nous voyons aussi que dépassé le groupe des 6 équivalents, ne sont présents que des noms et des verbes. Ces données semblent indiquer que la partie du discours à laquelle appartient un terme et le nombre d'équivalents auxquels il est associé n'a pas d'impact significatif sur le nombre de termes avec des équivalents identiques.

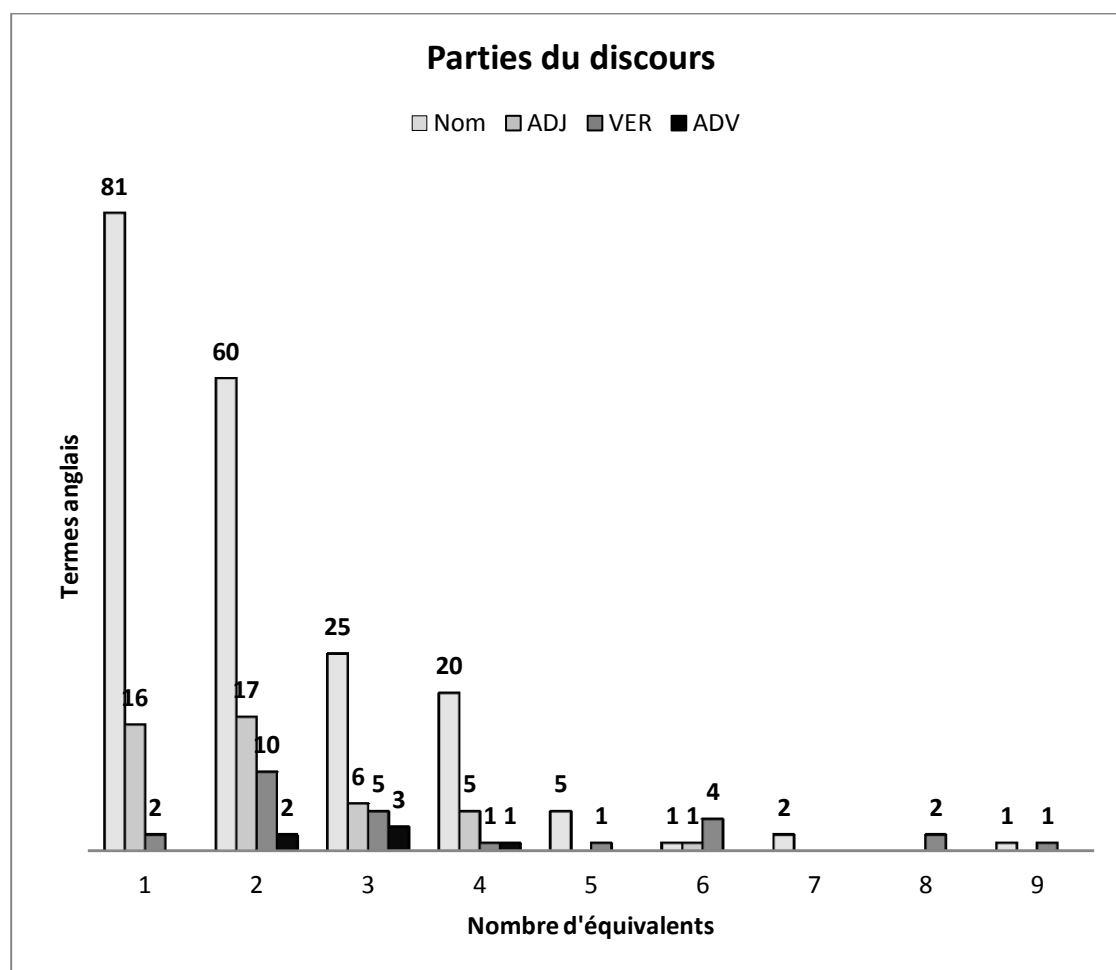


Figure 7.8. Répartition des termes anglais en fonction de leur partie du discours et de leur nombre d'équivalents

La Figure 7.9 illustre la répartition des 272 termes en fonction du critère de validation et du nombre d'équivalents. Dans cette figure, nous constatons que les termes validés avec les critères b), c) ou d) possèdent une nette tendance à être rendus par plusieurs équivalents identiques dans les deux corpus.

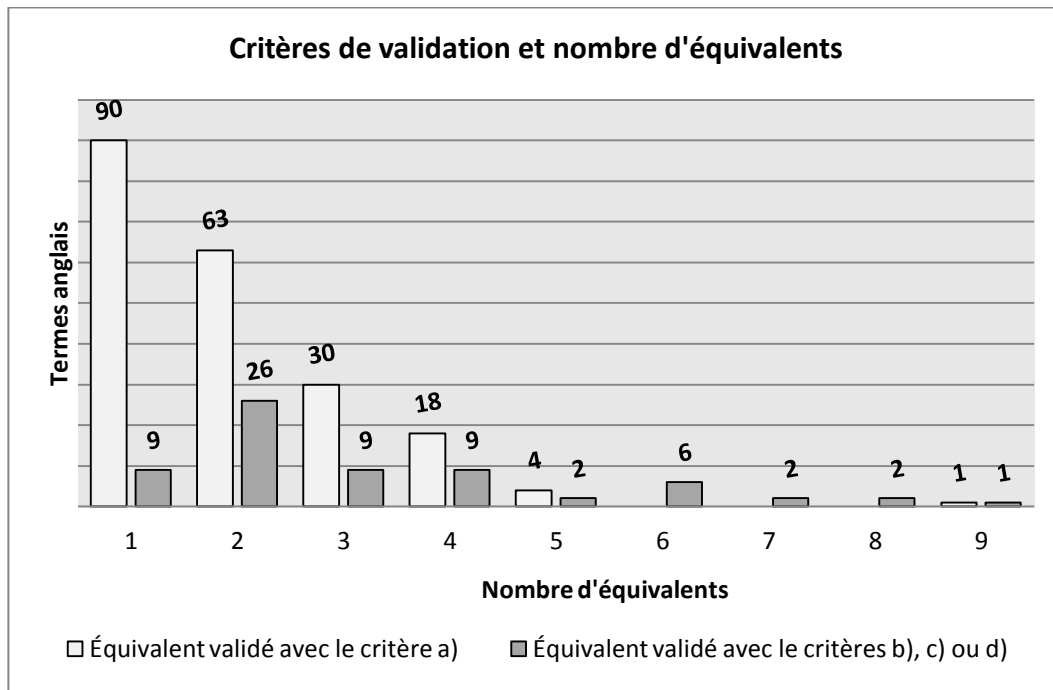


Figure 7.9. Répartition des termes anglais avec équivalents identiques en fonction du critère de validation

Au terme de cette analyse, nous avons observé que le nombre d'équivalents pour un même terme n'est pas un facteur déterminant pour que ce terme soit associé à des équivalents identiques dans les deux corpus (Figure 7.7). Autrement dit, qu'un terme soit associé à un ou plusieurs équivalents, il a autant de chance que ses équivalents soient identiques dans les deux corpus. La distribution des termes anglais en fonction de la partie du discours et du nombre d'équivalents nous apprend que les verbes et les adverbes, qui ont tendance à être associés à plusieurs équivalents, ne semblent pas influencer sur le nombre de termes associés à des équivalents identiques (Tableau 7.16 et Figure 7.8). Nous avons aussi vu que les termes sélectionnés avec le critère b), c) ou d) sont la plupart du temps rendus par plusieurs équivalents (Figure 7.9).

Dans la section qui suit (7.2.3), nous poursuivons notre analyse avec les termes ayant des équivalents différents ( $\pm$ ,  $\neq$ CP,  $\neq$ CC).

### 7.2.3 Termes anglais avec des différences d'équivalence dans les deux corpus

Sur les 339 termes anglais, 67 termes présentent des différences d'équivalence formelles ou quantitatives, ce qui équivaut à 19,8 % de l'ensemble des termes analysés. Afin de mieux interpréter ce pourcentage, nous avons effectué l'analyse qualitative des équivalents (cf. 7.1.2.2). Après notation des équivalents des 67 termes anglais, nous obtenons 55 termes avec des équivalents différents non critiquables dans le CP ou le CC ( $\pm$ ), 2 termes avec des équivalents critiquables dans le corpus parallèle ( $\neq$ CP) et 10 termes avec des équivalents critiquables dans le corpus comparable ( $\neq$ CC) (Tableau 7.17). Les  $\pm$  représentent 82,1 %, de l'ensemble des termes ayant des équivalents différents, alors que les  $\neq$ CP et les  $\neq$ CC correspondent respectivement à 3 % et 14,9 % (Tableau 7.17).

Tableau 7.17. Les 67 équivalents différents répartis par  $\pm$ ,  $\neq$ CP,  $\neq$ CC

Barème	Nombre	Pourcentage sur les 67 termes
$\pm$	55	82,1
$\neq$ CP	2	3
$\neq$ CC	10	14,9
<b>Total</b>	<b>67</b>	<b>100</b>

Si l'on reporte les chiffres de la colonne « nombre » du Tableau 7.17 sur l'ensemble des termes analysés, soit 339 termes anglais (Tableau 7.18), les  $\pm$  représentent 16,2 %, tandis que les  $\neq$ CP et les  $\neq$ CC, 0,6 % et 2,9 % respectivement, soit seulement 3,5 % des 339 termes.

Tableau 7.18. Pourcentage des équivalents différents sur les 339 termes anglais

Barème	Nombre	Pourcentage sur les 339 termes
$\pm$	55	16,2
$\neq$ CP	2	0,6
$\neq$ CC	10	2,9
<b>Total</b>	<b>67</b>	<b>19,7</b>



Les Figures 7.10a et 7.10b montrent la répartition des termes anglais du CP et du CC ayant des équivalents différents en fonction de la partie du discours et du nombre d'équivalents. Dans les deux corpus, aucun terme autre que le nom n'est associé à un seul équivalent, les verbes, les adjectifs et l'adverbe sont associés à au moins deux équivalents. Bien que les données soient beaucoup moins nombreuses (67) que pour les équivalents identiques (272), elles semblent suivre les mêmes tendances que celles observées pour les termes ayant des équivalents identiques, à savoir que la partie du discours à laquelle appartient un terme et le nombre d'équivalents auxquels il est associé n'a pas d'impact significatif sur le nombre de termes avec des équivalents différents.

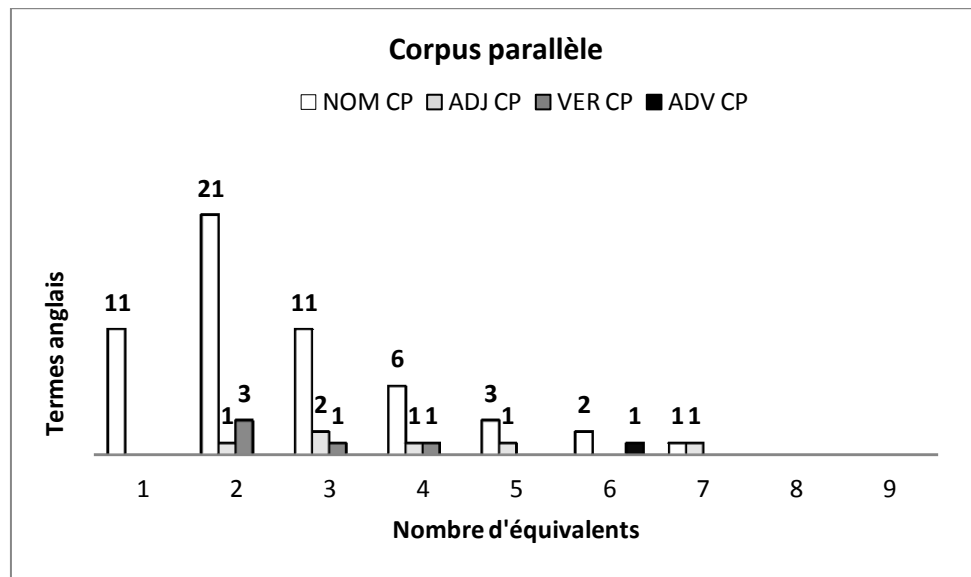


Figure 7.10a. Catégorisation des termes avec équivalents différents en fonction de la partie du discours (corpus parallèle).

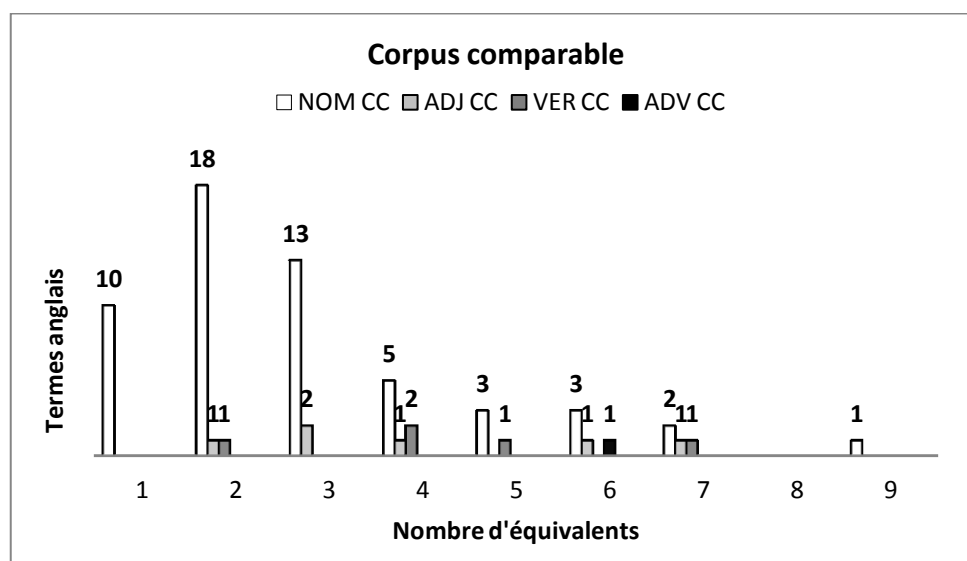


Figure 7.10b. Catégorisation des termes avec équivalents différents en fonction de la partie du discours (corpus comparable).

Le Tableau 7.19 affiche la répartition en fonction de la partie du discours et du barème de notation ( $\pm$ ,  $\neq$ CP,  $\neq$ CC). Les adjectifs et les verbes figurent relativement plus fréquemment dans la catégorie  $\neq$ CP et  $\neq$ CC que les noms, cependant la fréquence des termes possédant ces parties du discours est trop peu élevée pour qu'on puisse en tirer des conclusions générales.

Tableau 7.19. Répartition des 67 termes par partie du discours

Partie du discours	$\pm$	$\neq$ CP	$\neq$ CC	Total	Pourcentage
Nom	49	0	6	55	82,1
Adjectif	3	1	2	6	8,9
Verbe	2	1	2	5	7,5
Adverbe	1	0	0	1	1,5
<b>Total</b>	<b>55</b>	<b>2</b>	<b>10</b>	<b>67</b>	<b>100</b>

La Figure 7.11 illustre la répartition des équivalents différents du CP et du CC en fonction du critère de validation. Tout d'abord, nous remarquons que les termes dont les équivalents sont validés avec les critères b), c) ou d) sont toujours rendus par plusieurs équivalents dans les deux corpus (Figure 7.11). En effet, dans la colonne des termes ayant un seul équivalent, aucun n'a été validé par ce critère, il est vrai que les données sont trop

peu nombreuses pour donner un portrait exact. Cependant, ce dernier résultat combiné à celui de la Figure 7.9 montre que les critères b), c) ou d) ne semblent pas influencer sur le nombre de termes associés à des équivalents identiques ou différents.

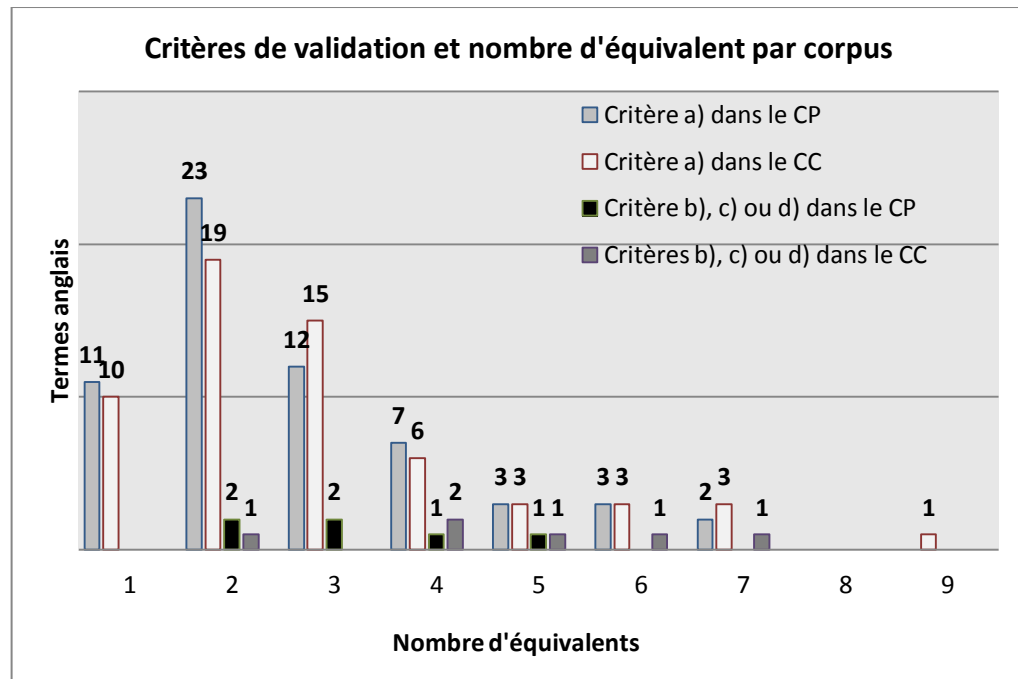


Figure 7.11. Répartition des termes avec équivalents différents du CP et du CC en fonction du critère de validation

Les sections qui suivent décrivent plus en détail les catégories de termes avec des équivalents différents ( $\pm$ ,  $\neq$ CP,  $\neq$ CC).

### 7.2.3.1 Résultats de l'analyse des termes anglais avec des équivalents différents non critiquables dans le CP ou le CC ( $\pm$ )<sup>109</sup>

La collecte des informations générales nous apprend que les 55 termes anglais  $\pm$  ont 68 équivalents différents, 48 sont des équivalents non typiques (ÉqNT) et 20 appartiennent à la classe des équivalents typiques (ÉqT) (cf. 7.1.2.1) (Tableau 7.20).

<sup>109</sup> Les  $\pm$  sont des termes anglais pour lesquels aucun des équivalents différents n'est un emprunt, un faux sens ou un cognat dont la fréquence se démarque dans le CC par rapport au CP et inversement.

Tableau 7.20. Équivalents classés en fonction des critères ÉqNT et ÉqT

Type d'équivalent	Nombre	Pourcentage
ÉqNT	48	70,6
ÉqT	20	29,4
<b>Total</b>	<b>68</b>	<b>100</b>

Sur les 68 équivalents différents, 40 équivalents sont peu fréquents (15 occurrences et moins), soit 58,8 %.

Enfin, 50 termes anglais présentent des différences d'équivalence unilatérales (cf. 7.1.2.4.1) et 5 des différences d'équivalence bilatérale (cf. 7.1.2.4.2). Les différences unilatérales se manifestent comme suit : 29 dans le CC et 21 dans le CP. Les fiches suivantes montrent des exemples de différence d'équivalence unilatérale et bilatérale. Les fiches 1 à 4 sont des différences unilatérales, les fiches 5 et 6 illustrent des cas de différence bilatérale.

**Fiche 1. Différence unilatérale**Terme anglais : *subsidence*-NOM**Équivalent CP***subsidence*-NOM***affaissement de terrain*-NOM****Équivalent CC***subsidence*-NOM**Fiche 2. Différence unilatérale**Terme anglais : *peat*-NOM**Équivalent CP***tourbe*-NOM***tourbeux*-ADJ****Équivalent CC***tourbe*-NOM**Fiche 3. Différence unilatérale**Terme anglais : *harvest*-NOM**Équivalent CP***récolte*-NOM*récolter*-VER**Équivalent CC***récolte*-NOM*récolter*-VER***cueillette*-NOM**

**Fiche 4. Différence unilatérale**Terme anglais : *landfill*-NOM

Équivalent CP

*décharge*-NOM

Équivalent CC

*décharge*-NOM***centre d'enfouissement*-NOM****Fiche 5. Différence bilatérale**Terme anglais : *woodland*-NOM

Équivalent CP

***région boisée*-NOM**

Équivalent CC

***surface boisée*-NOM****Fiche 6. Différence bilatérale**Terme anglais : *sequester*-VER

Équivalent CP

*piéger*-VER*fixer*-VER*piégeage*-NOM***séquestrer*-VER**

Équivalent CC

*piéger*-VER*fixer*-VER*piégeage*-NOM***capturer*-VER*****fixation*-NOM**

Au terme de l'analyse des termes avec des équivalents non critiquables, nous constatons que leurs équivalents sont dans la très grande majorité des équivalents non typiques, qu'ils sont assez souvent peu fréquents et que la différence unilatérale est de loin la plus fréquente.

### 7.2.3.2 Analyse des termes anglais avec équivalent différent critiquable dans le CP (≠CP)<sup>110</sup>

Les termes anglais avec un ou des équivalents critiquables dans le CP (≠CP) sont au nombre de 2 : *reflect*-VER et *anthropogenic*-ADJ. Un équivalent de ces termes anglais est un faux sens, *réfléter*-VER, et deux équivalents sont des emprunts fautifs, *anthropogène*-ADJ et *anthropogénique*-ADJ (Tableau 7.21).

La collecte d'informations générales des équivalents critiquables dans le corpus parallèle indique qu'ils sont tous des ÉqNT et des ÉqPF (Tableau 7.21).

<sup>110</sup> Les ≠CP sont des termes anglais pour lesquels l'équivalent différent est un emprunt, un faux sens ou un congnat dont la fréquence se démarque dans le CP par rapport au CC.

Tableau 7.21. Informations générales sur les ≠CP

Terme anglais	Équivalent	ÉqT/ÉqNT	ÉqPF/ÉqF
<i>reflect</i> -VER	<i>réfléter</i> -VER (6)	ÉqNT	ÉqPF
<i>anthropogenic</i> -ADJ	<i>anthropogène</i> -ADJ (11)	ÉqNT	ÉqPF
	<i>anthropogénique</i> -ADJ (4)	ÉqNT	ÉqPF

Les deux fiches suivantes expliquent les raisons pour lesquelles ces équivalents sont critiquables et montrent que les termes anglais *reflect*-VER et *anthropogenic*-ADJ sont des cas de différence d'équivalence bilatérale.

#### Fiche 1. Différence bilatérale

**Terme anglais :** *reflect*-VER

##### Équivalent CP

*réfléchir*-VER

***réfléter*-VER(6)**

##### Équivalent CC

*réfléchir*-VER

***renvoyer*-VER (6)**

Dans le cas du terme anglais *reflect*-VER ('renvoyer par réflexion'), l'équivalent *réfléter*-VER ('réfléchir de façon atténuée et plus ou moins vague') ne rend pas, dans ce contexte, le sens du terme anglais (Van Roey *et al.* 1991). L'équivalent *réfléter*-VER est néanmoins très peu fréquent (6) dans le CP. Les contextes suivants, relevés dans le CP, montrent *réfléter*-VER et *réfléchir*-VER.

CP : Once in the atmosphere, these gases trap and REFLECT heat back toward the Earth's surface...

CP : Une fois dans l'atmosphère, ces gaz piègent et REFLÈTENT la chaleur vers la surface de la Terre... (RNCA\_adaptation\_2004)

CP : ... clouds (which both REFLECT sunlight and absorb outgoing heat radiation)

CP : ... des nuages (qui RÉFLÉCHISSENT la lumière du Soleil et absorbent le rayonnement thermique réfléchi)... (RNCA\_impacts\_2008)

#### Fiche 2. Différence bilatérale

**Terme anglais :** *anthropogenic*-ADJ

##### Équivalent CP

*anthropique*-ADJ (230)

*humain*-ADJ

***anthropogène*-ADJ (11)**

***anthropogénique*-ADJ (4)**

##### Équivalent CC

*anthropique*-ADJ (233)

*humain*-ADJ

***homme*-NOM**

Les dictionnaires ne s'entendent pas toujours sur les équivalents du terme anglais *anthropogenic*-ADJ ('provoqué directement ou indirectement par l'action de l'homme'). IATE propose *anthropogénique*-ADJ, *anthropique*-ADJ et *anthropogène*-ADJ, ceux-là mêmes trouvés dans le CP. Toutefois, si l'on se fie à Termium, le seul équivalent valable est *anthropique*-ADJ et les deux autres sont des anglicismes à éviter. D'autre part, *anthropique*-ADJ est de loin le plus fréquent dans le CP et c'est le seul qui soit attesté dans Brodhag et Breuil (2009), Bioret *et al.* (2009) ou Dutuit et Gorenflot (2008) – tous des sources européennes. Nous considérons donc que les équivalents *anthropogène*-ADJ et *anthropogénique*-ADJ sont à éviter.

Le fait que les 2 termes anglais présentent des ressemblances formelles avec des mots qui existent dans la langue française a probablement influencé les traducteurs à choisir des équivalents fautifs. Cependant, l'emploi de ces équivalents est très peu fréquent dans le CP.

### 7.2.3.3 Résultats de l'analyse des termes anglais avec équivalent différent critiquable dans le CC ( $\neq$ CC)<sup>111</sup>

Les termes anglais comportant des  $\neq$ CC sont au nombre de 10 : *capture*-NOM; *capture*-VER; *deforestation*-NOM; *reforestation*-NOM; *global*-ADJ; *mitigate*-VER; *mitigation*-NOM; *monitoring*-NOM; *offshore*-ADJ; *permafrost*-NOM (Tableau 7.22).

La collecte des informations générales sur les 10 équivalents critiquables dans le CC montre que 5 d'entre eux appartiennent à la catégorie ÉqT et 5 à celle des ÉqNT (Tableau 7.22). D'autre part, 3 équivalents sont fréquents (ÉqF), 4 sont peu fréquents (ÉqPF) et 3 comptent un peu plus d'occurrences que les 15 occurrences pour être considérés comme peu fréquents.

Tableau 7.22. Informations générales sur les  $\neq$ CC

Terme anglais	Équivalent	ÉqT/ÉqNT	ÉqPF ou ÉqF
<i>capture</i> -NOM	<i>capture</i> -NOM	ÉqT	ÉqF
<i>capture</i> -VER	<i>capturer</i> -VER	ÉqT	ÉqPF
<i>deforestation</i> -NOM	<i>déforestation</i> -NOM	ÉqT	42 occ.
<i>reforestation</i> -NOM	<i>reforestation</i> -NOM	ÉqT	ÉqPF
<i>global</i> -ADJ	<i>global</i> -ADJ	ÉqNT	ÉqF
<i>mitigate</i> -VER	<i>mitiger</i> -VER	ÉqNT	ÉqPF
<i>mitigation</i> -NOM	<i>mitigation</i> -NOM	ÉqNT	35 occ.
<i>monitoring</i> -NOM	<i>monitoring</i> -NOM	ÉqNT	ÉqPF
<i>offshore</i> -ADJ	<i>offshore</i> -ADJ	ÉqNT	31 occ.
<i>permafrost</i> -NOM	<i>permafrost</i> -NOM	ÉqT	ÉqF

Les équivalents critiquables des 10 termes  $\neq$ CC présentent tous des ressemblances formelles avec le terme anglais (voir Tableau 7.22). *Capture*-NOM, *capture*-VER,

<sup>111</sup> Les  $\neq$ CC sont des termes pour lesquels l'équivalent différent est un emprunt, un faux sens ou un cognat dont la fréquence se démarque dans le CC par rapport au CP.

*deforestation*-NOM, *reforestation*-NOM et *permafrost*-NOM sont des cognats dont la fréquence est significativement plus élevée dans le CC que dans celui du CP. *Global*-ADJ, *mitigate*-VER, *mitigation*-NOM, *monitoring*-NOM et *offshore*-ADJ sont des anglicismes.

Les fiches 1 à 10 présentent les termes anglais ayant des équivalents critiquables dans le CC, 6 sont des différences d'équivalence unilatérale et 4 des différences d'équivalence bilatérales.

#### Fiche 1. Différence unilatérale

Terme anglais : *capture*-NOM

##### Équivalent CP

*piégeage*-NOM (145)

*piéger*-VER (14)

##### Équivalent CC

***capture***-NOM (120)

***capturer***-VER (8)

***captage***-NOM (11)

***capter***-VER (6)

*piégeage*-NOM (6)

*piéger*-VER (4)

Comme en témoignent les chiffres, la fréquence du cognat *capture*-NOM domine largement dans le CC, les autres équivalents étant très peu fréquents. Cet équivalent est en usage et acceptable (Brodhag et Breuil 2009 : 22). Il n'en demeure pas moins que l'équivalent *capture*-NOM est un cognat utilisé dans de nombreux textes du CC, alors qu'il ne l'est pas dans le CP.

CP : PIÉGEAGE et stockage du carbone (IPCC\_eau\_2008)

CC : la CAPTURE et le stockage du CO<sub>2</sub> (Campos\_captage-et-stockage\_2007)

#### Fiche 2. Différence unilatérale

Terme anglais : *capture*-VER

##### Équivalent CP

*piéger*-VER (30)

*piégeage*-NOM (10)

##### Équivalent CC

***capturer***-VER (8)

***capture***-NOM (120)

***capter***-VER (6)

***captage***-NOM (11)

*piéger*-VER (4)

*piégeage*-NOM (6)

Le terme anglais *capture*-VER n'est pas présent dans les banques de terminologie, mais il est morphologiquement relié au terme *capture*-NOM. *Capturer*-VER est complètement absent du CP.

CP : ...qui permettent de PIÉGER les émissions émanant de ces sources (IPCC\_dioxyde\_2005).

CC : volumes gigantesques de CO<sub>2</sub> CAPTURÉS à partir de sources d'émissions... (Senat\_energie\_2006)



**Fiche 3. Différence bilatérale****Terme anglais** : *deforestation*-NOM**Équivalent CP**      **Équivalent CC*****deforestation***-NOM (13)      ***deforestation***-NOM (42)*déboisement*-NOM (145)      *déboisement*-NOM (6)

Les chiffres montrent que le cognat *deforestation*-NOM est plus fréquent que *déboisement*-NOM dans le CC, alors que c'est l'inverse dans le CP. Selon Termium et IATE ces deux équivalents sont synonymes.

CP : Le terme "DÉBOISEMENT" peut être défini comme étant la transformation d'une terre forestière en terre non forestière (IPCC\_LULUCF\_2000).

CC : Parmi les risques réels, figure la destruction de la forêt tropicale qui privera la recherche pharmaceutique de la découverte de molécules pharmacologiques actives extraites des végétaux alors que cette DÉFORESTATION est surtout perçue par l'opinion comme une atteinte au "poumon de la planète", ce qui est une image inexacte (Senat\_ampleur\_2002).

**Fiche 4. Différence unilatérale****Terme anglais** : *reforestation*-NOM**Équivalent CP****Équivalent CC***reboisement*-NOM (70)*reboisement*-NOM (19)***reforestation***-NOM (6)

Les chiffres indiquent qu'il y a 6 occurrences de *reforestation*-NOM dans le CC et aucune dans le CP. Selon Termium et IATE, ces deux équivalents sont synonymes. Cependant, IATE précise que la *reforestation* se fait généralement sur une plus grande surface que le *boisement*.

CP : Les activités de REBOISEMENT peuvent aussi avoir des effets positifs sur la biodiversité (PNUE\_bioseq\_2009).

CC : ...mais les avantages présentés par les énergies renouvelables, solaire, éolienne, et la fixation du CO<sub>2</sub> par la REFORESTATION à grande échelle, sont mieux perçus (Campos\_captage-et-stockage\_2007).

**Fiche 5. Différence bilatérale**

**Terme anglais :** *global-ADJ*

**Équivalent CP**

**global-ADJ** (TleCP 415)

*mondial-ADJ* (TleCP 1505)

*planétaire-ADJ* (TleCP 278)

*globe-NOM* (TleCP 192)

**Équivalent CC**

**global-ADJ** (844)

*mondial-ADJ* (836)

*planétaire-ADJ* (136)

*globe-NOM* (101)

Lorsqu'on compare la fréquence de *global-ADJ* dans l'ensemble du CC et dans l'ensemble du CP (TleCP), nous remarquons un nombre anormalement élevé de cet équivalent dans le CC. La consultation des contextes du CC, nous montre que dans de nombreux contextes l'équivalent *global-ADJ* pourrait être avantageusement remplacé par *mondial-ADJ*. Selon Van Roey *et al.* (1991) et Termium, *global-ADJ* est à éviter lorsqu'il prend le sens de 'qui est relatif à l'ensemble de la planète'. Les contextes suivants relevés dans le CC sont des exemples dans lesquels *mondial-ADJ* serait plus approprié.

CC : Si le changement est GLOBAL, les impacts sont bien souvent locaux... (Boe\_hydrologie\_2007).

CC : ...une influence perceptible de l'homme sur le climat GLOBAL (CNRS-lettre14\_2002).

CC : ... le changement climatique GLOBAL auquel l'humanité se trouve désormais confrontée...(Senat\_ampleur\_2002)

CC : La modification de la composition de l'atmosphère entraîne un réchauffement GLOBAL (Senat\_energie\_2006).

**Fiche 6. Différence unilatérale**

**Termes anglais :** *mitigation-NOM*

**Équivalent CP**

*atténuer-VER*

*atténuation-NOM*

**Équivalent CC**

*atténuer-VER*

*atténuation-NOM*

**mitigation-NOM** (35)

**mitiger-VER** (4)

Dans Termium, le terme *mitigation-NOM* est à éviter. La Commission générale de terminologie et de néologie (France) l'accepte seulement dans le domaine des sciences et des technologies spatiales, domaine qui n'est pas relié aux changements climatiques. L'équivalent typique du terme anglais est *atténuation-NOM*. Le premier contexte suivant appartient au CP, le deuxième au CC.

CP : ...Ainsi, l'ATTÉNUATION des émissions de ces autres gaz est un élément important... (IPCC\_statilisationdesgaz\_1997).

CC : ...pour débattre autour d'interrogations diverses portant sur le changement climatique et les solutions de MITIGATION des émissions de CO2 (Campos\_captage-et-stockage\_2007).

**Fiche 7. Différence unilatérale****Termes anglais :** *mitigate*-VER**Équivalent CP***atténuer*-VER*atténuation*-NOM**Équivalent CC***atténuer*-VER*atténuation*-NOM***mitiger*-VER (4)*****mitigation*-NOM (35)**

Le terme anglais *mitigate*-VER n'est pas présent dans les banques de terminologie, mais il est morphologiquement relié au terme *mitigate*-NOM, par conséquent son équivalent typique est *atténuer*-VER. Le premier contexte suivant appartient au CP, le deuxième au CC.

CP : Le piégeage et le stockage du CO2 sont en particulier replacés dans le contexte des autres solutions envisageables pour ATTÉNUER les changements climatiques (IPCC\_dioxyde\_2005).

CC : ...sont favorables à l'utilisation de la séquestration du CO2 pour MITIGER le changement climatique (Campos\_captage-et-stockage\_2007).

**Fiche 8. Différence bilatérale****Termes anglais :** *monitoring*-NOM**Équivalent CP***surveillance*-NOM*suivi*-NOM*contrôle*-NOM*surveiller*-VER***observation*-NOM*****évaluation*-NOM****Équivalent CC***surveillance*-NOM*suivi*-NOM*contrôle*-NOM*surveiller*-VER***monitoring*-NOM (4)**

Selon Termium *monitoring* est à éviter. La Commission générale de terminologie et de néologie (France) propose *monitorage* au lieu de l'équivalent étranger *monitoring*. Les deux contextes suivants montrent *surveillance*-NOM dans le CP et *monitoring*-NOM dans le CC.

CP : ...qui combine protection de l'habitat, SURVEILLANCE et recherche (RNCA\_impacts\_2008).

CC : Les trois principales catégories de coûts sont le MONITORING, les travaux de protection de la côte et de recherche (ONERC\_impact\_2008).

**Fiche 9. Différence unilatérale****Termes anglais :** *offshore*-ADJ**Équivalent CP***en mer*-LOC*au/du large*-LOC**Équivalent CC***en mer*-LOC*au/du large*-LOC***offshore*-ADJ (31)**

Selon Termium *offshore*-ADJ est un anglicisme. La Commission générale de terminologie et de néologie (France) précise que ce terme est un équivalent étranger et propose la locution *en mer*. Les contextes suivants illustrent *offshore*-ADJ dans le CC et *en mer*-LOC dans le CP.

CP : l'exploration et la production pétrolière EN MER sont sensibles aux changements... (RNCA\_impacts\_2008).

CC : Des réserves prouvées de pétrole se trouvent en OFFSHORE (Senat\_energie\_2006).

**Fiche 10. Différence bilatérale**Termes anglais : *permafrost*-NOM**Équivalent CP***pergélisol*-NOM (308)***permafrost*-NOM** (16)***pergélisolé*-ADJ****Équivalent CC***pergélisol*-NOM (96)***permafrost*-NOM** (87)

Comme l'indiquent les chiffres entre parenthèses, l'équivalent *permafrost*-NOM est pratiquement aussi fréquent que *pergélisol*-NOM dans le CC alors que dans le CP la fréquence de *permafrost*-NOM est de loin supérieure à celle de *permafrost*-NOM. Selon Termium, *permafrost* est un anglicisme à éviter. La Commission générale de terminologie et de néologie (France), quant à elle, admet *permafrost*, mais recommande *permagel* ou *pergélisol*.

Les 10 termes pour lesquels il y a des équivalents critiquables dans le CC sont soit des cognats dont la fréquence se démarque dans le CC par rapport à celle du CP, soit des anglicismes. Par conséquent, les termes anglais pour lesquels il y a des équivalents critiquables sont plus fréquents dans le CC (10) que dans le CP (2).

Au sujet des anglicismes, on pourrait se demander si la variation régionale entre en ligne de compte dans les résultats. En ce qui concerne le CC, les anglicismes ont été trouvés uniquement dans des textes d'origine européenne. Pour ce qui est du CP, étant donné que les textes du corpus parallèle sont majoritairement internationaux, il est raisonnable de supposer que les traducteurs sont en grande partie d'origine européenne. Après, avoir vérifié si les notions pour lesquelles il y a eu des équivalents critiquables dans le CC se trouvaient également dans les textes internationaux du CP, nous avons constaté que les traducteurs de ces textes n'ont pas utilisé les équivalents critiquables trouvés dans le CC.

### 7.3 Conclusion du chapitre

Dans ce chapitre, nous avons tout d'abord décrit la méthodologie suivie pour effectuer l'analyse comparative des équivalents dans le corpus parallèle et le corpus comparable (7.1). Ensuite, à la section 7.2.1, nous avons constaté que sur les 343 termes anglais analysés 272 étaient des termes avec équivalents identiques, 67 des termes avec des équivalents différents et seulement 4 termes anglais n'avaient pas d'équivalent dans le corpus comparable. Dans cette même section, nous avons observé que les termes

sélectionnés avec le critère a) et les termes appartenant à la partie du discours du nom étaient relativement plus fréquents dans la catégorie « équivalents différents ».

L'analyse des termes avec des équivalents identiques (7.2.2) nous a appris, plus particulièrement, que le nombre d'équivalents pour un même terme n'est pas un facteur déterminant pour qu'un terme soit représenté par des équivalents identiques dans les deux corpus. Autrement dit, un terme anglais avec plusieurs équivalents a tout autant de chance qu'un terme avec un seul équivalent d'avoir des équivalents identiques dans les deux corpus.

Nous avons également vu que 67 termes anglais présentent des différences d'équivalence dans les deux corpus (7.2.3). Parmi ces termes, 55 appartiennent à la catégorie des termes avec équivalents non critiquables ( $\pm$ ). Il y a donc au total 12 termes anglais pour lesquels nous avons trouvé des équivalents à critiquer ( $\neq$ CP,  $\neq$ CC), 2 dans le CP et 10 dans le CC, soit seulement 3,5 % des 339 termes anglais ayant des équivalents dans les deux corpus.

Les équivalents des termes  $\pm$  ont la particularité d'être généralement des équivalents non typiques et peu fréquents. En outre, la très grande majorité se présente sous forme de différences d'équivalence unilatérale plutôt que de différences bilatérales. Les 2 termes  $\neq$ CP ont des équivalents très peu fréquents (11 occurrences ou moins). Enfin, parmi les 10 termes avec équivalents critiquables 4 sont très peu fréquents et 6 sont relativement fréquents à très fréquents.

Considérant le nombre de termes analysés (343), les termes présentant des équivalents critiquables dans les deux corpus sont relativement peu élevés (12), tandis que le nombre de termes présentant des similitudes d'équivalence entre les corpus est très élevé (272 équivalents identiques et 55 équivalents non critiquables). Par conséquent, l'analyse comparative décrite dans ce chapitre confirme notre hypothèse selon laquelle la terminologie employée dans les traductions de textes spécialisés ne se démarque pas autant qu'on ne le suppose généralement de celle des textes originaux. D'ailleurs, il est à noter que c'est dans le corpus comparable que nous avons trouvé le plus grand nombre d'équivalents critiquables, ce qui contredit les jugements négatifs à l'endroit de l'utilisation des textes

traduits. Les auteurs des textes originaux, tout autant que les traducteurs, doivent se méfier des pièges interlinguistiques. L'introduction des emprunts des faux sens et des termes à éviter n'est pas juste le fait des traducteurs comme on le suppose généralement. Les auteurs de textes scientifiques lisent de nombreux ouvrages et participent à des colloques en anglais. De ce fait, il est normal qu'ils soient tout autant influencés par la terminologie anglaise que les traducteurs.

## **DEUXIÈME VOLET :**

ANALYSE COMPARATIVE DES ÉQUIVALENTS  
FRANÇAIS ASSOCIÉS À UN MÊME TERME ANGLAIS

## **8 MÉTHODOLOGIE ET RÉSULTATS DE L'ANALYSE COMPARATIVE DES ÉQUIVALENTS FRANÇAIS ASSOCIÉS À UN MÊME TERME ANGLAIS**

Le deuxième volet de la recherche concerne l'analyse comparative des équivalents associés à un même terme anglais. Au Chapitre 6, nous avons observé que 67,1 % des termes anglais ont plusieurs équivalents, que ce soit dans le corpus parallèle ou le corpus comparable (*cf.* 6.2 et 6.3). Dans cette section, nous nous interrogeons tout particulièrement sur la nature de ces équivalents. Est-ce que les équivalents d'un même terme anglais sont généralement des synonymes? Sinon, quelle est leur caractéristique? Quel lien entretiennent-ils avec l'équivalent typique dans les dictionnaires?

À partir de ce point, notre étude ne tient plus compte de la distinction corpus parallèle et corpus comparable<sup>112</sup>, mais de la différence entre les équivalents français d'un même terme anglais. La section 8.1 décrit la méthode suivie pour analyser les équivalents d'un même terme et la section 8.2 présente les résultats de cette analyse. Enfin, à la section 8.3, nous voyons les ÉqT dont la partie du discours est différente du terme anglais et nous apportons des observations sur les équivalents des adverbes et sur les équivalents des éléments et composés chimiques.

### **8.1 Méthodologie de l'analyse comparative des équivalents associés à un même terme anglais**

Tout d'abord, pour comparer les équivalents d'un même terme anglais entre eux, nous nous appuyons sur la notion d'équivalent typique (ÉqT) et d'équivalent non typique (ÉqNT) abordée dans le Chapitre 5 (*cf.* 5.5.2). À titre d'exemple, dans le Tableau 8.1, nous voyons que des termes anglais peuvent être associés à un (*desertification*-NOM) ou plusieurs équivalents (*plant*-NOM, *simulate*-VER, *biodiversity*-NOM). Par exemple, le terme anglais

---

<sup>112</sup> À plus forte raison que dans le premier volet de l'analyse, nous en sommes arrivée à la conclusion que les différences qualitatives d'équivalence entre les deux types de corpus sont minimales.



*plant*-NOM a pour équivalents *plante*-NOM, *végétal*-ADJ, *végétal*-NOM et *flore*-NOM. Toutefois, dans les dictionnaires, ces équivalents ne sont pas tous des ÉqT, des quatre équivalents du terme anglais seul l'équivalent *plante*-NOM peut être qualifié d'ÉqT, les autres sont des ÉqNT. De même, *simulate*-VER a pour ÉqT *simulate*-VER et pour ÉqNT *simulé*-ADJ et *simulation*-NOM. Le terme anglais *biodiversity*-NOM, quant à lui, est associé à deux ÉqT, *biodiversité*-NOM et *diversité biologique*-NOM. Enfin, le terme anglais *desertification*-NOM est associé seulement à un ÉqT.

Tableau 8.1. Exemples de termes anglais et de leurs équivalents

Terme anglais	Équivalents	ÉqT/ÉqNT
<i>plant</i> -NOM	<i>plante</i> -NOM <i>végétal</i> -ADJ <i>végétal</i> -NOM <i>flore</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT
<i>simulate</i> -VER	<i>simuler</i> -VER <i>simulé</i> -ADJ <i>simulation</i> -NOM	ÉqT ÉqNT ÉqNT
<i>biodiversity</i> -NOM	<i>biodiversité</i> -NOM <i>diversité biologique</i> -NOM	ÉqT ÉqT
<i>desertification</i> -NOM	<i>désertification</i> -NOM	ÉqT

Une fois les équivalents des 343 termes anglais identifiés en tant qu'ÉqT ou ÉqNT<sup>113</sup>, nous les avons classés selon la typologie présentée au Tableau 8.2. La typologie que nous avons adoptée mesure en quelque sorte l'écart d'équivalence entre l'équivalent typique et les autres équivalents trouvés pour un même terme anglais. Autrement dit, nous voulons avoir une idée du degré de liberté ou de contrainte dont disposent les auteurs quant au choix des termes pour exprimer une même notion. Pour effectuer la classification, nous nous appuyons sur ce que nous avons observé dans les dictionnaires de l'Annexe C. Ainsi, dans le groupe des équivalents appartenant à la même partie du discours, les ÉqT *biodiversité*-NOM et *diversité biologique*-NOM sont qualifiés d'équivalents synonymes, alors que l'ÉqNT *flore*-NOM par rapport à l'ÉqT *plante*-NOM est un holonyme (Tableaux 8.1 et 8.2). De leur côté, dans le groupe des équivalents transcatégoriels, l'ÉqNT

<sup>113</sup> Voir les tableaux de l'Annexe E et de l'Annexe F.

*végétal*-ADJ est un équivalent transcatégoriel par rapport à l'ÉqT *plante*-NOM et les ÉqNT *simulé*-ADJ et *simulation*-NOM sont des équivalents transcatégoriels morphologiquement reliés à l'ÉqT *simuler*-VER (Tableaux 8.1 et 8.2).

Les équivalents des 99 termes anglais ne comportant qu'un seul et même équivalent dans les deux corpus (cf. 7.2.2), par exemple *désertification*-NOM (Tableau 8.1), ne font pas partie de cette analyse étant donné qu'ils ne peuvent être comparés à eux-mêmes et qu'ils sont tous des ÉqT. De même, les équivalents des 3 termes anglais *biota*-NOM, *biome*-NOM et *unsustainable*-NOM (cf. 7.2.1) qui n'ont qu'un seul équivalent dans le CP et aucun dans le CC ne sont pas pris en compte dans notre analyse. En conséquence, sur les 343 termes anglais, 241 termes sont analysés dans le deuxième volet de l'étude, soit 70,3 % des termes.

Tableau 8.2. Typologie des équivalents français associés à un même terme anglais

<b>Équivalent appartenant à la même partie du discours que l'ÉqT</b>	
1	Équivalents synonymes
2	Équivalent quasi-synonyme par rapport à l'ÉqT
3	Équivalent hyperonyme par rapport à l'ÉqT
4	Équivalent hyponyme par rapport à l'ÉqT
5	Équivalent holonyme par rapport à l'ÉqT
6	Équivalent méronyme par rapport à l'ÉqT
7	Équivalent dont le système conceptuel présente une divergence par rapport à celui de l'autre langue
8	Équivalent présentant un autre point de vue par rapport à l'ÉqT
<b>Équivalent transcatégoriel par rapport à l'ÉqT</b>	
1	Équivalent transcatégoriel par rapport à l'ÉqT et morphologiquement relié à celui-ci
2	Équivalent transcatégoriel par rapport à l'ÉqT, mais relié non morphologiquement à l'ÉqT

Dans les sections qui suivent, nous décrivons en détail la typologie du Tableau 8.2. La section 8.1.1 se penche sur le groupe des équivalents de la même partie du discours que l'ÉqT. La section 8.1.2 s'intéresse au groupe des équivalents transcatégoriels par rapport à l'ÉqT. Ces deux groupes, comme le montre le Tableau 8.2, sont eux-mêmes subdivisés en sous-groupes. Par ailleurs, nous avons exclu de l'analyse les 8 équivalents anglicismes ou

faux-amis<sup>114</sup>, car ces équivalents ont été jugés fautifs lors de l'analyse comparative décrite au Chapitre 7 (cf. 7.2.3.2 et 7.2.3.3).

### 8.1.1 Équivalent de la même partie du discours que l'ÉqT

Le groupe des équivalents qui appartiennent à la même partie du discours que le ou les équivalents typiques se divise en 8 catégories (voir Tableau 8.2). Dans les sections qui suivent, nous définissons chacune des catégories, nous présentons quelques exemples dans un tableau, puis nous illustrons la catégorie à l'aide de quelques contextes.

#### 8.1.1.1 Équivalents synonymes

« Fondamentalement, la relation de *synonymie* est celle qu'entretiennent deux unités lexicales (deux termes, dans le cas qui nous intéresse) qui ont le même sens. » (L'Homme 2004 : 94). « Il y a synonymie lorsque deux ou plusieurs termes différents sont affectés à une seule notion » (Felber : 1987 : 153). Dans cette thèse, à l'instar de l'approche lexico-sémantique ou de Felber, nous ne tenons pas compte des marques d'usage dans notre définition de la synonymie. Pour déterminer si deux équivalents sont synonymes, nous nous appuyons sur les ressources terminologiques et autres dictionnaires cités à l'Annexe C. Les équivalents synonymes appartiennent nécessairement à la même partie du discours et sont tous des ÉqT du terme anglais. Le Tableau 8.3 présente quelques synonymes relevés dans le corpus.

Tableau 8.3. Quelques exemples de synonymes relevés dans le corpus

Terme anglais	Équivalent synonyme (ÉqT)	Équivalent synonyme (ÉqT)	Source
<i>biodiversity</i> -NOM	<i>biodiversité</i> -NOM	<i>diversité biologique</i> -NOM	Termium
<i>capture</i> -NOM	<i>capture</i> -NOM	<i>piégeage</i> -NOM	Brodhag et Breuil (2009 : 22)
<i>capture</i> -NOM	<i>captage</i> -NOM	<i>piégeage</i> -NOM	Brodhag et Breuil (2009 : 22)
<i>chemical</i> -NOM	<i>produit chimique</i> -NOM	<i>substance chimique</i> -NOM	IATE
<i>diversion</i> -NOM	<i>dérivation</i> -NOM	<i>détournement</i> -NOM	Termium

<sup>114</sup> *Refléter*-VER, *anthropogénique*-NOM, *anthropogène*-NOM, *mitigation*-NOM, *mitiger*-VER, *global*-ADJ, *monitoring*-NOM et *offshore*-ADJ.

<i>drought</i> -NOM	<i>période de sécheresse</i> -NOM	<i>sécheresse</i> -NOM	Termium
<i>frost</i> -NOM	<i>gel</i> -NOM	<i>gelée</i> -NOM	Termium
<i>global</i> -ADJ	<i>mondial</i> -ADJ	<i>planétaire</i> -ADJ	Termium
<i>nutrient</i> -NOM	<i>élément nutritif</i> -NOM	<i>nutriant</i> -NOM	Termium
<i>nutrient</i> -NOM	<i>élément nutritif</i> -NOM	<i>nutriment</i> -NOM	Termium
<i>permafrost</i> -NOM	<i>pergélisol</i> -NOM	<i>permafrost</i> -NOM	FranceTerme
<i>resilience</i> -NOM	<i>capacité d'adaptation</i> -NOM	<i>résilience</i> -NOM	IATE
<i>subsidence</i> -NOM	<i>affaissement de terrain</i> -NOM	<i>subsidence</i> -NOM	Termium
<i>thaw</i> -NOM	<i>dégel</i> -NOM	<i>fonte</i> -NOM	Termium
<i>watershed</i> -NOM	<i>bassin</i> -NOM	<i>bassin versant</i> -NOM	C.N.F.S.H.
<i>wetland</i> -NOM	<i>milieu humide</i> -NOM	<i>zone humide</i> -NOM	Termium
<i>wildfire</i> -NOM	<i>feu de forêt</i> -NOM	<i>incendie de forêt</i> -NOM	Termium
<i>world</i> -NOM	<i>globe</i> -NOM	<i>monde</i> -NOM	Petit Robert

Les contextes suivants montrent les synonymes *mondial*-ADJ (2) et *planétaire*-ADJ (2) du terme anglais *global*-ADJ (1). On remarquera que dans cette phrase, le même terme est utilisé en anglais (*global*-NOM), alors qu'en français, probablement pour éviter une répétition, deux termes synonymes sont employés (*mondial*-NOM et *planétaire*-NOM).

- 1) It provides information relevant to decision-making in regard to safeguarding the ozone layer and the GLOBAL climate system: two GLOBAL environmental issues involving complex scientific and technical considerations. (IPCC\_ozone\_2005)
- 2) Il renferme des informations utiles à la prise de décision pour la sauvegarde de la couche d'ozone et la protection du système climatique MONDIAL : deux enjeux environnementaux PLANÉTAIRES ayant diverses implications complexes, tant sur le plan scientifique que sur le plan technique. (IPCC\_ozone\_2005)

Les phrases suivantes montrent que les équivalents *pergélisol*-NOM (1) et *permafrost*-NOM (2) sont synonymes. Le premier contexte provient de la traduction d'un texte canadien, le deuxième d'un document rédigé à l'intention du Sénat français.

- 1) Une fonte à grande échelle du PERGÉLISOL entraîne des répercussions mondiales, car le processus libérera des GES (méthane) dans l'atmosphère contribuant davantage au réchauffement de l'atmosphère. (NRTEE\_true\_north\_2009)

- 2) Autre menace, elle aussi redoutable, la fonte du PERMAFROST pourrait également libérer le méthane qu'il contient. (Senat\_energie\_2006)

### 8.1.1.2 Équivalent quasi-synonymes par rapport à l'ÉqT

« [...] deux *quasi-synonymes* sont des unités lexicales partageant une partie de leurs composantes sémantiques, mais pas toutes. » (L'Homme 2004 : 295). Pour valider les quasi-synonymes du corpus, nous nous sommes appuyée sur les ressources terminologiques et autres dictionnaires généraux de l'Annexe C. Notons au passage, que les sources terminologiques offrent rarement des quasi-synonymes, le GDT étant la seule source terminologique de l'Annexe C qui le fasse. De plus, cette catégorie n'est pas nécessairement exclusive. En effet, un quasi-synonyme figurera en priorité dans l'une des catégories suivante s'il remplit les conditions de sélection : hyperonymes, hyponymes et « équivalent dont le système conceptuel présente une divergence par rapport à celui de l'autre langue ». Le Tableau 8.4 offre quelques exemples de quasi-synonymes trouvés dans le corpus.

Tableau 8.4. Exemples de quasi-synonymes tirés du corpus

Terme anglais	Équivalent quasi-synonyme	Équivalent typique	Source
<i>change</i> -NOM	<i>variation</i> -NOM	<i>changement</i> -NOM	Antidote
<i>coastline</i> -NOM	<i>côte</i> -NOM	<i>littoral</i> -NOM	IATE
<i>consequence</i> -NOM	<i>répercussion</i> -NOM	<i>conséquence</i> -NOM	Antidote
<i>consequence</i> -NOM	<i>effet</i> -NOM	<i>conséquence</i> -NOM	Antidote
<i>consequence</i> -NOM	<i>incidence</i> -NOM	<i>conséquence</i> -NOM	Antidote
<i>contamination</i> -NOM	<i>pollution</i> -NOM	<i>contamination</i> -NOM	Antidote
<i>damage</i> -NOM	<i>dégât</i> -NOM	<i>dommage</i> -NOM	Antidote
<i>decrease</i> -VER	<i>baisser</i> -VER	<i>diminuer</i> -VER	Antidote
<i>degradation</i> -NOM	<i>détérioration</i> -NOM	<i>dégradation</i> -NOM	Antidote
<i>disaster</i> -NOM	<i>catastrophe</i> -NOM	<i>désastre</i> -NOM	Antidote
<i>effect</i> -NOM	<i>impact</i> -NOM	<i>effet</i> -NOM	Antidote
<i>effect</i> -NOM	<i>conséquence</i> -NOM	<i>effet</i> -NOM	Antidote
<i>environment</i> -NOM	<i>milieu</i> -NOM	<i>environnement</i> -NOM	Antidote
<i>extinction</i> -NOM	<i>disparition</i> -NOM	<i>extinction</i> -NOM	Antidote
<i>fire</i> -NOM	<i>incendie</i> -NOM	<i>feu</i> -NOM	Petit Robert
<i>flooding</i> -NOM	<i>débordement</i> -NOM	<i>inondation</i> -NOM	Antidote
<i>food</i> -NOM	<i>nourriture</i> -NOM	<i>aliment</i> -NOM	Antidote
<i>fragmentation</i> -NOM	<i>morcellement</i> -NOM	<i>fragmentation</i> -NOM	Antidote
<i>grassland</i> -NOM	<i>pâturage</i> -NOM	<i>prairie</i> -NOM	Antidote

<i>harmful</i> -ADJ	<i>néfaste</i> -ADJ	<i>nocif</i> -ADJ	Antidote
<i>hazard</i> -NOM	<i>risque</i> -NOM	<i>danger</i> -NOM	Antidote
<i>impact</i> -NOM	<i>incidence</i> -NOM	<i>impact</i> -NOM	Antidote
<i>increase</i> -NOM	<i>hausse</i> -NOM	<i>augmentation</i> -NOM	Antidote
<i>increase</i> -NOM	<i>accroissement</i> -NOM	<i>augmentation</i> -NOM	Antidote
<i>negative</i> -ADJ	<i>néfaste</i> -ADJ	<i>néгатif</i> -ADJ	Antidote
<i>projection</i> -NOM	<i>prévision</i> -NOM	<i>projection</i> -NOM	Antidote
<i>region</i> -NOM	<i>zone</i> -NOM	<i>région</i> -NOM	Petit Robert
<i>scarcity</i> -NOM	<i>rareté</i> -NOM	<i>pénurie</i> -NOM	Petit Robert
<i>scenario</i> -NOM	<i>hypothèse</i> -NOM	<i>scénario</i> -NOM	Antidote
<i>soil</i> -NOM	<i>terre</i> -NOM	<i>sol</i> -NOM	Antidote
<i>storage</i> -NOM	<i>entreposage</i> -NOM	<i>stockage</i> -NOM	Antidote
<i>variation</i> -NOM	<i>fluctuation</i> -NOM	<i>variation</i> -NOM	Antidote
<i>vary</i> -VER	<i>fluctuer</i> -VER	<i>varier</i> -VER	Antidote
<i>vulnerable</i> -ADJ	<i>fragile</i> -ADJ	<i>vulnérable</i> -ADJ	Petit Robert

Les quasi-synonymes s’emploient pour diverses raisons, en particulier pour éviter les répétitions. Par exemple, le terme anglais *consequence*-NOM peut avoir pour équivalent *conséquence*-NOM (1), *effet*-NOM (2), *répercussion*-NOM (3) ou *incidence*-NOM (4).

- 1) À terme, les CONSÉQUENCES du changement climatique entraînent des modifications des propriétés et des rendements des sols. (CES\_agriculture\_2008)
- 2) Les probabilités de déclenchement de ces événements sont mal comprises, mais on devrait en tenir compte, vu la gravité de leurs EFFETS. (IPCC\_bilan\_2001)
- 3) Ces RÉPERCUSSIONS seront considérables, et c'est probablement sur les populations les plus vulnérables que leur poids se fera le plus sentir. (ONU\_overview\_2008)
- 4) Les INCIDENCES sur le climat et sur le niveau de la mer de telles incertitudes sont extrêmement faibles; (IPCC\_emissions\_1997)

### 8.1.1.3 Équivalent hyperonyme par rapport à l'ÉqT

Un *hyperonyme* désigne un générique (L’Homme 2004 : 91) – c’est-à-dire que l’hyperonyme est une lexie dont le sens est moins riche que celui de ses hyponymes (Polguère 2003 : 120). Par exemple, le terme *fruit* est un hyperonyme de l’hyponyme

*pomme*. L'hyperonyme entretient avec son ou ses hyponymes un rapport de générique à spécifique. Le Tableau 8.5 montre des exemples d'hyperonymes tirés du corpus.

Tableau 8.5. Exemples d'hyperonymes trouvés dans le corpus

Terme anglais	Équivalent hyperonyme	Équivalent typique
<i>electricity</i> -NOM	<i>énergie</i> -NOM	<i>électricité</i> -NOM
<i>forest</i> -NOM	<i>peuplement</i> -NOM	<i>forêt</i> -NOM
<i>migrate</i> -VER	<i>déplacer (se)</i> -VER	<i>migrer</i> -VER
<i>migration</i> -NOM	<i>déplacement</i> -NOM	<i>migration</i> -NOM
<i>plant</i> -NOM	<i>végétal</i> -NOM	<i>plante</i> -NOM
<i>rainfall</i> <sub>1</sub> -NOM (chute de pluie)	<i>précipitation</i> -NOM	<i>pluie</i> -NOM
<i>river</i> -NOM	<i>cours d'eau</i> -NOM	<i>rivière</i> -NOM <i>fleuve</i> -NOM
<i>runoff</i> -NOM	<i>écoulement</i> -NOM	<i>ruissellement</i> -NOM

Dans l'exemple suivant, nous constatons que dans le contexte français l'équivalent hyperonyme *énergie*-NOM (1) remplace le terme typique *électricité*-NOM.

- 1) Les futures répercussions sur le transport, le transport d'ÉNERGIE et les communications auront des conséquences d'ordre général, (RNCA\_impacts\_2008)
- 2) Future disruptions to transportation, ELECTRICITY transmission and communication will have widespread implications, (RNCA\_impacts\_2008)

De même, dans le contexte français suivant, le traducteur a eu recours à l'hyperonyme *précipitations*-NOM (1) au lieu de l'équivalent typique *pluie*-NOM.

- 1) Chaleurs, sécheresses, PRÉCIPITATIONS et inondations extrêmes affectent de nombreuses régions en Europe. (AEE\_signals\_2009)
- 2) Extremes of heat and drought, RAIN and flooding are affecting many parts of Europe. (AEE\_signals\_2009)

#### 8.1.1.4 Équivalent hyponyme par rapport à l'ÉqT

Inversement au cas précédent, un *hyponyme* désigne un spécifique. « L'hyponyme possède toutes les composantes sémantiques de l'hyperonyme et a une ou quelques composantes supplémentaires » (L'Homme 2004 : 92). On peut donc dire que son sens est plus riche que celui de son hyperonyme (Polguère 2003 : 120). Par exemple, le terme *chien* est un hyponyme de l'hyperonyme *canidé*. Le Tableau 8.6 illustre des exemples d'hyponymes trouvés dans le corpus.

Tableau 8.6. Exemples d'hyponymes trouvés dans le corpus

Terme anglais	Équivalent hyponyme	Équivalent typique
<i>agriculture-NOM</i>	<i>culture-NOM</i>	<i>agriculture-NOM</i>
<i>groundwater-NOM</i>	<i>nappe phréatique-NOM</i>	<i>eau souterraine-NOM</i>
<i>migration-NOM</i>	<i>exode-NOM</i>	<i>migration-NOM</i>
<i>pest-NOM</i>	<i>animal nuisible-NOM</i>	<i>ravageur-NOM</i> (insecte, animal, etc. qui détruit)
<i>road-NOM</i>	<i>chemin-NOM</i>	<i>route-NOM</i>
<i>specie-NOM</i>	<i>essence-NOM</i> (espèce d'un arbre)	<i>espèce-NOM</i>
<i>stream-NOM</i>	<i>rivière-NOM</i>	<i>cours d'eau-NOM</i>
<i>wetland-NOM</i>	<i>marais-NOM</i>	<i>zone humide-NOM</i>

Le contexte français suivant montre que dans certaines situations le terme anglais *energy-NOM* (2) peut être traduit par l'hyponyme *électricité-NOM* (1).

- 1) Ce processus valorise la chaleur dégagée lors de la production d'ÉLECTRICITÉ (AEE\_signals\_2009)
- 2) This process harnesses the heat that is a by-product of ENERGY production. (AEE\_signals\_2009)

Dans l'exemple suivant, nous voyons que l'équivalent hyponyme *pluie-NOM* (1) remplit, pour ce contexte, le même rôle que son hyperonyme *précipitation-NOM* (2).

- 1) (notamment les PLUIES diluviennes au Venezuela en 1999 et 2003),... (IPCC\_eau\_2008)



- 2) (e.g., heavy PRECIPITATIONS in Venezuela 1999 and 2005);... (IPCC\_eau\_2008)

### 8.1.1.5 Équivalent holonyme par rapport à l'ÉqT

Dans une relation méronymique<sup>115</sup>, un *holonyme* est un terme qui désigne un tout (L'Homme 2004 : 98). Par exemple le terme *bicyclette* est l'holonyme du terme *guidon*. Le Tableau 8.7 illustre des exemples d'holonymes relevés dans le corpus.

Tableau 8.7. Exemples d'holonymes trouvés dans le corpus

Terme anglais	Équivalent holonyme	Équivalent typique
<i>animal-NOM</i>	<i>faune-NOM</i>	<i>animal-NOM</i>
<i>human-NOM</i>	<i>population-NOM</i>	<i>humain-NOM</i>
<i>plante-NOM</i>	<i>flore-NOM</i>	<i>plante-NOM</i>
<i>tree-NOM</i>	<i>forêt-NOM</i>	<i>arbre-NOM</i>

Le contexte suivant montre que les équivalents holonymes *flore-NOM* (1) et *faune-NOM* (1) peuvent remplacer avantageusement les ÉqT *plante-NOM* et *animal-NOM*. Généralement, les holonymes sont utilisés à la place d'un ÉqT qui serait mis au pluriel.

- 1) Des facteurs climatiques influent également sur la productivité de la FAUNE et de la FLORE et sur d'autres fonctions des écosystèmes. (IPCC\_biodiversité\_2002)
- 2) Climatic factors also affect PLANT and ANIMAL productivity and other ecosystem functions. (IPCC\_biodiversité\_2002)

### 8.1.1.6 Équivalent méronyme par rapport à l'ÉqT

Dans une relation méronymique, un méronyme est un terme qui désigne la partie d'un tout (L'Homme 2004 : 98). Par exemple, le terme *roue* est un méronyme du terme *voiture* (Cruse 1986 : 157–180). Le Tableau 8.8 illustre trois exemples de méronymes relevés dans le corpus.

<sup>115</sup> « Les relations méronymiques sont fondées sur les notions vagues de proximité ou d'association dans l'espace. Elles engagent un terme dénotant un tout et un ou plusieurs termes dénotant des parties » (L'Homme 2004 : 98).

Tableau 8.8. Trois exemples de méronymes relevés dans le corpus

Terme anglais	Équivalent méronyme	Équivalent typique
<i>livestock</i> -NOM	<i>animal</i> -NOM	<i>bétail</i> -NOM
<i>forest</i> -NOM	<i>arbre</i> -NOM	<i>forêt</i> -NOM
<i>aerosol</i> -NOM	<i>particule</i> -NOM	<i>aérosol</i> -NOM

Les équivalents méronymes sont généralement au pluriel dans le texte cible. Dans l'exemple suivant, le terme anglais *forest*-NOM (1) est rendu par l'équivalent méronyme *arbre*-NOM (2) au pluriel au lieu de l'équivalent typique *forêt*-NOM.

- 1) ...there are still large tracts of dead or dying FOREST, killed by budworm, that pose a substantial fire hazard. (RNCA\_impacts\_2008)
- 2) ...il reste encore de vastes zones ravagées par cet insecte où les ARBRES morts ou moribonds constituent un risque d'incendie. (RNCA\_impacts\_2008)

Dans le contexte français suivant, c'est l'équivalent méronyme *animal*-NOM (1) au pluriel qui remplace l'ÉqT *bétail*-NOM.

- 1) (comme l'indique la flèche bidirectionnelle entre la production des végétaux et des ANIMAUX), puisque les cultures nourrissent les ANIMAUX et que les déchets animaux servent d'engrais. (CCE\_trends\_2003)
- 2) (as the bidirectional arrow between crop and LIVESTOCK agriculture indicates) since crops are used to feed LIVESTOCK and animal waste is used to fertilize crops. (CCE\_trends\_2003)

#### 8.1.1.7 Équivalent dont le système conceptuel présente une divergence par rapport à celui de l'autre langue

Comme expliqué dans le Chapitre 2, il arrive qu'une notion soit désignée par un seul terme dans une langue alors qu'elle est représentée par au moins deux termes dans une autre (Svensén 2009 : 254). À cette occasion, nous avons donné l'exemple du terme anglais *river*-NOM dont les équivalents typiques sont *rivière*-NOM ou *fleuve*-NOM (Figure 2.2). Afin de choisir le bon équivalent, il faut que le contexte soit suffisamment explicite, sinon,

le traducteur doit utiliser, lorsqu'il existe, un hyperonyme, *cours d'eau*-NOM dans ce cas-ci. En plus du terme *river*-NOM, le Tableau 8.9 présente deux autres termes anglais appartenant à cette catégorie, *biofuel*-NOM et *fuel*-NOM. En ce qui concerne le terme anglais *biofuel*-NOM, IATE fait l'observation suivante « The English term “biofuel” corresponds in some languages (e.g. French, problem language) to both “biocombustible” [...] and “biocarburant” [IATE:860487], which are not synonyms ». De même, pour le terme *fuel*-NOM, IATE fait la même distinction entre *carburant*-NOM et *combustible*-NOM. Toutefois, ces deux termes anglais n'ont pas d'équivalent hyperonyme dans le corpus.

Tableau 8.9. Trois exemples de termes anglais dont le système conceptuel présente une divergence par rapport à celui de l'autre langue

Terme anglais	Équivalent
<i>river</i> -NOM	<i>rivière</i> -NOM <i>fleuve</i> -NOM
<i>biofuel</i> -NOM	<i>biocarburant</i> -NOM <i>biocombustible</i> -NOM
<i>fuel</i> -NOM	<i>combustible</i> -NOM <i>carburant</i> -NOM

Les contextes suivants illustrent ce type d'équivalence à l'aide du terme anglais *fuel*-NOM (1, 3) et des équivalents *combustible*-NOM (2) et *carburant*-NOM (4).

- 1) In addition to reducing losses from forest fires, the study suggests that these FUEL treatments may also increase the total annual allowable cut.
- 2) L'étude suggère qu'en plus de réduire les pertes dues aux feux de forêt, le traitement des matières COMBUSTIBLES pourrait augmenter la possibilité de coupe annuelle. (RNCA\_adaptation\_2004)
- 3) Work on the development of alternative FUELS for aviation shows promise and is continuing apace.
- 4) Les travaux de mise au point de CARBURANTS de substitution pour l'aviation sont prometteurs et se poursuivent à un rythme accéléré. (ONU\_overview\_2008)

### 8.1.1.8 Équivalent présentant un autre point de vue par rapport à l'ÉqT

Les équivalents de cette catégorie n'appartiennent à aucun des groupes précédents. Ces équivalents ne sont pas attestés dans les sources terminologiques et autres dictionnaires cités à l'Annexe C. Dans les énoncés traitant d'un même sujet, ces équivalents ont la particularité de présenter les choses sous un angle différent, sans pour autant changer le sens de l'idée principale. Nous avons observé ce phénomène autant dans le corpus parallèle que dans le corpus comparable.

Tableau 8.10. Exemples d'équivalents présentant un point de vue différent par rapport à l'ÉqT

Terme anglais	Équivalent	Équivalent typique
<i>assessment</i> -NOM	<i>étude</i> -NOM	<i>évaluation</i> -NOM
<i>concentration</i> -NOM (CO <sub>2</sub> )	<i>teneur</i> -NOM	<i>concentration</i> -NOM
<i>emit</i> -VER	<i>rejeter</i> -VER	<i>émettre</i> -VER
<i>environmental</i> -ADJ	<i>écologique</i> -ADJ	<i>environnemental</i> -ADJ
<i>extreme</i> -ADJ	<i>intense</i> -ADJ	<i>extrême</i> -ADJ
<i>forested</i> -ADJ	<i>forestier</i> -ADJ	<i>boisé</i> -ADJ
<i>forestry</i> -NOM	<i>forêt</i> -NOM	<i>foresterie</i> -NOM
<i>hazard</i> -NOM	<i>phénomène</i> -NOM	<i>danger</i> -NOM
<i>infestation</i> -NOM	<i>prolifération</i> -NOM	<i>infestation</i> -NOM
<i>intense</i> -ADJ	<i>grave</i> -ADJ	<i>intense</i> -ADJ
<i>intensity</i> -NOM	<i>intense</i> -NOM	<i>intensité</i> -NOM
<i>livestock</i> -NOM	<i>élevage</i> -NOM	<i>bétail</i> -NOM
<i>mapping</i> -NOM	<i>carte</i> -NOM	<i>cartographie</i> -NOM
<i>moisture</i> -NOM	<i>eau</i> -NOM	<i>humidité</i> -NOM
<i>mortality</i> -NOM	<i>décès</i> -NOM	<i>mortalité</i> -NOM
<i>ocean</i> -NOM	<i>mer</i> -NOM	<i>océan</i> -NOM
<i>outbreak</i> -NOM	<i>invasion</i> -NOM	<i>prolifération</i> -NOM
<i>release</i> -NOM	<i>émission</i> -NOM	<i>libération</i> -NOM
<i>release</i> -NOM	<i>rejet</i> -NOM	<i>libération</i> -NOM
<i>release</i> -VER	<i>rejeter</i> -VER	<i>libérer</i> -VER
<i>scarcity</i> -NOM	<i>raréfaction</i> -NOM	<i>pénurie</i> -NOM
<i>sea</i> -NOM	<i>océan</i> -NOM	<i>mer</i> -NOM
<i>stress</i> -NOM	<i>menace</i> -NOM	<i>stress</i> -NOM
<i>stress</i> -NOM	<i>agression</i> -NOM	<i>stress</i> -NOM
<i>uptake</i> -NOM	<i>fixation</i> -NOM	<i>absorption</i> -NOM

Dans les contextes traitant des principales causes de la présence des gaz à effet de serre dans l'atmosphère, les équivalents *libération*-NOM (1, 3) ou *rejet*-NOM (2, 4) sont fréquemment employés. Ces équivalents de sens différent ne changent pas l'idée véhiculée. Les exemples suivants montrent que ces deux équivalents peuvent être utilisés indifféremment, même si on a tendance à utiliser l'équivalent *rejet*-NOM lorsque les activités humaines sont en cause et *libération*-NOM pour les événements naturels.

- 1) Le changement d'affectation des terres - et plus notablement le déboisement - entraîne une augmentation des gaz à effet de serre, principalement par la LIBÉRATION du carbone stocké dans la biomasse. (PNUE\_bioseq\_2009)
- 2) Sous l'effet du réchauffement, les REJETS de carbone terrestre à partir des zones à pergélisol et les émissions de méthane en provenance des hydrates présents dans les sédiments côtiers contribueraient encore à augmenter la concentration des gaz à effet de serre dans l'atmosphère... (IPCC\_bilan\_2001)
- 3) Donc, les émissions liées aux LIBÉRATIONS des CFC et HCFC présents dans les stocks (comme les systèmes de réfrigération ou les mousses) ne sont traitées ni par le Protocole de Montréal ni par la Convention sur le climat et le Protocole de Kyoto. (IPCC\_ozone\_2005)
- 4) En février 2007, le GIEC notait que le réchauffement climatique résultait "vraisemblablement" du REJET de gaz à effet de serre par les activités humaines. (ONERC\_impact\_2008)

Les deux paires de contextes suivants montrent que les équivalents *cartographie*-NOM (2) et *carte*-NOM (4) peuvent jouer le même rôle dans certaines situations.

- 1) MAPPING of flood-susceptible areas is a significant tool in reducing flood risk,... (RNCA\_impacts\_2008)
- 2) La CARTOGRAPHIE de régions susceptibles d'être inondées est un outil important de réduction des risques associés aux inondations,... (RNCA\_impacts\_2008)
- 3) Flood-risk MAPPING for Annapolis Royal, NS... (RNCA\_impacts\_2008)

- 4) CARTE du risque d'inondation pour Annapolis Royal, en Nouvelle-Écosse...  
(RNCA\_impacts\_2008)

Les exemples suivants montrent les équivalents *stress*-NOM (1) et *agression*-NOM (2) dans des contextes expliquant les conséquences causées par des températures anormalement élevées.

- 1) ...disparition des espèces les plus sensibles au STRESS thermique au profit d'espèces plus robustes;... (ONERC\_partie2\_2009)
- 2) ...il y aura une baisse de la productivité des sols en raison des AGRSSIONS thermiques et du stress hydrique... (ONU\_desertification\_2007)

### 8.1.2 Équivalent transcatégoriel par rapport à l'ÉqT

Les équivalents transcatégoriels sont des équivalents appartenant à une autre partie du discours que l'équivalent typique et se divisent en deux catégories : les équivalents transcatégoriels par rapport à l'ÉqT et morphologiquement relié à celui-ci; les équivalents transcatégoriels non morphologiquement reliés à l'ÉqT. Dans les sections qui suivent, nous décrivons ces deux types d'équivalents.

#### 8.1.2.1 Équivalent transcatégoriel par rapport à l'ÉqT et morphologiquement relié à celui-ci

Un équivalent transcatégoriel par rapport à l'ÉqT et morphologiquement relié à celui-ci est un équivalent dont la partie du discours est différente de celle de l'ÉqT, mais qui lui ressemble sur le plan de la forme (Tableau 8.11).

Tableau 8.11. Exemples d'équivalents transcatégoriels par rapport à l'ÉqT et morphologiquement reliés à celui-ci

Terme anglais	Équivalent	Équivalent typique
<i>absorb</i> -VER	<i>absorption</i> -NOM	<i>absorber</i> -VER
<i>adaptation</i> -NOM	<i>adapter</i> -VER	<i>adaptation</i> -NOM
<i>assessment</i> -NOM	<i>évaluer</i> -VER	<i>évaluation</i> -NOM
<i>assess</i> -VER	<i>évaluation</i> -NOM	<i>évaluer</i> -VER
<i>atmosphere</i> -NOM	<i>atmosphérique</i> -ADJ	<i>atmosphère</i> -NOM
<i>capture</i> -VER	<i>piégeage</i> -NOM	<i>piéger</i> -VER

<i>change</i> -NOM	<i>changer</i> -VER	<i>changement</i> -NOM
<i>change</i> -VER	<i>changement</i> -NOM	<i>changer</i> -VER
<i>climate</i> -NOM	<i>climatique</i> -ADJ	<i>climat</i> -NOM
<i>cooling</i> -NOM	<i>refroidir</i> -VER	<i>refroidissement</i> -NOM
<i>coral</i> -NOM	<i>corallien</i> -ADJ	<i>corail</i> -NOM
<i>decrease</i> -VER	<i>diminution</i> -NOM	<i>diminuer</i> -VER
<i>ecological</i> -ADJ	<i>écologie</i> -NOM	<i>écologique</i> -ADJ
<i>emit</i> -VER	<i>émission</i> -NOM	<i>émettre</i> -VER
<i>erode</i> -VER	<i>érosion</i> -NOM	<i>éroder</i> -VER
<i>flood</i> -VER	<i>inondation</i> -NOM	<i>inonder</i> -VER
<i>food</i> -NOM	<i>alimentaire</i> -ADJ	<i>aliment</i> -NOM
<i>food</i> -NOM	<i>alimentaire</i> -ADJ	<i>aliment</i> -NOM
<i>glacial</i> -ADJ	<i>glace</i> -NOM	<i>glaciaire</i> -ADJ
<i>glacier</i> -NOM	<i>glaciaire</i> -ADJ	<i>glacier</i> -NOM
<i>hydrological</i> -ADJ	<i>hydrologie</i> -NOM	<i>hydrologique</i> -ADJ
<i>ice</i> -NOM	<i>glaciaire</i> -ADJ	<i>glace</i> -NOM
<i>increase</i> -NOM	<i>augmenter</i> -VER	<i>augmentation</i> -NOM
<i>melt</i> -VER	<i>fonte</i> -NOM	<i>fondre</i> -VER
<i>migration</i> -NOM	<i>migrer</i> -VER	<i>migration</i> -NOM
<i>mitigation</i> -NOM	<i>atténuer</i> -VER	<i>atténuation</i> -NOM
<i>mountain</i> -NOM	<i>montagneux</i> -ADJ	<i>montagne</i> -NOM
<i>mountainous</i> -ADJ	<i>montagne</i> -NOM	<i>montagneux</i> -ADJ
<i>negatively</i> -ADV	<i>nuire</i> -VER	<i>négativement</i> -ADV
<i>ocean</i> -NOM	<i>océanique</i> -ADJ	<i>océan</i> -NOM
<i>release</i> -NOM	<i>libérer</i> -VER	<i>libération</i> -NOM
<i>seasonal</i> -ADJ	<i>saison</i> -NOM	<i>saisonnier</i> -ADJ
<i>simulated</i> -ADJ	<i>simulation</i> -NOM	<i>simulé</i> -ADJ
<i>stabilization</i> -NOM	<i>stabiliser</i> -VER	<i>stabilisation</i> -NOM
<i>stabilize</i> -VER	<i>stabilisation</i> -NOM	<i>stabiliser</i> -VER
<i>stratospheric</i> -ADJ	<i>stratosphère</i> -NOM	<i>stratosphérique</i> -ADJ
<i>tropospheric</i> -ADJ	<i>troposphère</i> -NOM	<i>troposphérique</i> -ADJ
<i>variability</i> -NOM	<i>varier</i> -VER	<i>variabilité</i> -NOM
<i>vulnerability</i> -NOM	<i>vulnérable</i> -ADJ	<i>vulnérabilité</i> -NOM
<i>vulnerable</i> -ADJ	<i>vulnérabilité</i> -NOM	<i>vulnérable</i> -ADJ
<i>warm</i> -VER	<i>réchauffement</i> -NOM	<i>réchauffer</i> -VER

Les 3 contextes suivants illustrent dans l'ordre le terme anglais *adaptation*-NOM (1), son ÉqT *adaptation*-NOM (2) et l'ÉqNT trancatégoriel *adapter*-VER (1). Ces exemples montrent que la même notion peut être exprimée par un équivalent sous une forme nominale ou verbale.

- 1) Forest loss could therefore be irreversible if ADAPTATION is slow or only reactive. (RNCA\_impacts\_2008)
- 2) Les espèces végétales peuvent s'ADAPTER aux changements climatiques en migrant à des rythmes qui varient de 4 à 200km par siècle. (ONERC\_impact\_2008)
- 3) Certains considèrent que le rythme du changement climatique permettra l'ADAPTATION des espèces à travers notamment leur migration,... (Senat\_amlour\_2002)

Dans les 2 paires de contextes qui suivent, nous voyons que les termes anglais *cooling*-NOM (1) et *melt*-VER (3) sont rendus en français par des équivalents appartenant à une autre partie du discours *refroidir*-VER (2) et *fonte*-NOM (4).

- 1) ...to induce artificial COOLING of the permafrost under warming conditions. (RNCA\_impacts\_2008)
- 2) ...pour REFROIDIR artificiellement le pergélisol dans l'espoir de... (RNCA\_impacts\_2008)
- 3) As permafrost MELTS, more mercury is released to rivers and the ocean... (RNC\_impacts\_2008)
- 4) La FONTE du pergélisol libère dans les cours d'eau et l'océan de plus en plus de mercure... (RNC\_impacts\_2008)

Certains équivalents transcatégoriels sont privilégiés, voire imposés, par une collocation. Dans l'exemple (1), la collocation *changement* privilégie l'ÉqNT *climatique*-ADJ du terme anglais *climate*-ADJ. Dans l'ensemble du corpus *changement climatique* est présent 7939 fois alors que *changement du climat* n'apparaît que 47 fois. La collocation *modèle*, elle aussi, impose pratiquement toujours l'équivalent *océanique*-ADJ.

- 1) CLIMATE change => changement CLIMATIQUE
- 2) OCEAN model => modèle OCÉANIQUE



### 8.1.2.2 Équivalent transcatégoriel par rapport à l'ÉqT, mais non morphologiquement relié à l'ÉqT

Il existe également des équivalents transcategoriels qui ne sont pas morphologiquement reliés à l'équivalent typique (Tableau 8.12).

Tableau 8.12. Exemples d'équivalents transcategoriels par rapport à l'ÉqT

Terme	Équivalent	Équivalent typique
<i>air</i> -NOM	<i>atmosphérique</i> -ADJ	<i>air</i> -NOM
<i>change</i> -NOM	<i>varier</i> -VER	<i>changement</i> -NOM
<i>decrease</i> -VER	<i>réduction</i> -NOM	<i>diminuer</i> -VER
<i>fishery</i> <sub>2</sub> -NOM	<i>halieutique</i> -ADJ	<i>pêche</i> -NOM
<i>fish</i> -NOM	<i>piscicole</i> ADJ	<i>poisson</i> -NOM
<i>fish</i> -NOM	<i>halieutique</i> ADJ	<i>poisson</i> -NOM
<i>forest</i> -NOM	<i>boisé</i> -ADJ	<i>forêt</i> -NOM
<i>freshwater</i> -NOM	<i>dulcicole</i> -ADJ	<i>eau douce</i> -NOM
<i>global</i> -ADJ	<i>planète</i> -NOM	<i>mondial</i> -ADJ
<i>heat</i> -NOM	<i>thermique</i> ADJ	<i>chaleur</i> -NOM
<i>impact</i> -NOM	<i>affecter</i> -VER	<i>impact</i> -NOM
<i>increase</i> -VER	<i>accroissement</i> -NOM	<i>augmenter</i> -VER
<i>increase</i> -VER	<i>hausse</i> -NOM	<i>augmenter</i> -VER
<i>increase</i> -VER	<i>intensification</i> -NOM	<i>augmenter</i> -VER
<i>livestock</i> -NOM	<i>animal</i> -ADJ	<i>bétail</i> -NOM
<i>moisture</i> -NOM	<i>hydrique</i> ADJ	<i>humidité</i> -NOM
<i>plant</i> -NOM	<i>végétal</i> -ADJ	<i>plante</i> -NOM
<i>population</i> -NOM	<i>démographique</i> ADJ	<i>population</i> -NOM
<i>projection</i> -NOM	<i>prévoir</i> -VER	<i>projection</i> -NOM
<i>project</i> -VER	<i>prévision</i> -NOM	<i>prévoir</i> -VER
<i>radiation</i> -NOM	<i>radiatif</i> -ADJ	<i>rayonnement</i> -NOM
<i>reduce</i> -VER	<i>diminution</i> -NOM	<i>réduire</i> -VER
<i>reduce</i> -VER	<i>baisse</i> -NOM	<i>réduire</i> -VER
<i>release</i> -VER	<i>rejet</i> -NOM	<i>libérer</i> -VER
<i>resilient</i> -ADJ	<i>résister</i> -VER	<i>résilient</i> -ADJ
<i>river</i> -NOM	<i>hydrographique</i> ADJ	<i>rivière</i> -NOM <i>fleuve</i> -NOM
<i>sequester</i> -VER	<i>piégeage</i> -NOM	<i>séquestrer</i> -VER
<i>specie</i> -NOM	<i>taxinomique</i> - ADJ	<i>espèce</i> -NOM
<i>stream</i> -NOM	<i>fluvial</i> -ADJ	<i>cours d'eau</i> -NOM
<i>water</i> -NOM	<i>hydrique</i> ADJ	<i>eau</i> -NOM
<i>weather</i> -NOM	<i>météorologique</i> -ADJ	<i>temps</i> -NOM
<i>weather</i> -NOM	<i>climatique</i> -ADJ	<i>temps</i> -NOM
<i>wind</i> -NOM	<i>éolien</i> ADJ	<i>vent</i> -NOM
<i>wood</i> -NOM	<i>ligneux</i> ADJ	<i>bois</i> -NOM

On fait appel à ce type d'équivalent par choix stylistique, pour éviter des répétitions ou encore parce qu'une collocation privilégie cet équivalent. Par exemple, les termes anglais *freshwater*-NOM (1), *fish*-NOM (3), *air*-NOM (5) et *global*-ADJ (7) des exemples suivants sont respectivement rendus en français par un ÉqNT transcatégoriel, *dulcicole*-ADJ (2), *halieutique*-ADJ (4), *atmosphérique*-ADJ (6) et *planète*-NOM (8).

- 1) Climate change will affect the structure and function of Arctic FRESHWATER ecosystems. (RNCA\_impacts\_2008)
- 2) Le changement climatique aura un effet sur la structure et la fonction des écosystèmes DULCICOLES de l'Arctique. (au lieu *écosystème d'eau douce*) (RNCA\_impacts\_2008)
- 3) In addition, there are other important non-climatic factors affecting FISH resources... (ECA\_changclim\_2005)
- 4) En outre, les ressources HALIEUTIQUES sont touchées par d'autres... (au lieu de *ressources en poissons*) (ECA\_changclim\_2005)
- 5) Health problems related to AIR pollution (RNCA\_impacts\_2008)
- 6) Problèmes de santé liés à la pollution ATMOSPHERIQUE (au lieu de *pollution de l'air*) (RNCA\_impacts\_2008)
- 7) The dominant impact of GLOBAL warming will...(IPCC\_bilan\_2001)
- 8) L'effet dominant du réchauffement de la PLANÈTE consistera... (au lieu de *réchauffement mondial*) (IPCC\_bilan\_2001)

## **8.2 Résultats de l'analyse comparative des équivalents associés à un même terme anglais**

Les 241 termes anglais analysés dans le deuxième volet de notre recherche sont associés à 758 équivalents, dont 312 ÉqT et 446 ÉqNT (Tableau 8.13).

Tableau 8.13. Répartition des ÉqT et des ÉqNT

241 termes anglais	
Type d'équivalents	Nombre
ÉqT	312
ÉqNT	446
<b>Total</b>	<b>758</b>

Étant donné que nous avons effectué une analyse comparative des équivalents associés à un même terme anglais, 241 ÉqT ont servi de base de référence pour catégoriser les 517 restants. Le Tableau 8.14 montre la distribution des 517 équivalents en fonction de la typologie décrite à la section méthodologie de ce chapitre (8.1). Dans ce tableau, nous observons que les équivalents appartenant à la même partie du discours sont un peu plus nombreux que les équivalents transcatégoriels, 272 contre 245, soit une différence de 27.

Parmi les équivalents appartenant à la même partie du discours, les quasi-synonymes sont en plus grand nombre (116), suivent les équivalents synonymes (68) et les équivalents présentant un autre point de vue par rapport à l'ÉqT (61). Les autres types d'équivalence sont en nombre nettement inférieur, 9 ou 3. Il est à noter que les équivalents synonymes (68) et les équivalents dont le système conceptuel présente des divergences dans les deux langues (3) sont constitués uniquement d'ÉqT. Ces 71 équivalents ajoutés aux 241 ÉqT qui ont servi de base de référence représentent un total de 312 ÉqT (Tableau 8.13). Les équivalents de toutes les autres catégories sont des ÉqNT (Tableau 8.14).

Les équivalents transcatégoriels par rapport à l'équivalent typique sont au nombre de 245. Même s'ils ne sont pas les plus nombreux, ils représentent une part importante des équivalents. Naturellement, ces équivalents sont tous des ÉqNT. De plus, les équivalents de cette catégorie sont le plus souvent reliés morphologiquement à l'ÉqT (158) (Tableau 8.14).

Tableau 8.14. Résultats de l'analyse des équivalents

Équivalent appartenant à la même partie du discours que l'ÉqT		Nombre
1	Équivalents synonymes	68
2	Équivalent quasi-synonyme par rapport à l'ÉqT	116
3	Équivalent hyperonyme par rapport à l'ÉqT	9
4	Équivalent hyponyme par rapport à l'ÉqT	9
5	Équivalent holonyme par rapport à l'ÉqT	3
6	Équivalent méronyme par rapport à l'ÉqT	3
7	Équivalent dont le système conceptuel présente une divergence par rapport à celui de l'autre langue	3
8	Équivalent présentant un autre point de vue par rapport à l'ÉqT	61
<b>Sous-total</b>		<b>272</b>
Équivalent transcategoriel par rapport à l'ÉqT		
1	Équivalent transcategoriel par rapport à l'ÉqT et morphologiquement relié à celui-ci	158
2	Équivalent transcategoriel par rapport à l'ÉqT, mais non morphologiquement relié à l'ÉqT	87
<b>Sous-total</b>		<b>245</b>
<b>Total</b>		<b>517</b>

En tout, 169 ÉqT ont servi de base de référence pour catégoriser les 245 équivalents transcategoriels. La différence s'explique par le fait que certains termes anglais sont associés à plusieurs équivalents, par exemple, le terme anglais *capture-VER* dont l'ÉqT est *piéger-VER* est aussi associé aux ÉqNT *piégeage-NOM*, *captage-NOM*, *captation-NOM* et *capture-NOM*. Sur les 169 ÉqT, 34 sont des adjectifs, 4 des adverbes, 103 des noms et 28 des verbes (Tableau 8.15).

Tableau 8.15. Répartition des ÉqT associés à des équivalents transcategoriels

Équivalent typique	Nombre	Pourcentage
adjectif	34	20,1
adverbe	4	2,4
nom	103	61
verbe	28	16,5
<b>Total</b>	<b>169</b>	<b>100</b>

Les adjectifs sont associés à 50 équivalents transcatégoriels, les adverbes à 11, les noms à 130 et les verbes à 54 (Tableau 8.15). Le rapport entre le nombre d'ÉqT et les équivalents transcatégoriels auxquels ils sont associés est plus élevé chez les adverbes. C'est-à-dire que l'adverbe, par rapport à son nombre, est la partie du discours la plus productive en équivalents transcatégoriels<sup>116</sup>. Les verbes suivent avec un rapport de 1,9. Viennent ensuite les adjectifs (1,5) et enfin les noms (1,3), ces derniers étant relativement les moins productifs des quatre parties du discours (Tableau 8.16).

Tableau 8.16. Nombre de passage d'une partie du discours à une autre

Équivalent typique	Nombre	Nombre d'équivalents transcatégoriels	Rapport entre ÉqT et Équivalents transcatégoriels
adjectif	34	50	1,5
adverbe	4	11	2,75
nom	103	130	1,3
verbe	28	54	1,9
<b>Total</b>	<b>169</b>	<b>245</b>	

Chez les adjectifs, le passage à une autre partie du discours se fait le plus souvent vers le nom, 44 fois. Pour les adverbes, le passage se fait le plus souvent vers l'adjectif, 7 fois. Le nom, quant à lui, est plus productif en adjectifs (87) qu'en verbes (43). Enfin, pour le verbe, le passage se fait généralement vers le nom (50). Le Tableau 8.17 illustre le nombre de passage vers une autre partie du discours pour les adjectifs, les adverbes, les noms et les verbes de notre corpus. Pour terminer, nous constatons que dans notre corpus, le nom et le verbe ne présentent pas de passage vers l'adverbe.

<sup>116</sup> Il faut tout de même souligner que les adverbes sont très peu fréquents dans notre analyse.

Tableau 8.17. Typologie des équivalents transcatégoriels

Équivalent typique	Nombre	Équivalent non typique	Nombre	morphologiquement relié	non morphologiquement relié
adjectif	34	adverbe	2	2	0
		nom	44	34	10
		verbe	4	3	1
adverbe	4	adjectif	7	4	3
		nom	3	1	2
		verbe	1	0	1
nom	103	adjectif	87	57	30
		adverbe	0	0	0
		verbe	43	27	16
verbe	28	adjectif	4	3	1
		adverbe	0	0	0
		nom	50	27	23
	<b>166</b>		<b>245</b>	<b>158</b>	<b>87</b>

### 8.3 Autres observations sur l'équivalence

Au cours de la classification des équivalents, nous avons constaté trois phénomènes qui méritent une attention particulière. Le premier concerne les équivalents français dont la partie du discours est différente de celle du terme anglais, le deuxième touche les adverbes et le troisième se rapporte aux éléments et composés chimiques.

#### 8.3.1 ÉqT français dont la partie du discours est différente de celle du terme anglais

Les termes anglais ayant des ÉqT appartenant à une autre partie du discours ne sont pas nombreux dans notre corpus, mais nous en avons relevé 5 (Tableau 8.18). On remarquera que 2 sont des adjectifs et 3 des adverbes. Dans le cas des termes anglais *environmentally-ADV* et *worldwide-ADV*, nous n'avons trouvé dans notre corpus aucune occurrence d'*environnementalement* ou de *mondialement*.

Tableau 8.18. ÉqT dont la partie du discours est différente du terme anglais

Terme anglais	Équivalent français
<i>environmentally</i> -ADV	<i>environnemental</i> -ADJ
<i>seasonally</i> -ADV	<i>saison</i> -NOM <i>saisonnier</i> -ADJ
<i>tidal</i> -ADJ	<i>marée</i> -NOM
<i>upstream</i> -ADJ	<i>amont</i> -NOM
<i>worldwide</i> -ADV	<i>planète</i> -NOM <i>mondial</i> -ADJ

### 8.3.2 Équivalents des termes anglais adverbes

Dans le corpus parallèle, les 7 termes anglais adverbes sont plus souvent traduits par un adjectif, un nom ou une locution que par un adverbe. D'ailleurs, le terme anglais *seasonally* n'a pas d'équivalent adverbe. Les 7 termes anglais adverbes analysés dans cette étude sont *environmentally*, *globally*, *locally*, *negatively*, *regionally*, *seasonally*, *worldwide*. Ces termes sont reportés dans le Tableau 8.19 avec leurs équivalents respectifs.

Tableau 8.19. Équivalents de 7 termes anglais adverbes

Terme anglais	Équivalent
<i>environmentally</i> -ADV	<i>environnement</i> -NOM (39) <i>environnemental</i> -ADJ (6) <i>écologiquement</i> -ADV (10)
<i>globally</i> -ADV	<i>mondial</i> -ADJ (26) <i>au niveau mondial</i> -LOC (13) <i>à l'échelle mondiale</i> -LOC (17) <i>au plan mondial</i> -LOC (1) <i>point de vue mondial</i> -LOC (1) <i>dans le monde</i> -LOC (14) <i>du monde</i> -LOC (1) <i>globe</i> -NOM (20) <i>planétaire</i> -ADJ (1) <i>à l'échelle de la planète</i> -LOC (8) <i>d'ordre planétaire</i> (1) <i>planète</i> -NOM (10) <i>global</i> -ADJ (10) <i>mondialement</i> -ADV (9) <i>globalement</i> -ADV (8)
<i>locally</i> -ADV	<i>au niveau local</i> -LOC (4) <i>à l'échelle locale</i> -LOC (11)

	<i>local</i> -ADJ (5) <i>localement</i> -ADV (14)
<i>negatively</i> -ADV	<i>négatif</i> -ADJ (7) <i>de manière négative</i> -LOC (2) <i>de façon négative</i> -LOC (1) <i>néfaste</i> -ADJ (12) <i>négativement</i> -ADV (6) <i>nuire</i> -VER (6) <i>défavorable</i> -NOM (5)
<i>regionally</i> -ADV (64)	<i>régional</i> -ADJ (19) <i>région</i> (4) <i>d'une région à l'autre</i> (8) <i>selon les régions</i> (12) <i>entre les régions</i> (1) <i>en fonction des régions</i> (1) <i>à l'échelle régionale</i> (8) <i>sur le plan régional</i> (1) <i>au niveau régional</i> (2) <i>régionalement</i> (5)
<i>seasonally</i> -ADV	<i>saisonnier</i> -ADJ (16) <i>saison</i> -NOM (5)
<i>worldwide</i> -ADV	<i>monde</i> -NOM (19) <i>monde entier</i> -LOC (19) <i>à travers le monde</i> -LOC (7) <i>autour du monde</i> -LOC (1) <i>mondial</i> -ADJ (23) <i>planète</i> -NOM (4)

### 8.3.3 Équivalents des éléments et composés chimiques

Dans le corpus parallèle, nous avons observé que certains éléments et composés chimiques écrits au long en anglais peuvent avoir pour équivalent leur symbole chimique (formule chimique), alors que d'autres ne l'ont jamais. Dans l'exemple suivant, on voit que le terme anglais *carbon dioxide* a pour équivalent français sa formule chimique. Nous n'avons pas vraiment d'explications pour ce phénomène. On pourrait penser qu'il n'est pas nécessaire de traduire au long ces éléments ou composés chimiques lorsqu'ils sont souvent cités. Cependant, cette hypothèse ne se vérifie pas toujours lorsqu'on observe leur fréquence. Par exemple, la fréquence de N<sub>2</sub>O (130) n'est pas tellement plus basse que celle de CH<sub>4</sub> (177), pourtant seule la formule chimique CH<sub>4</sub> sert d'équivalent au terme anglais *methane*-NOM.



CP : In addition, the oceans constitute an important sink of atmospheric CARBON DIOXIDE.

CP : De plus, les océans constituent un puits de CO<sub>2</sub> important (Unesco\_patrimoine\_2009).

Le Tableau 8.20 présente quelques éléments et composés chimiques. Les trois premières colonnes contiennent dans l'ordre : le terme anglais, l'équivalent français, la formule chimique. La dernière colonne indique par oui ou par non si la formule chimique sert d'équivalent au terme écrit au long dans le texte source.

Tableau 8.20. Éléments et composés chimiques et équivalence

Terme anglais	Équivalent français	Formule chimique	Oui/non
<i>carbon dioxide</i>	<i>dioxyde de carbone</i>	CO <sub>2</sub>	oui
<i>carbon monoxide</i>	<i>monoxyde de carbone</i>	CO	non
<i>hydrocarbon</i>	<i>hydrocarbures</i>	HC	non
<i>methane</i>	<i>méthane</i>	CH <sub>4</sub>	oui
<i>nitrous oxide</i>	<i>oxyde nitreux</i>	N <sub>2</sub> O	non
<i>nitrogen oxide</i>	<i>oxyde d'azote</i>	NO <sub>x</sub>	non
<i>ozone</i>	<i>ozone</i>	O <sub>3</sub>	non
<i>sulphur dioxide</i>	<i>dioxyde de soufre</i>	SO <sub>2</sub>	non

## 8.4 Conclusion du chapitre

Dans ce chapitre, nous avons présenté une analyse comparative des équivalents associés à un même terme anglais afin de voir en quoi les différents équivalents d'un terme se distinguent.

Dans la première partie du chapitre, nous avons décrit la méthodologie suivie pour analyser les équivalents. Premièrement, nous avons expliqué que les équivalents d'un même terme sont comparés à l'ÉqT du terme anglais<sup>117</sup>. Deuxièmement, nous avons distingué les équivalents appartenant à la même partie du discours que l'ÉqT des équivalents d'une autre partie du discours. La première catégorie est divisée en 8 sous-

<sup>117</sup> Les termes n'ayant qu'un même équivalent dans le corpus parallèle et le corpus comparable n'ont pas été analysés dans ce chapitre puisqu'ils ne peuvent être comparés à eux-mêmes.

catégories et la deuxième en 2 sous-catégories (Tableau 8.2). Les sous-catégories ont été décrites et des exemples venant du corpus ont été présentés.

La deuxième partie du chapitre présente les résultats de l'analyse comparative. Pour un total de 758 équivalents, nous avons recensé 312 ÉqT et 446 ÉqNT (Tableau 8.13). Les équivalents appartenant à la même partie du discours sont au nombre de 272 et les équivalents transcatégoriels de 245. Même si le nombre de ce dernier groupe est inférieur au premier, il est relativement élevé (Tableau 8.14). Parmi les équivalents de la même partie du discours, les quasi-synonymes sont de loin les plus fréquents, suivent les synonymes et les équivalents présentant un autre point de vue par rapport à l'ÉqT. Dans les 5 autres catégories, la fréquence est très peu élevée, 27 en tout (Tableau 8.14). Parmi les équivalents transcatégoriels, qu'ils soient ou non morphologiquement reliés à l'ÉqT, nous constatons qu'ils représentent un nombre important des équivalents. Nous avons aussi observé que les adverbes sont les plus susceptibles d'être associés à des équivalents transcatégoriels; suivis par les verbes, les adjectifs et enfin les noms.

La classification présentée dans ce chapitre s'apparente aux typologies de l'équivalence du Chapitre 2 de la façon suivante. Premièrement, les ÉqT font partie dans la très grande majorité des équivalences exactes. Les ÉqT qui font exception à la règle sont ceux qui sont associés aux termes anglais dont le système conceptuel présente des divergences avec celui du français (Svensén 2009 : 254). Deuxièmement, les ÉqNT dont la partie du discours est identique à celle de l'ÉqT font partie des équivalences partielles. Enfin, les ÉqNT dont la partie du discours est différente à celle de l'ÉqT entreraient, entre autres, dans les procédés obliques de Vinay et Darbelnet (1958) ou les « shift » de Catford (1965).

Dans la troisième partie du chapitre, nous nous sommes arrêtée sur trois aspects de l'équivalence observés dans notre corpus; à savoir les 5 cas d'ÉqT dont la partie du discours est différente de celle du terme anglais; les nombreux équivalents d'un même terme anglais adverbe et les équivalents des éléments et composés chimiques.

Les résultats de l'analyse comparative des équivalents nous montrent que de nombreux termes anglais peuvent être rendus par plusieurs équivalents. Nous avons aussi

constaté que les synonymes ne sont pas la catégorie la plus fréquente. D'autre part, les équivalents appartenant à une autre partie du discours constituent une part importante des équivalents.

Il est à noter que les dictionnaires spécialisés valident en général les synonymes, parfois les quasi-synonymes, mais ne valident pas les autres types d'équivalence décrits dans notre analyse comparative des équivalents. Cela est tout à fait normal, énumérer tous les équivalents non typiques d'un terme serait une entreprise trop difficile à réaliser, car il faudrait connaître d'avance tous les contextes possibles et imaginables dans lesquels ces notions peuvent être utilisées. Autrement dit, la situation de communication impose le choix des équivalents, qu'ils soient « découverts » ou « créés » (Adamska-Salaciak (2010 : 399–400) (*cf.* 2.2.5). Ainsi, la typologie de l'équivalence en terminologie présentée dans ce chapitre montre la richesse des moyens linguistiques mis à la disposition des traducteurs ou des auteurs pour transposer une notion.

Les auteurs traitant de l'extraction bilingue sur corpus comparable décrivent rarement ce qu'ils entendent exactement par "équivalence". En tout cas, ils n'envisagent pratiquement jamais d'extraire des équivalents appartenant à une autre partie du discours. Les résultats de cette thèse montrent que les listes de termes produites par les extracteurs sur corpus comparables doivent être analysées à la lumière de ce nous avons observé en corpus. Ainsi, pour mieux évaluer les outils d'extraction, il serait utile de tenir compte de tous les cas de figure présentés dans ce chapitre, en particulier l'équivalence transcatégorielle.

Par ailleurs, en extraction bilingue de termes sur corpus comparable, pour évaluer les résultats offerts par un outil, les candidats équivalents sont généralement ordonnés selon les valeurs statistiques obtenues, en espérant que l'équivalent du terme source soit situé au premier rang. Ainsi, on estime la performance d'un outil sur le nombre d'équivalents qui se trouve situé au premier rang, plus le nombre des équivalents occupant le premier rang est élevé, meilleure sera la performance. Selon nous, cette méthode d'évaluation n'est pas appropriée, car elle ne tient pas compte des phénomènes de l'équivalence énumérés dans ce chapitre. Lorsqu'il existe plusieurs équivalents pour un même terme, ce type de

classification n'est pas valable, par exemple, dans notre corpus l'équivalent le plus fréquent du terme anglais *climate*-NOM est l'adjectif *climatique*-ADJ. Conséquemment, l'outil ne se tromperait pas s'il présentait cet équivalent au premier rang; or il risquerait d'être pénalisé au moment de la validation puisque l'équivalent typique du terme anglais est *climat*-NOM.

## 9 CONCLUSION GÉNÉRALE

Cette thèse se divisait en deux volets principaux. D'une part, nous voulions réaliser une analyse comparative de l'équivalence terminologique en corpus parallèle et en corpus comparable, afin d'en vérifier les différences et les similitudes. D'autre part, après avoir observé en corpus que les termes d'une langue source sont souvent associés à plusieurs équivalents en langue cible, nous désirions procéder à l'analyse des équivalents d'un même terme anglais.

Pour atteindre nos objectifs, après avoir évoqué quelques approches théoriques de la terminologie, nous avons choisi d'asseoir notre étude sur la base de la terminologie textuelle et de l'optique lexico-sémantique. Ces deux approches étaient particulièrement appropriées pour le type d'étude que nous projetions. On se souviendra, entre autres, que les deux approches prennent pour point de départ le texte pour sélectionner les termes et recueillir des renseignements sur ces termes. Par ailleurs, notre recherche se distingue des autres études par le fait que nous avons étudié uniquement les termes simples. Il est également à noter que nous avons travaillé de l'anglais vers le français. Ce choix s'est tout naturellement imposé lors de la construction du sous-corpus parallèle. En effet, parmi les paires de textes comportant un texte source et une traduction, les traductions de l'anglais vers le français sont plus nombreuses que l'inverse.

En vue de mieux cerner du point de vue notionnel l'équivalence, nous avons exploré dans le détail les fondements théoriques de l'équivalence en traductologie, en lexicographie et en terminologie. Au terme de cette exploration, nous avons pu constater qu'il existe plusieurs définitions de l'équivalence au sein de chacune de ces disciplines et qu'aucune ne fait consensus. De la traductologie, nous avons lu avec intérêt les travaux de Vinay et Darbelnet (1958) et de Catford (1965) puisqu'ils abordent certains phénomènes de l'équivalence que nous avons observés en corpus. De la lexicographie, nous avons retenu plus particulièrement les enseignements sur le non-isomorphisme des systèmes linguistiques, sur l'équivalence explicative et sur l'équivalence pragmatique. Enfin, de la terminologie, nous avons suivi avec intérêt les travaux des auteurs ayant déjà fait des observations de l'équivalence en corpus (*cf.* 2.3.2). Ces différentes approches nous ont

permis de tenir compte dans nos analyses des contextes dans lesquels les équivalents s'inséraient, nous affranchissant du même coup de la terminologie classique afin d'observer certains phénomènes courants en corpus habituellement passés sous silence.

Le chapitre traitant des textes électroniques et des corpus avait pour premier objectif de définir et de décrire les types de corpus avec lesquels nous avons travaillé dans cette étude, soit les corpus parallèles et les corpus comparables. Le deuxième objectif était d'analyser les critiques touchant les corpus parallèles pour en connaître l'importance et pour relever les raisons évoquées par les auteurs pour critiquer les traductions. Tous d'abord, nous avons constaté que lorsqu'il y avait des commentaires sur les traductions, ils étaient plus souvent négatifs que positifs. Ensuite, il leur était reproché de contenir des emprunts fautifs, des faux sens et des cognats dont la fréquence est plus élevée que dans les originaux.

À cause du nombre important des termes analysés dans notre étude, nous nous sommes largement appuyée sur les outils informatiques. Nous avons donc consacré tout un chapitre à la description des différentes techniques d'extraction lexicale et terminologique (*cf.* 3). Cette étude nous a permis de mieux comprendre le fonctionnement de ces outils et d'en connaître les possibilités et les limites. Nous avons abordé en premier les extracteurs de termes sur corpus monolingue. Nous sommes passée ensuite à l'extraction bilingue sur corpus parallèle et nous avons terminé avec l'extraction bilingue sur corpus comparable. Nous avons utilisé les outils appartenant aux deux premiers groupes, mais pas celui du dernier. Cependant, comme l'extraction en corpus comparable est encore du domaine de la recherche, nous avons particulièrement développé sa description afin de dégager en quoi une étude comme la nôtre pourrait servir cette technique.

Pour éviter de répéter les informations et faciliter la lecture de cette thèse, nous avons réuni dans le Chapitre 5 la méthodologie commune aux deux volets de la recherche. Dans cette partie nous avons construit un corpus original. En effet, celui-ci se divise en trois parties, le sous-corpus anglais faisant office de relais entre les deux sous-corpus français. Le repérage des équivalents dans le corpus parallèle a été relativement facile à réaliser. Par contre, dans le corpus comparable, il a fallu élaborer une méthodologie assez

complexe pour permettre le ratissage systématique du corpus afin de repérer les équivalents. Cependant, grâce aux ressources terminologiques, aux trois types de documents et aux sept stratégies, nous sommes assurée d'avoir la plus grande couverture possible.

Les résultats du premier volet ont confirmé notre hypothèse de départ. Nous avons même été agréablement surprise de voir que la terminologie employée dans les deux types de corpus était très similaire. En fait, nous avons constaté que les auteurs des textes originaux de notre corpus sont ceux qui introduisaient le plus d'emprunts fautifs. Nous pensons que cette situation peut s'expliquer par le fait que les auteurs de ces textes lisent de nombreux ouvrages et participent à des colloques en anglais. Il est par conséquent normal qu'ils soient influencés par la terminologie anglaise. De leur côté, les traducteurs ont généralement reçu une formation pour éviter les pièges étudiés dans cette thèse. Il est également intéressant de souligner que sans une analyse comparative entre un corpus parallèle et un corpus comparable, nous n'aurions probablement pas remarqué l'emploi quasi systématique des cognats *capture* et *déforestation* dans le corpus comparable. Bien entendu, d'autres questions terminologiques auraient pu être abordées, par exemple des auteurs comme Vinay et Darbelnet (1958) font remarquer que l'anglais utilise des « mots communs » (*snow*) là où le français emploie des « mots savants » (*nival*). Il serait donc intéressant dans une future étude de voir si les traductions contiennent autant de mots savants que les originaux. De même, des faits de langues autres que terminologiques pourraient également être examinés. Par exemple, il serait intéressant d'effectuer une comparaison de l'emploi des prépositions, de la syntaxe ou de la typographie. Cependant, nous pouvons déjà conclure que les traductions sont tout aussi valables que les originaux pour étudier la terminologie d'un domaine. En fait, selon nous, ce qui permet d'assembler un corpus de bonne qualité ce sont les critères de sélection des textes.

La typologie de l'équivalence présentée dans le deuxième volet de la recherche illustre la richesse des moyens empruntés pour exprimer des situations identiques. Il est intéressant de constater que les équivalents synonymes n'en forment pas le groupe le plus important. Ce qui est particulièrement marquant dans nos résultats, ce sont les nombreux

cas de passages transcatégoriels. À notre avis, ce n'est que dans les corpus qu'il est possible d'observer l'ampleur de ce phénomène. En ce sens, notre analyse des équivalents représente une contribution importante à la réflexion sur l'équivalence en terminologie.

Les résultats de notre typologie montrent l'importance de la prise en compte des équivalents transcatégoriels en traitement automatique du langage pour évaluer la performance des outils d'extraction bilingue sur corpus comparable et sur corpus parallèle. Les évaluations tiennent rarement compte de ce genre d'équivalence. Il est probable que si on le faisait, les résultats annoncés seraient encore meilleurs. Notre étude montre clairement que pour mieux découvrir ce que contient un corpus, il faut bien connaître les mécanismes de l'équivalence. Par notre étude, nous avons démontré qu'il est préférable d'éviter d'établir des listes d'équivalents au préalable et d'effectuer une validation manuelle en corpus des résultats offerts par les outils. Dans le même ordre d'idée, la typologie de l'équivalence que nous avons proposée, montre que les personnes œuvrant dans le domaine langagier (ex. terminologues, traducteurs) devraient avoir accès à des outils qui tiennent compte de tous les types d'équivalence, même des équivalents non typiques.

En ce qui concerne l'élaboration d'outils d'extraction bilingues sur corpus comparable, nous pensons qu'il serait utile de travailler avec un corpus construit sur le même modèle que le nôtre, à savoir en trois parties. Ainsi, les résultats du corpus comparable pourraient être comparés à ceux du corpus parallèle. Une variante peut-être encore plus intéressante serait de travailler sur un corpus parallèle comme si c'était un corpus comparable. Pour expérimenter un outil ayant à travailler sur les corpus comparables les textes ne seraient pas alignés. Toutefois, on aurait la possibilité de tester les résultats plus facilement ce qui permettrait de trouver des solutions pour tenir compte des phénomènes observés dans cette étude. Enfin, force est de constater que l'énoncé de Rapp (1999 : 520) (*cf.* 4.3.)<sup>118</sup> doit être élargi si l'on veut trouver tous les types d'équivalents.

---

<sup>118</sup> If – for example – in a text of one language two words A and B co-occur more often than expected by chance, then in a text of another language those words that are translations of A and B should also co-occur more frequently than expected. (Rapp 1999 : 520)



Toute étude répond à des questions, mais surtout en appelle beaucoup d'autres. Notre recherche ne fait pas exception à cette règle. Bien que nous ayons réalisé un grand nombre d'analyses, plusieurs autres nous viennent à l'esprit. Une première serait d'effectuer une étude de l'équivalence du français vers l'anglais afin de vérifier si notre typologie de l'équivalence présente les mêmes particularités que celles que nous avons élaborée. Une autre serait d'examiner les équivalents des termes complexes. De plus, étant donné que nous avons observé des traces de variation régionale et de variation diachronique, nous pourrions construire des corpus permettant d'effectuer ce type d'analyse. Enfin, les expérimentations présentées dans cette étude demanderaient à être étendues vers des domaines différents de celui avec lequel nous avons travaillé et vers d'autres couples de langues afin de vérifier si nous observerions les mêmes phénomènes que dans notre étude.

## BIBLIOGRAPHIE

- Adamska-Salaciak, A. 2010. « Examining equivalence ». *International Journal of Lexicography* 23(4), 387–409.
- Ahrenberg, L., M. Andersson et M. Merkel. 1998. « A simple hybrid aligner for generating lexical correspondences in parallel texts ». Dans *Proceedings of 36<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*. 29–35. Montréal, Canada.
- Ahronian, C. 2005. *Les noms composés anglais, français et espagnol du domaine d'Internet*. Thèse de doctorat en Lexicologie et Terminologie Multilingue – Traduction, Université Lumière Lyon 2.
- Al-Kasimi, A. 1977. *Linguistics and Bilingual Dictionaries*. Leiden : Brill.
- Anderman, G. 2007. « Linguistics and translation ». Dans Kuhlweiczak, P. et K. Littau (dir.). *A Companion to Translation Studies (Topics in Translation)*. 45–62. Clevedon : Multilingual Matters.
- Anthony, L. 2007. *AntConc3.1.2 A Freeware Concordance Program*. <http://www.antlab.sci.waseda.ac.jp/> (page consultée le 21 juin 2010).
- Baker, M. 1993. « Corpus linguistics and translation studies: implications and applications ». Dans Baker, M., G. Francis et E. Tognini-Bonelli (dir.). *Text and Technology: in Honour of John Sinclair*. 233–52. Amsterdam : Benjamins.
- Baker, M. 1995. « Corpora in translation studies: An overview and some suggestions for future research ». *Target* 7(2), 223–243.
- Baker, M. 1996. « Corpus-based translation studies: The challenges that lie ahead ». Dans Somers, H. (dir.). *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager*. 175–186. Amsterdam & Philadelphia : John Benjamins.
- Baker, M. 1998. « Réexplorer la langue de la traduction : une approche par corpus ». *Meta* 43(4), 480–485.

- Baker, M. 2004. « The status of equivalence in translation studies: an appraisal ». Dans Bravo, J. M. (dir.). *A New Spectrum of Translation Studies*. 63–71. Valladolid: Universidad de Valladolid.
- Bassnett, S. et H. Trivedi (dir.). 1999. *Postcolonial Translation: Theory and Practice*. Londres/New York : Pinter.
- Bastin, G., Á. Echeverri et Á. Campo. 2004. « La traduction-appropriation, une arme idéologique. Le cas de l'émancipation de l'Amérique hispanique », version révisée et mise à jour de l'article « La traducción en América Latina: propia y apropiada ». Dans Pagni, A. (dir.). *Estudios* 24, 69–94. Caracas : Universidad Simón Bolívar.
- Bédard, C. 1986. *La traduction technique : principes et pratique*. Brossard (Québec) : Linguattech.
- Bekavac, B., P. Osenova, K. Simov et M. Tadić. 2004. « Making monolingual corpora comparable: a case study of Bulgarian and Croatian ». Dans *Proceedings of the 4<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'04)*. 1187–1190. Lisbonne, Portugal.
- Bernot, D. 2000. « La lexicographie bilingue : un chemin semé d'embûches d'une culture à une autre ». Dans Szende, T. (dir.). *Dictionnaires bilingues : Méthodes et contenus*. 53–60. Paris : Honoré Champion.
- Biber, D. 1993. « Using register-diversified corpora for general language studies ». *Computational Linguistics* 19(2), 243–258.
- Bodson, C. 2005. « Termes et relations sémantiques en corpus spécialisés : rapport entre patrons de relations sémantiques (PRS) et types sémantiques (TS) ». Thèse de doctorat, Département de linguistique et de traduction, Université de Montréal.
- Bourigault, D. 1993. « Analyse syntaxique locale pour le repérage de termes complexes dans un texte ». *Traitement automatique des langues* 34(2), 105–117.
- Bourigault, D. et M. Slodzian. 1999. « Pour une terminologie textuelle ». *Terminologies Nouvelles Spécial TIA* 19, 29–32.
- Bourigault, D., C. Chodkiewicz et J. Humbley. 1999. « Construction d'un lexique bilingue des droits de l'homme à partir de l'analyse automatique d'un corpus aligné ». *Terminologies Nouvelles Spécial TIA* 19, 69–77.

- Boutin-Quesnel, R., N. Bélanger, N. Kerpan et L.-J. Rousseau. 1985. *Vocabulaire systématique de la terminologie*. Québec : Les publications du Québec.
- Bowker, L. et J. Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York : Routledge.
- Brown, P., J. Lai, R. Mercer. 1991. « Aligning Sentences in Parallel Corpora ». Dans *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL'91)*. 169–176. Morristown, New Jersey.
- Brown, P. F., S. A. D. Pietra, V. J. D. Pietra et L. R. Mercer. 1993. « The mathematics of statistical machine translation: parameter estimation ». *Computational Linguistics* 19(2), 263–311.
- Bureau de la traduction. 2007. *Le Pavel, didacticiel de terminologie*. [http://www.collectionscanada.gc.ca/webarchives/20071116012251/http://www.termiumplus.gc.ca/didacticiel\\_tutorial/francais/lecon1/page1\\_2\\_5\\_f.html](http://www.collectionscanada.gc.ca/webarchives/20071116012251/http://www.termiumplus.gc.ca/didacticiel_tutorial/francais/lecon1/page1_2_5_f.html) (page consultée le 30 septembre 2012).
- Cabré, M. T. 1998. *La terminologie. Théorie, méthode et applications*, traduit du catalan, adapté et mis à jour par Cormier, M. et J. Humbley. Ottawa : Presses de l'Université d'Ottawa (Regards sur la traduction) et Paris : Armand Colin (U - Linguistique).
- Cabré, M. T. 2000. « Terminologie et linguistique : la théorie des portes ». *Terminologie nouvelle* 21, 10–15.
- Cabré, M. T. 2005. « Lexicographie versus terminographie: comment les technologies déplacent leur affrontement ». Dans Béjoint, H et F. Maniez (dir.). *De la mesure dans les termes : hommage à Philippe Thoiron*. 189–210. Lyon : Presses universitaires de Lyon.
- Carl, M., E. Rascu, J. Haller et P. Langlais. 2004. « Abducing term variant translations in aligned texts ». *Terminology* 10(1), 103–133.
- Carreño, I. 2004. *Analyse de la variation terminologique en corpus parallèle anglais-espagnol et de son incidence sur l'extraction de termes bilingues*. Thèse de maîtrise. Université de Montréal : Département de linguistique et de traduction.
- Catford, J. 1965. *A Linguistic theory of translation : an essay in applied linguistics*. London : Oxford University Press.

- Celotti, N. 2002. « La culture dans les dictionnaires bilingues : où, comment, laquelle? » *Revue de didactologie des langues-cultures* 4(128), 455–466.
- Chamberlain, L. 1988. « Gender and the metaphors of translation ». *Signs : Journal of Women and Culture in Society* 13(3), 454–472.
- Chesterman, A. 2004. « Beyond the particular ». Dans Mauranen, A. et P. Kujamäki (dir.). *Translation Universals. Do they exist?* 33–49. Amsterdam/Philadelphia : John Benjamins.
- Chesterman, A. 2008. « Translation data problems ». Dans Lindstedt, J. *et al.* (dir.). *S ljubovŔju k slovu. Festschrift for Arto Mustajoki*. 17–26. Helsinki: Department of Slavonic and Baltic Languages and Literatures.
- Chiao, Y.-C. 2004. *Extraction lexicale bilingue à partir de textes médicaux comparables : application à la recherche d'information translangue*. Thèse de doctorat en informatique. Université Pierre et Marie Curie, Paris 6.
- Chiao, Y.-C. et P. Zweigenbaum. 2002. « Looking for candidate translational equivalents in specialized, comparable corpora ». Dans *Proceedings of the American Medical Informatics Association 2002 Annual Symposium*. 150–154. San Antonio, Texas.
- Chiao, Y.-C., J.-D. Sta et P. Zweigenbaum. 2004. « A novel approach to improve word translation extraction from non-parallel, comparable corpora ». Dans *Actes International Joint Conference on Natural Language Processing*. Hainan, Chine.
- Chiao, Y.C., O. Kraif, D. Laurent, T. M. H. Nguyen, N. Semmar, F. Stuck, J. Véronis et W. Zaghouani. 2006. « Evaluation of multilingual text alignment systems: the ARCADE II project ». Dans *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC'06)*. Gène, Italie.
- Chomsky, N. 1957. *Syntactic Structures*. La Hague : Mouton.
- Choueka, Y., S. Klein et E. Neuwitz. 1983. « Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus ». *Journal of the Association for Literary and Linguistic Computing* 4(1), 34–38.
- Church, K. 1993. « Char align: A program for Aligning Parallel Texts at the Character Level ». Dans *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics (ACL-93)*. 1–8. Columbus, Ohio.

- Church, K. et P. Hanks. 1990. « Word association norms, mutual information, and lexicography ». *Computational Linguistics* 16(1), 22–29.
- Condamines, A. 2003. « *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques* ». Habilitation à diriger les recherches. <http://w3.erss.univ-tlse2.fr:8080/index.jsp?perso=acondami&subURL=publications.html> (page consultée le 12 juillet 2010).
- Condamines, A. 2005. « Linguistique de corpus et terminologie ». Dans Depecker, L. (dir.). *Langages, La terminologie : nature et enjeux* 157, 36–47. Paris : Larousse.
- Confiant, R. 2003. « La traduction en milieu diglossique ». Dans *Atelier de recherche du sur l'enseignement du créole et du français dans l'espace américano-caraiïbe (AREC-F)*. 1–17.
- Corpas Pastor, G., R. Mitkov, N. Afzal et V. Pekar. 2008. « Translation universals: do they exist? A corpus-based NLP study of convergence and simplification ». Dans *Proceedings of the 6<sup>th</sup> International Conference on Language Resources and Evaluation (LREC'08) Workshop on Building and Using Comparable Corpora*. Marrakech, Maroc. [http://www.amtaweb.org/papers/2.05\\_Corpas.pdf](http://www.amtaweb.org/papers/2.05_Corpas.pdf) (page consultée le 17 juillet 2011).
- Cronin, M. 2003. *Translation and Globalization*. Londre/New York: Routledge.
- Cruse, D.A. 1986. *Lexical Semantics*. London / New York / New Rochelle /Melbourne / Sydney : Cambridge University Press.
- Dagan, I. et K. Church. 1994. « Termight: Identifying and translating technical terminology ». Dans *Proceedings of the 4th Conference on Applied Natural Language Processing (ANLP'94)*. 34–40. Stuttgart, Allemagne.
- Daille, B., É. Gaussier et J.-M. Langé. 1994. « Towards automatic extraction of monolingual and bilingual terminology ». Dans *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*. 515–521. Kyoto, Japon.
- Déjean, H. et É. Gaussier. 2002. « Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables ». *Lexicometrica No spécial 2002*, 1–21.

- Déjean, H., E. Gaussier, J.-M. Renders et F. Sadat. 2005. « Automatic processing of multilingual medical terminology: applications to thesaurus enrichment and cross-language information retrieval ». *Artificial Intelligence in Medicine* 33, 111–224.
- Deléger, L. 2009. *Exploitation de corpus parallèles et comparables pour la détection de correspondances lexicales : application au domaine médical*. Thèse de doctorat. Université Pierre et Marie Curie – Paris 6.
- Delisle, J. 2003. *La traduction raisonnée*. Ottawa : Les Presses de l'Université d'Ottawa.
- Diki-Kidiri, M. 2008. *Le vocabulaire scientifique dans les langues africaines : Pour une approche culturelle de la terminologie*. Paris : Karthala.
- Drouin, P. 2002. *Acquisition automatique de termes : l'utilisation des pivots lexicaux spécialisés*. Thèse de doctorat. Département de linguistique et de traduction, Université de Montréal.
- Drouin, P. 2003. « Term extraction using non-technical corpora as a point of leverage ». *Terminology* 9(1), 99–117.
- Drouin, P. et F. Doll. 2008. « Quantifying termhood through corpus comparison ». Dans *Terminology and Knowledge Engineering (TKE-2008)*. 101–206. Copenhagen Business School, Copenhagen.
- Dubuc, R. 2002. *Manuel pratique de terminologie*, 4e édition. Montréal : Linguatech.
- Dunning, T. 1993. « Accurate methods for the statistics of surprise and coincidence ». *Computational Linguistics* 19(1), 61–74.
- Estopà, R. 2001. « Les unités de signification spécialisées élargissant l'objet du travail en terminologie ». *Terminology* 7(2), 217–237.
- Felber, E. 1987. *Manuel de terminologie*. Paris : Organisation des Nations Unies pour l'éducation, la science et la culture.
- Flotow, L. von. 1991. « Feminist translation: Contexts, practices dans theories ». *TTR : traduction, terminologie, rédaction* 4(2), 69–84.
- Foo, J. et M. Merkel. 2010. « Computer aided term bank creation and standardization. Building standardized term banks through automated term extraction and advanced editing tools ». Dans Thelen, M. et F. Steurs (dir.). *Terminology in Everyday Life*. 163–180. Amsterdam/Philadelphie : John Benjamins.

- Fung, P. 1995. « Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus ». Dans Yarovsky, D. et K. Church (dir.). *Proceedings of the 3rd Workshop on Very Large Corpora (VLC'95)*. 173–183. Boston, Mass.
- Fung, P. 1998. « A statistical view on bilingual lexicon extraction : From parallel corpora to non-parallel corpora ». Dans *Proceedings of the Conference of the association of Translation in the americas (AMTA'98)*. 1–17. Langhorne, Pennsylvanie.
- Fung, P. et K. Church. 1994. « K-vec: A new approach for aligning parallel texts ». Dans *Proceedings of the 15th International Conference on Computational Linguistics*. 1096\_1102. Kyoto, Japan.
- Fung, P. et K. McKeown. 1997. « Finding Terminology Translations from Non-Parallel Corpora ». Dans *Proceedings of the 5th Annual Workshop on Very Large Copora 1*, 192–202. Hong Kong.
- Fung, P. et L. Y. Yee. 1998. « An IR approach for translating new words from nonparallel, comparable text ». Dans *Proceedings of the 17th international conference on Computational linguistics*. 414–420. Montréal, Canada.
- Gale, W. et K. W. Church. 1991. « A program for aligning sentences in bilingual corpora ». Dans *Proceeding of the 29th Annual Meeting of the Association for Computational Linguistics*. 177–184. Berkeley, California.
- Gamallo Otero, P. 2007. « Learning bilingual lexicons from comparable English and Spanish corpora ». Dans *Machine Translation Summit'07*. 191–198. Copenhagen, Denmark.
- Gamallo Otero, P. 2008. « Evaluating two different methods for the task of extracting bilingual lexicons from comparable corpora ». Dans *Proceedings of the 6th Edition of the Language Resources and Evaluation Conference (LREC'08) – Workshop on Comparable Corpora*. 19–26. Marrakech, Maroc.
- Garera, N., C. Callison-Burch et D. Yarowsky. 2009. « Improving translation lexicon induction from monolingual corpora via dependency contexts and part-of-speech equivalences ». Dans *Proceedings of the Thirteenth Conference on Computational Natural Language Learning, (CoNLL'09)*. 129–137. Boulder, Colorado.



- Gaudin, F. 1993. *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Rouen : Presses de l'Université de Rouen.
- Gaussier, É. 1998. « Flow network models for word alignment and terminology extraction from bilingual corpora ». Dans Boitet, C. (dir.). *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics (ACL) and the 17th International Conference on Computational Linguistics (COLING)*. 444–450. Montréal, Canada.
- Gaussier, É. 2001. « General considerations on bilingual terminology extraction ». Dans Bourigault, D., C. Jacquemin et M.-C. L'Homme (dir.). *Recent Advances in Computational Terminology*. 167–182. Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Gaussier, É. 2005. *Contributions à l'accès à l'informatique documentaire*. Mémoire de synthèse présenté en vue de l'obtention d'une habilitation à diriger des recherches – spécialité informatique.
- Gaussier, É., J.-M. Renders, I. Matveeva, C. Goutte et H. Déjean. 2004. « A geometric view on bilingual lexicon extraction from comparable corpora ». Dans *Proceedings of the 42<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (ACL'04)*. 526–533. Barcelone, Espagne.
- GDT. *Grand dictionnaire terminologique de l'Office de la langue français*. <http://www.olf.gouv.qc.ca/ressources/gdt.html>. (page consultée le 16 juillet 2011).
- Gellerstam, M. 1996. « Translations as a source for cross-linguistic studies ». Dans Aijmer, K., B. Altenberg et M. Johansson (dir.). *Languages in contrast*. 53–62. Lund : Lund University Press.
- Gémar, J.-C. 2002. « Le plus et le moins-disant culturel du texte juridique. Langue, culture et équivalence ». *Meta* 47(2), 163–176.
- Gentzler, E. 1993. *Contemporary Translation Theories*. London/New York : Routledge.
- Godard, B. 1984. « Translating and sexual difference ». *Resources for Feminist Research* 13(3), 13–16.
- Gœuriot, L. 2009. *Découverte et caractérisation des corpus comparables*. Thèse de doctorat. Université de Nantes : département d'informatique.

- Gosciny, R. et A. Uderzo. 1964. *Astérix gladiateur*. Paris : Dargaud S.A.
- Gurrutxaga, A., X. Saralegi et S. Ugartetxea. 2006. « ELeXBI, a basic tool for bilingual term extraction from Spanish-Basque parallel corpora ». Dans *Actes du 12e Congrès international de lexicographie (EURALEX,06)*. 159–165. Torino, Italie.
- Haghighi, A., P. Liang, T. Berg-Kirkpatrick et D. Klein. 2008. « Learning bilingual lexicons from monolingual corpora ». Dans *Proceedings of the 46<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL'08)*. 771–779. Columbus, Ohio, États-Unis.
- Harris, B. 1988. « Bi-text: A new concept in translation theory ». *Language Monthly* 54, 8–10.
- Hartmann, R. R. K. 1996. « Contrastive textology and corpus linguistics on the value of parallel texts ». *Language Sciences* 18(3–4), 947–957.
- House, J. 2008. « Beyond intervention: Universals in translation? » *Trans-kom* 1(1). [http://www.trans-kom.eu/bd01nr01/trans-kom\\_01\\_01\\_02\\_House\\_Beyond\\_Intervention.20080707.pdf](http://www.trans-kom.eu/bd01nr01/trans-kom_01_01_02_House_Beyond_Intervention.20080707.pdf) (page consultée le 9 janvier 2010).
- Hull, D. A. 1998. « A practical approach to terminology alignment ». Dans Bourigault, D., C. Jacquemin et M.-C. L'Homme (dir.). *Proceedings of the first workshop on Computational Terminology (COMPUTERM'98)*. 1–7. Montréal, Canada.
- Hull, D. A. 2001. « Software tools to support the construction of bilingual terminology lexicons ». Dans Bourigault, D., C. Jacquemin et M.-C. L'Homme (dir.). *Recent advances in Computational Terminology*. 167–182. Amsterdam/ Philadelphia : John Benjamins Publishing Company.
- Isabelle, P. 1992. « La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie ». *Meta* 37(4), 721–731.
- Ismail, A. et S. Manandhar. 2009. « Utilizing contextually relevant terms in bilingual lexicon extraction ». Dans *NAACL-2009 Workshop on Unsupervised and Minimally Supervised Learning off Lexical Semantics*. Presentation PowerPoint 1–37. Boulder, Colorado, États-Unis.

- ISO 1087-1. 2000. *Terminology work – Vocabulary – Part 1 : Theory and application. Travaux terminologiques – Vocabulaire – Partie 1 : Théorie et application.*
- ITWG (International Terminology Working Group). 1999. *Guidelines for Forming Language Equivalents: A Model Based on the Art & Architecture Thesaurus.* [http://www.pearltrees.com/#/N-u=1\\_250643&N-p=20987329&N-fa=2552276&N-f=1\\_2900800&N-play=1&N-s=1\\_2900800](http://www.pearltrees.com/#/N-u=1_250643&N-p=20987329&N-fa=2552276&N-f=1_2900800&N-play=1&N-s=1_2900800) (page consultée le 13 juin 2011).
- Jakobson, R. 2004 [1959]. « On linguistic aspects of translation ». Dans Venuti, L. (dir.). *The Translation Studies Reader.* 113–118. London/New York : Routledge.
- Johansson, S. 2000. « Towards a multilingual corpus for contrastive analysis and translation studies ». Reports from the project *Languages in Contrast* (Språk i kontrast). [www.hf.uio.no/ilos/forskning/prosjekter/sprik/docs/pdf/sj/johansson1.pdf](http://www.hf.uio.no/ilos/forskning/prosjekter/sprik/docs/pdf/sj/johansson1.pdf) (page consultée le 17 juillet 2011).
- Jousse, A.-L. 2002. *Dérivation sémantique et morphologique de termes, analyse en corpus specialise et modélisation au moyen des fonctions lexicales.* Mémoire de maîtrise en linguistique. Université du Maine, Le Mans.
- Kay, M. et M. Röscheisen. 1988. « Text-Translation alignment ». *Rapport technique, Xerox Palo Alto Research Center.*
- Kilgarriff, A. 2007. « Googleology is bad science ». *Computational Linguistics* 33(1), 147–151.
- Kit, C. et X. Liu. 2008. « Measuring mono-word termhood by rank difference via corpus comparison ». *Terminology*, 14(2), 204–229.
- Knight, K. et J. Graehl. 1998. « Machine transliteration ». Dans Cohen, P. R. et W. Wahlster (dir.). *Proceedings of the 3<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics (ACL'97).* 128–135. Madrid, Espagne.
- Kocourek, R. 1991. *La langue française de la technique et de la science. Vers une linguistique de la langue savante.* Wiesbaden : Oscar Brandstetter.
- Koehn, P. et K. Knight. 2002. « Learning a translation lexicon from monolingual corpora ». Dans *Proceedings of the ACL'02 workshop on Unsupervised lexical acquisition.* 9–16. Philadelphia, Philadelphie, États-Unis.

- Koller, W. 1979. *Einführung in die Übersetzungswissenschaft*. Heidelberg/Wiesbaden : Quelle and Meyer.
- Kraif, O. 1999. « Identification des cognats et alignement bi-textuel : une étude empirique ». Dans *Actes de la 6e conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'99)*. 205–214. Cargèse, France. [http://www.atala.org/doc/actes\\_taln/AC\\_0008.pdf](http://www.atala.org/doc/actes_taln/AC_0008.pdf) (page consultée le 31 mars 2007).
- Kraif, O. 2000. « Extraction automatique de correspondances lexicales évaluation d'indices et d'algorithmes ». Dans *Actes de la 7e conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'00)*. 225–236. Lausanne, Suisse.
- Kraif, O. 2001. *Constitution et exploitation de bi-textes pour l'Aide à la traduction*. Thèse de doctorat, sous la dir. de H. Zinglé, Université de Nice, Sophia Antipolis.
- Kraif, O. 2008. « Extraction automatique de lexique bilingue : application pour la recherche d'exemples en lexicographie ». Dans Maniez, F., P. Dury, N. Arlin et C. Rougemont (dir.). *Corpus et dictionnaires de langue de spécialités, Travaux du CRTT*. 67–87. Grenoble : Presses universitaires de Grenoble
- Kraif, O. 2010. *Page personnelle, Alinea, aide d'Alinea*. <http://w3.u-grenoble3.fr/kraif/index.php> (page consultée le 19 juin 2010).
- Kraif, O. et B. Chen. 2004. « Combining clues for lexical level aligning using the Null hypothesis approach ». Dans *Proceedings of Coling 2004*. 1261–1264. Genève, Suisse.
- L'Homme, M.-C. 2004. *La terminologie : principes et techniques*. Montréal : Les Presses de l'Université de Montréal.
- L'Homme, M.-C. 2005. « Sur la notion de “termes” ». *Meta* 50(4), 1112–1132.
- L'Homme, M.-C. 2005a. « Glossaire des termes de traductique ». *TRA2000A*. Montréal : Université de Montréal.
- Labbé, C. et D. Labbé. 2001. « Que mesure la spécificité du vocabulaire? ». *Lexicometria* 3. <http://lexicometrica.univ-paris3.fr/article/numero3Resumes.htm> (page consultée le 12 mai 2011).

- Langlais, P. 1997. « Alignement de corpus bilingues : intérêt, algorithmes et évaluation ». Dans *Bulletin de Linguistique Appliquée et Générale, numéro Hors Série*. 245–254. Université de Franche-Comté, France.
- Langlais, P., F. Yvon et P. Zweigenbaum. 2009. « Improvements in analogical learning: application to translating multi-terms of the medical domain ». Dans *Proceedings 12th Conference of the European Chapter of the ACL (EACL 2009)*. 487–495. Athènes, Grèce.
- Laroche, A. et P. Langlais. 2010. « Revisiting context-based projection methods for term-translation spotting in comparable corpora ». Dans *Proceedings of the 23<sup>rd</sup> International Conference on Computational Linguistics (COLING '10)*. 617–625. Beijing, Chine.
- Laviosa, S. 1997. « How comparable can ‘comparable corpora’ be? » *Target* 9, 289–319.
- Laviosa, S. 1998. « Core patterns of lexical use in a comparable corpus of English narrative prose ». *Meta* 43(4), 557–570.
- Laws, F., L. Michelbacher, B. Dorow, C. Scheible, U. Heid et H. Schütze. 2010. « A linguistically grounded graph model for bilingual lexicon extraction ». Dans *Proceedings of the 23rd International Conference on Computational Linguistics Posters (2010)*. 614–622. Beijing, Chine.
- Le nouveau petit Robert. 2002. *Dictionnaire de la langue française*. Paris : Dictionnaire le Robert.
- Le Serrec, A. 2008. *Étude sur l'équivalence de termes extraits automatiquement d'un corpus parallèle : contribution à l'extraction terminologique bilingue*. Mémoire de maîtrise en traduction. Université de Montréal : Département de linguistique et de traduction.
- Le Serrec, A., M.-C. L'Homme, P. Drouin et O. Kraif. 2010. « Automating the compilation of specialized dictionaries: use and analysis of term extraction and lexical alignment ». *Terminology* 16(1), 77–106.
- Léon, S. 2008. *Acquisition automatique de traduction d'unités lexicales complexes à partir du Web*. Thèse de doctorat. Université de Provence : Cognition, Langage, Éducation.

- Leonardi, V. 2000. « Equivalence in translation : Between myth and reality ». *Translation Journal* 4(4), <http://accurapid.com/journal/14equiv.htm> (page consultée le 29 novembre 2009).
- Liberman, M. 2004. « 46 somali words for camel ». *Language Log*. [http://itre.cis.upenn.edu/~myl/language-log/archives/2004\\_02.html](http://itre.cis.upenn.edu/~myl/language-log/archives/2004_02.html) (page consultée le 28 mai 2011).
- Lison, P. 2005. « Extraction de terminologies bilingues à partir de corpus comparables ». Rapport final du projet de TAL. Université Catholique de Louvain.
- Lotbinière-Harwood, S. (de). 1991. *Re-belle et infidèle : la traduction comme pratique de réécriture au féminin. The Body Bilingual : Translation as a Rewriting in the Feminine*. Montréal : Women's Press/Édition du remue-ménage.
- Lyons, J. 1970. *Linguistique général – Introduction à la linguistique théorique*. Paris : Larousse.
- Maia, B. 2003. « What are comparable corpora? ». Dans *Proceedings of the pre-conference workshop on Multilingual Corpora: Linguistic requirements and technical perspectives*. 27–31. Lancaster, Grande Bretagne.
- Malik, A., C. Boitet et P. Bhattacharyya. 2008. « Hindi Urdu Machine Transliteration using Finite-state Transducers ». Dans *Proceedings of the 22<sup>nd</sup> International Conference on Computational Linguistics (COLING'08)*. 537–544. Manchester, UK.
- Maniez, F. 2001. « Extraction d'une phraséologie bilingue en langue de spécialité : corpus parallèles et corpus comparables ». *Meta* 46(3), 552–563.
- Mauranen, A. 2002. « Will 'translationese' ruin a contrastive study? » *Languages in Contrast* 2(2), 161–185.
- McEnery, A. et Z. Xiao. 2007. « Parallel and comparable corpora : What are they up to? » Dans *Incorporating Corpora: translation and the linguist. Translating Europe. Multilingual Matters*. 2–13. Clevedon, Grande-Bretagne.
- Morin, E. 2009. « Apport d'un corpus comparable déséquilibré à l'extraction de lexiques bilingues ». Dans *Actes, 16e Conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'09)*. Senlis, France.

- Morin, E. et Daille. 2006. « Comparabilité de corpus et fouille terminologique multilingue ». *Traitement automatique des langues (TAL)* 47(1), 113–136.
- Morin, E., B. Daille, K. Takeuchi et K. Kageura. 2007. « Bilingual terminology mining – Using brain, not brawn comparable corpora ». Dans *Proceedings of the 45<sup>th</sup> Annual Meeting of the Association of Computational Linguistics*. 664–671. Prague, République Chèque.
- Morin, E., S. Dufour-Kowalski et B. Daille. 2004. « Extraction de terminologies bilingues à partir de corpus comparables ». Dans *Actes de la 11<sup>e</sup> Conférence annuelle sur le Traitement Automatique des Langues Naturelles (TALN'04)*. 309–318. Fès, Maroc.
- Nakagawa, H. et T. Mori. 2002. “A simple but powerful automatic term extraction method”. Dans *Proceedings of the International Conference On Computational Linguistics (COLING'02)*. 29–35. Taipei.
- Nakamura-Delloye, Y. 2005. « Système AIALeR : Alignement au niveau phrastique des textes parallèles français-japonais ». Dans *Actes de la conférence Traitement Automatique des Langues Naturelles (TALN'05)*. 285–294. Dourdan, France.
- Névéol, A. et S. Ozdowska. 2005. « Terminologie médicale bilingue anglais–français : usages clinique et législatif ». *Glottopol* 8. 5–21.
- Nida, E. 1945. « Linguistics and ethnology in translation problems ». *Word* 1, 194–208.
- Nida, E. 1964. *Toward a Science of Translating, with Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden, Holland : Brill.
- Niranjana, T. 1992. *Siting Translation: History, Post-structuralism and the Colonial Text*. Berkeley : University of California Press.
- Ozdowska, S. 2006. *ALIBI, un système d'ALIGNement Bilingue à base de règles de propagation syntaxique*. Thèse de doctorat, Université de Toulouse II-Le Mirail.
- Ozdowska, S. et D. Bourigault. 2004. « Détection de relations bilingues entre termes à partir d'une analyse syntaxique de corpus ». Dans *Actes du 14<sup>e</sup> Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'04)*. Toulouse, France.
- Ozdowska, S. et V. Claveau. 2005. « Alignement de mots par apprentissage artificiel de règles de propagation syntaxique en corpus ». Dans *Actes de la conférence*

- Traitement automatique des langues naturelles (TALN'05)*. 243–252. Dourdan, France.
- Pearson, J. 1998. *Terms in Context*. Amsterdam/Philadelphia : John Benjamins.
- Pearson, J. 2000. « Une tentative d'exploitation bi-directionnelle d'un corpus bilingue ». Dans *Cahiers de grammaire 25*, « *Sémantique et Corpus* », 53–69.
- Philip, G. 2009. « Arriving at equivalence: Making a case for comparable general reference corpora in translation studies ». Dans Beeby, A., P. Rodríguez Inés et P. Sánchez-Gijón (dir.). *Corpus Use and Translating*. 59–73. Amsterdam/Philadelphie : John Benjamins
- Pic, E. 2007-2008. « Faire de la terminologie en droit? » *Cahiers du CIEL*. <http://www.eila.univ-paris-diderot.fr/recherche/clillac/ciel/cahiers/2007-2008> (page consultée le 16 juillet 2011).
- Polguère, A. 2003. *Lexicologie et sémantique lexicale – Notions fondamentales*. Montréal : Les presses de l'Université de Montréal.
- Prochasson, E. 2008. « Observation des translittérations dans un corpus comparable spécialisé trilingue ». *Journée des Doctorants de l'École doctorale STIM*.
- Prochasson, E. 2009. *Alignement multilingue en corpus comparables spécialisés*. Thèse de doctorat en informatique. Université de Nantes : École doctorale STIM « Sciences et technologies de l'information et des matériaux ».
- Prochasson, E. 2009a. « Alignement multilingue en corpus comparables spécialisés ». *Présentation PowerPoint de la Thèse de doctorat*. 1–52.
- Prochasson, E. et E. Morin. 2009. « Influence des points d'ancrage pour l'extraction lexicale bilingue à partir de corpus comparables spécialisés ». Dans *Actes de la Conférence sur le Traitement Automatique des Langues Naturelles (TALN'09)*. Senlis, France.
- Pullum, G. K. 1991. « The great Eskimo vocabulary hoax and other irreverent essays on the study of language ». Chicago: University of Chicago Press.
- Ramavonirina, O., E. Mbodj, A. Baboya et M. Diki-Kidiri. 2007. « Terminologie culturelle : concepts fondamentaux ». *Colloque international de Terminologie*. Gatineau Canada.



- Rapp, R. 1995. « Identifying word translations in non-parallel texts ». Dans *Proceedings of the 33<sup>rd</sup> Annual Meeting of the Association for Computational Linguistics (ACL '95)*. 320–322. Boston, Massachusetts.
- Rapp, R. 1999. « Automatic Identification of word translations from unrelated English and German corpora ». Dans *Proceedings of the 37<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL '99)*. 519–526. College Park, MD, États-Unis.
- Robert & Collins. 2003. *Le grand Robert et Collons électronique*. Version 1.0. Paris : Dictionnaire Le Robert.
- Rondeau, G. 1981. *Introduction à la terminologie*. Montréal : Centre éducatif et culturel.
- Rosset, S. et M. Jardino. 2008. « Comparaison de documents : mesures de similarité et mesures de distance ». Site Internet de Sophie Rosset, du groupe Traitement du langage parlé. [www.limsi.fr/Individu/rosset/similarite2.pdf](http://www.limsi.fr/Individu/rosset/similarite2.pdf) (page consultée le 17 juillet 2011).
- Russell, W. A. 1970. « The complete German language norms for responses to 100 words from the Kent-Rosanoff word association test ». Dans Postman, L. et G. Keppel (dir.). *Norms of word association*. 53–94. New York: Academic Press.
- Sadat, F., M. Yoshikawa et S. Uemura. 2003. « Learning bilingual translations from comparable corpora to cross-language information retrieval: Hybrid statistics-based and linguistics-based approach ». Dans *Proceedings of the 6<sup>th</sup> International Workshop on Information Retrieval with Asian Languages*. 57–64. Sapporo, Japon.
- Sager, J. C. 1994. « Reflections on terminological translation equivalents ». *Lebende Sprachen* 2(94), 55–57.
- Sanz Vicente, L. et J. García Palacios. 2010. « Proposals to standardize remote sensing terminology in Spanish ». Dans Thelen, M. et F. Steurs (dir.). *Terminology in Everyday Life*. 195–210. Amsterdam/Philadelphie : John Benjamins.
- Schmid, H. 1994. « Probabilistic part-of-speech tagging using decision trees ». Dans *Proceedings of International Conference on New Methods on Language Proceeding*. 44–49. Manchester, Royaume-Uni.
- Shahzad, I., K. Ohtake, S. Masuyama et K. Yamamoto. 1999. « Identifying translations of compound using non-aligned corpora ». Dans *Proceedings of the Workshop on*

- Multilingual Information Processing and Asian Language Processing*. 108–113. Beijing, Chine.
- Shao, L. et H. T. Ng. 2004. « Mining new word translations from comparable corpora ». Dans *Proceedings of the 20<sup>th</sup> International Conference on Computational Linguistics (COLING'04)*. 618–624. Genève, Suisse.
- Simard, M., G. Foster et P. Isabelle. 1992. « Using cognates to align sentences in bilingual corpora ». Dans *Proceedings of the 4<sup>th</sup> International Conference on Theoretical and Methodological Issues in Machine Translation (TMI'92)*. 67–81. Montréal, Canada.
- Simon, S. 1996. *Gender in Translation: Cultural Identity and the Politics of Transmission*. Londre/New York : Routledge.
- Sinclair, J. 1996. « Preliminary recommendations on Corpus Typology ». *Rapport technique. EAGLES*. <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html> (page consultée le 28 décembre 2009).
- Snell-Hornby, M. 1988. *Translation Studies: An Integrated Approach*. Amsterdam/Philedelphia : John Benjamins.
- Spivak, G. 2004. « The politics of translation ». Dans Venuti, L. (dir.). *The Translation Studies Reader*. 397–416. London/New York : Routledge.
- Svensén, B. 1993. *Principles and Methods of Dictionary-Making*. Oxford/New York : Oxford University Press.
- Svensén, B. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-making*. Cambridge : Cambridge University Press.
- Tanaka, K. et H. Iwasaki. 1996. « Extraction of lexical translations from non-aligned corpora ». Dans *Proceedings of the 16<sup>th</sup> International Conference on Computational Linguistics (COLING'96)* 2, 580–585. Copenhague, Danemark.
- Temmerman, R. 2000. « Une théorie réaliste de la terminologie : le sociocognitivism ». *Terminologies Nouvelles* 21, 58–64.
- Teubert, W. 1996. « Comparable or parallel corpora? » *International Journal of Lexicography* 9(3), 238–264.
- Teubert, W. 2001. « Corpus linguistics and lexicography ». *International Journal of Corpus Linguistics* 6 (special issue), 125–153.

- Tiedemann, J. 2003. *Recycling Translations. Extraction of Lexical Data from Parallel Corpora and their Application in Natural Language Processing*. Thèse de doctorat, Université d'Uppsala, Uppsala.  
<http://publications.uu.se/theses/abstract.xsql?dbid=3791> (page consultée le 24 juillet 2007).
- Tirkkonen-Condit, S. 2002. « Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language ». *Target* 14(2), 207–220.
- Tournier, J. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris/Genève : Champion-Slatkine.
- Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam, Philadelphia : John Benjamins.
- Tran, T. D., A. Burgun et N. Garcelon. 2003. « Acquisition semi-automatique de terminologie bilingue en biologie moléculaire à partir des corpus comparables ». Dans *Terminologie et Intelligence Artificielle (TIA'03)*. 166–175. Strasbourg, France.
- Truffaut, L. 2004. *Abécédaire de la traduction professionnelle*, vol. 2. Bruxelles : Les Éditions du Hazard.
- Tymoczko, M. 1998. « Computerized corpora and Translation Studies ». *Meta* 43(4), 652–659.
- Van Campenhoudt, M. 1996. « Réseau notionnel, intelligence artificielle et équivalence en terminologie multilingue : essai de modélisation ». Dans Clas A., P. Thoiron et H. Béjoint (dir.). *Lexicomatique et dictionnaires, 4<sup>e</sup> journées scientifiques du réseau thématique Lexicologie, terminologie, traduction*, Université Lumière (Lyon II), 28-30 septembre 1995. 281–306. Montréal : AUPELF-UREF et Beyrouth, F.M.A.
- Van Campenhoudt, M. 2001. « Pour une approche sémantique du terme et de ses équivalents ». *International Journal of Lexicography* 14(3), 181–209.
- Van Campenhoudt, M. 2003. « L'évolution des dictionnaires de traduction du domaine maritime au XIX<sup>e</sup> siècle : aux sources de *De la quille à la pomme de mâ*t ». Dans *Chronique d'histoire maritime*, revue de la Société française d'histoire maritime 52, 83–97.

- Van Der Eijk, P. 1993. « Automating the acquisition of bilingual terminology ». Dans *Proceedings of the 6<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*. 113–119. Utrecht, Pays-Bas.
- Vandaele, S. 2001. « Bibliographie sélective pour la traduction des sciences biomédicales ». *Meta* 46(1), 117–127.
- Vandaele, S. et M. Raffo. 2008. « Signification lexicales et notionnelles dans les domaines de spécialité ». Article accepté en avril 2008, à paraître (Festschrift en l'honneur d'Henri Béjoint. Université Lumière Lyon-2, Lyon).
- Véronis, J. 2000. « Alignement de corpus multilingues ». Dans Pierrel, J.-M. (dir.). *Ingénierie des langues*. 151–172. Paris : Éditions Hermès.
- Véronis, J. 2000a. *Parallel Text Processing : Alignment and Use of Translation Corpora*. Dordrecht/Boston/London : Kluwer Academic Publishers.
- Vieira, E. 1999. « Liberating calibans: Readings of antropofagia and Haroldo de Campos' Poetics of transcreation ». Dans Basnett, S. et H. Trivedi (dir.). *Post-Colonial Translation: Theory and Practice*. 95–113. Londres/New York : Pinter.
- Vinay, J.-P. et J. Darbelnet. 1977 [1958]. *Stylistique comparée du français et de l'anglais*. Paris : Les Éditions Didier.
- Vintar, Š. 2010. « Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation ». *Terminology* 16(2), 141–158.
- Wong, W., W. Liu et M. Bennamoun. 2008. « Determination of unithood and termhood for term recognition ». Dans Song, M. et Y. Wu (dir.). *Handbook of Research on Text and Web Mining Technologies*. 500–529. Hershey (PA) : IGI Global.
- Wu, D. 2000. « Bracketing and aligning words and constituents in parallel text using Stochastic Inversion Transduction Grammars ». Dans Véronis, J. (dir.). *Parallel Text Processing : Alignment and Use of Translation Corpora*. 139–167. Dordrecht : Kluwer Academic Publishers.
- Wüster, E. 1981. « L'étude générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses ». Dans Rondeau, G. et H. Felber (dir.). *Textes choisis de terminologie. I. Fondements théoriques de la terminologie*. 55–114. Québec : GIRSTERM.

- Yoda, L. A. 2005. *La traduction médicale du français vers le mooré et le bisa: un cas de communication interculturelle au Burkina Faso*. Thèse de doctorat, Groningen (Pays –Bas), Rijksuniversiteit Groningen.
- Yu, K. et J. Tsujii. 2009. « Bilingual dictionary extraction from Wikipedia ». Dans *Machine Translation Summit XII*. Ottawa, Ontario, Canada. <http://www.mt-archive.info/MTS-2009-TOC.htm> (page consultée le 17 juillet 2011).
- Zanettin, F. 1998. « Bilingual corpora and the training of translators ». *Meta* 43, 616–630.
- Zang, X et A. Fang. 2010. « An ATE System based on Probabilistic Relations between Terms and Syntactic Functions ». Dans *10th International Conference on Statistical Analysis of Textual Data*. Sapienza, Italie.
- Zgusta, L. 1971. *Manual of lexicography*. La Hague : Mouton.
- Zgusta, L. 1987. « Translational equivalence in the bilingual dictionary ». *Lexicographica Series Maior* 1, 147–154.
- Zweigenbaum, P. 2006. « Corpus parallèles et corpus comparables : introduction ». <http://www.limsi.fr/~pz/p11m2r-2006/corpus-paralleles.pdf> (page consultée le 29 juillet 2011).
- Zweigenbaum, P. et B. Habert. 2006. « Faire se rencontrer les parallèles : regards croisés sur l’acquisition lexicale monolingue et multilingue ». *Glottopol – revue de sociolinguistique en ligne* 8, 22–44.



## ANNEXES

### Annexe A : Bibliographie des corpus

#### Bibliographie des textes anglais et français du corpus parallèle

CC_AEE_signals_2009	Agence européenne pour l'environnement. 2009. « Thèmes environnementaux clés pour l'Europe ». <a href="http://www.eea.europa.eu/fr/publications/signals-2009">http://www.eea.europa.eu/fr/publications/signals-2009</a> (page consultée le 9 mars 2010).
CC_AEE_vulnerability_2005	Agence européenne pour l'environnement. 2005. « Changements climatiques et inondations liées aux rivières et fleuves en Europe ». <a href="http://reports.eea.europa.eu/index_table?lang=French">http://reports.eea.europa.eu/index_table?lang=French</a> (page consultée le 11 janvier 2007)
CC_CCE_carbone_2001	CCE. 2001. « Le Mexique et les nouveaux marchés du carbone ». <a href="http://www.cec.org/Page.asp?PageID=122&amp;ContentID=2602&amp;SiteNodeID=437&amp;AA_SiteLanguageID=2">http://www.cec.org/Page.asp?PageID=122&amp;ContentID=2602&amp;SiteNodeID=437&amp;AA_SiteLanguageID=2</a> (page consultée le 20 mars 2010).
CC_CCE_trends_2003	CCE. 2003. « Comprendre et prévoir les changements environnementaux en Amérique du Nord ». <a href="http://www.cec.org/Page.asp?PageID=122&amp;ContentID=2601&amp;SiteNodeID=437&amp;AA_SiteLanguageID=2">http://www.cec.org/Page.asp?PageID=122&amp;ContentID=2601&amp;SiteNodeID=437&amp;AA_SiteLanguageID=2</a> (page consultée le 16 mars 2010).
CC_CE_changclim_2002	EUROPA. 2002. « Opter pour un avenir plus vert ». <a href="http://ec.europa.eu/publications/booklets/move/32/txt_fr.pdf">http://ec.europa.eu/publications/booklets/move/32/txt_fr.pdf</a> (page consultée le 6 janvier 2007).
CC_CES_afrique_2007	Nations Unies, Conseil économique et social. 2007. « Débat interactif: Afrique croissance de l'agriculture africaine dans le contexte du changement climatique et de la désertification. Quel doit être le rôle des partenaires de développement? ». <a href="http://unbisnet.un.org/indexf.htm">http://unbisnet.un.org/indexf.htm</a> (page consultée le 6 mars 2010).
CC_CES_agriculture_2008	Nations Unies, Conseil économique et social. 2008. « Questions relatives à l'agriculture, au développement rural, à la sécheresse, à la désertification et aux sols, qui ont une incidence sur le développement durable dans la région de la CEE : réalisations, tendances et problèmes ». <a href="http://unbisnet.un.org/indexf.htm">http://unbisnet.un.org/indexf.htm</a> (page consultée le 6 mars 2010).
CC_CES_ozone_2007	Nations Unies, Conseil économique et social. 2007. « Organe exécutif de la convention sur la pollution atmosphérique transfrontière à longue distance ». <a href="http://unbisnet.un.org/indexf.htm">http://unbisnet.un.org/indexf.htm</a> (page consultée le 6 mars 2010).
CC_ECA_changclim_2005	Environnement Canada. 2005. « Une introduction au changement climatique – une perspective canadienne ». <a href="http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html">http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html</a> (page consultée le 15 janvier 2007).
CC_ECA_extremes_1998	Environnement Canada. 1998. « Phénomènes météorologiques extrêmes et changement climatique ». <a href="http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/understanding/ccd/ccd_9801/CCD_9801_f.pdf">http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/understanding/ccd/ccd_9801/CCD_9801_f.pdf</a> (page consultée le 20 janvier 2007).
CC_ECA_FAQ_2002	Environnement Canada. 2002. « FAQ - Foires aux questions – SCC ».

	<a href="http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html">http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html</a> (page consultée le 15 janvier 2007).
CC_ECA_projections_2000	Environnement Canada. 2000. « Projections du climat du Canada. CCD 00-01 ». <a href="http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html">http://www.msc-smc.ec.gc.ca/education/scienceofclimatechange/publications/reports_papers/index_f.html</a> (page consultée le 15 janvier 2007).
CC_IPCC_attenuation_2001	Groupe d'experts intergouvernemental sur l'évolution du climat. 2001. « Bilan 2001 des changements climatiques : mesures d'atténuations ». <a href="http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg3sum.pdf">http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg3sum.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_aviation_1999	Groupe d'experts intergouvernemental sur l'évolution du climat. 1999. « L'aviation et l'atmosphère planétaire ». <a href="http://www.grida.no/climate/ipcc/spmpdf/av-f.pdf">http://www.grida.no/climate/ipcc/spmpdf/av-f.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_bilan_2001	Groupe d'experts intergouvernemental sur l'évolution du climat. 2001. « Bilan 2001 des changements climatiques : Conséquences, adaptation et vulnérabilité ». <a href="http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg2sum.pdf">http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg2sum.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_biodiversité_2002	Groupe d'experts intergouvernemental sur l'évolution du climat. 2002. « Les changements climatiques et la biodiversité ». <a href="http://www.ipcc.ch/pub/tpbiodiv_f.pdf">http://www.ipcc.ch/pub/tpbiodiv_f.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_dioxyde_2005	Groupe d'experts intergouvernemental sur l'évolution du climat. 2005. « Piégeage et stockage du dioxyde de carbone ». <a href="http://www.ipcc.ch/activity/srccs/IPCC%20F.pdf">http://www.ipcc.ch/activity/srccs/IPCC%20F.pdf</a> (page consultée le 07 octobre 2006).
CC_IPCC_eau_2008	Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). 2008. « Le changement climatique et l'eau ». <a href="http://www.grida.no/climate/ipcc_tar/wg1/pdf/WG1_TAR-FRONT.PDF">http://www.grida.no/climate/ipcc_tar/wg1/pdf/WG1_TAR-FRONT.PDF</a> (page consultée le 3 avril 2010).
CC_IPCC_émissions_1997	Groupe d'experts intergouvernemental sur l'évolution du climat. 1997. « Incidences des propositions de limitation des émissions de CO <sub>2</sub> ». <a href="http://www.ipcc.ch/pub/IPCCTP.IV(F).pdf">http://www.ipcc.ch/pub/IPCCTP.IV(F).pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_GES_2000	Groupe d'experts intergouvernemental sur l'évolution du climat. 2000. « Scénario d'émission ». <a href="http://www.grida.no/climate/ipcc/spmpdf/sres-f.pdf">http://www.grida.no/climate/ipcc/spmpdf/sres-f.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_LULUCF_2000	Groupe d'experts intergouvernemental sur l'évolution du climat. 2000. « L'utilisation des terres, le changement d'affectation des terres et la foresterie ». <a href="http://www.grida.no/climate/ipcc/spmpdf/srl-f.pdf">http://www.grida.no/climate/ipcc/spmpdf/srl-f.pdf</a> (page consultée le 07 octobre 2006).
CC_IPCC_methodologie_2000	Groupe d'experts intergouvernemental sur l'évolution du climat. 2000. « Questions méthodologiques et technologiques dans le transfert de technologie ». <a href="http://www.grida.no/climate/ipcc/spmpdf/srtt-f.pdf">http://www.grida.no/climate/ipcc/spmpdf/srtt-f.pdf</a> (page consultée le 07 octobre 2006).
CC_IPCC_modele_1997	Groupe d'experts intergouvernemental sur l'évolution du climat. 1997. « Introduction aux modèles climatiques simples employés dans le



	Deuxième Rapport ». <a href="http://www.ipcc.ch/pub/IPCCTP.II(F).pdf">http://www.ipcc.ch/pub/IPCCTP.II(F).pdf</a> (page consultée le 6 janvier 2007) d'évaluation du GIE, . <a href="http://www.ipcc.ch/pub/IPCCTP.II(F).pdf">http://www.ipcc.ch/pub/IPCCTP.II(F).pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_ozone_2005	Groupe d'experts intergouvernemental sur l'évolution du climat. 2005. « Préservation de la couche d'ozone et du système climatique planétaire: Questions relatives aux hydrofluorocarbures et aux hydrocarbures perfluorés ». <a href="http://www.ipcc.ch/activity/specialrprt05/IPCC-TEAP_F.pdf">http://www.ipcc.ch/activity/specialrprt05/IPCC-TEAP_F.pdf</a> (page consultée le 07 octobre 2006).
CC_IPCC_region_1997	Groupe d'experts intergouvernemental sur l'évolution du climat. 1997. « Evaluation de la vulnérabilité ». <a href="http://www.grida.no/climate/ipcc/spmpdf/region-f.pdf">http://www.grida.no/climate/ipcc/spmpdf/region-f.pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_stabilisationdesgaz_1997	Groupe d'experts intergouvernemental sur l'évolution du climat. 1997. « Stabilisation des gaz atmosphériques à effet de serre : conséquences physiques biologiques et socio-économiques ». <a href="http://www.ipcc.ch/pub/IPCCTP.III(F).pdf">http://www.ipcc.ch/pub/IPCCTP.III(F).pdf</a> (page consultée le 7 octobre 2006).
CC_IPCC_synthese_2001	Groupe d'experts intergouvernemental sur l'évolution du climat. 2001. « Bilan 2001 des changements climatiques : Rapport de synthèse ». <a href="http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg1sum.pdf">http://www.grida.no/climate/ipcc_tar/vol4/french/pdf/wg1sum.pdf</a> (page consultée le 7 octobre 2006).
CC_NRTEE_options_1999	NRTEE. 1999. « Programme canadien d'échange de droits d'émission de gaz à effet de serre : options envisagées ». <a href="http://www.nrtee-trnee.com/fra/publications.o.ptions-echange-droits-emission/index-options-echange-droits-emission.php">http://www.nrtee-trnee.com/fra/publications.o.ptions-echange-droits-emission/index-options-echange-droits-emission.php</a> (page consultée le 20 mars 2010).
CC_NRTEE_true_north_2009	NRTEE. 2009. « Table ronde nationale sur l'environnement et l'économie ». <a href="http://www.nrtee-trnee.com/fra/index.php">http://www.nrtee-trnee.com/fra/index.php</a> (page consultée le 9 mars 2010).
CC_UNU_overview_2008	Nations Unies, Assemblée générale. 2008. « Aperçu des activités menées par le système des Nations Unies concernant les changements climatiques ». <a href="http://unbisnet.un.org/indexf.htm">http://unbisnet.un.org/indexf.htm</a> (page consultée le 2 mars 2010).
CC_PNUE_bioseq_2009	PNUE. 2009. « Le rôle des écosystèmes dans l'atténuation des changements climatiques ». <a href="http://www.unep.org/ecosystemmanagement/Publications/Publication/tabid/439/language/en-US/Default.aspx?BookID=4027">http://www.unep.org/ecosystemmanagement/Publications/Publication/tabid/439/language/en-US/Default.aspx?BookID=4027</a> (page consultée le 16 mars 2010).
CC_RNCA_adaptation_2004	RNCA. 2004. « Impacts et adaptation liés au changement climatique : perspective canadienne ». <a href="http://adaptation.nrcan.gc.ca/perspective/index_f.php">http://adaptation.nrcan.gc.ca/perspective/index_f.php</a> (page consultée le 24 mars 2010).
CC_RNCA_impacts_2008	RNCA. 2008. « Vivre avec les changements climatiques au Canada : édition 2007 ». <a href="http://adaptation.nrcan.gc.ca/assess/2007/index_f.php">http://adaptation.nrcan.gc.ca/assess/2007/index_f.php</a> (page consultée le 24 mars 2010).
CC_UNESCO_aworldofscience_2006	UNESCO. 2006. « Planète science, Bulletin trimestriel d'information sur les sciences exactes et naturelles, vol. 4, no 1 mars 2006 ». <a href="http://unesdoc.unesco.org/images/0014/001432/143225f.pdf">http://unesdoc.unesco.org/images/0014/001432/143225f.pdf</a> (page consultée le 7 octobre 2006).

CC_UNESCO_patrimoine_2009	UNESCO. 2009. « Études de cas Changement climatique et patrimoine mondial ». <a href="http://whc.unesco.org/en/activities/473/">http://whc.unesco.org/en/activities/473/</a> (page consultée le 7 mars 2010).
CC_UNFCCC_action-concertee_2007	UNFCCC. 2007. « Rapport des cofacilitateurs du dialogue pour une action concertée à long terme destinée à permettre de faire face aux changements climatiques par un renforcement de l'application de la Convention ». <a href="http://unbisnet.un.org/">http://unbisnet.un.org/</a> (page consultée le 6 mars 2010).
CC_UNFCCC_resilience_2009	UNFCCC. 2009. « Rapport de l'atelier technique sur le renforcement de la résilience économique aux changements climatiques et sur la réduction de la dépendance à l'égard des secteurs économiques vulnérables, y compris par une diversification économique ». <a href="http://unbisnet.un.org/indexf.htm">http://unbisnet.un.org/indexf.htm</a> (page consultée le 6 mars 2010).
CC_UNFPA_femmes_2009	United Nations Population Fund (UNFPA). 2009. « Les femmes en première ligne ». <a href="Http://www.unfpa.org/public/publications/pid/4028">Http://www.unfpa.org/public/publications/pid/4028</a> (page consultée le 16 juin 2010).

## Bibliographie des textes français du corpus comparable

CC_AEE_Alpes_2009.fr	Agence européenne pour l'environnement. 2009. « Changement climatique régional et adaptation : les Alpes face au changement des ressources en eau ». <a href="http://www.eea.europa.eu/fr/publications/changement-climatique-regional-et-adaptation-2014-les-alpes-face-au-changement-des-ressources-en-eau">http://www.eea.europa.eu/fr/publications/changement-climatique-regional-et-adaptation-2014-les-alpes-face-au-changement-des-ressources-en-eau</a> (page consultée le 16 février 2010).
CC_Andre_energie_2001.fr	André, J.-C. 2001. « Réflexions sur le calendrier du débat énergie et climat ». <a href="http://www.sciencedirect.com/science">http://www.sciencedirect.com/science</a> (page consultée le 21 février 2010).
CC_Boe_hydrologie_2007.fr	Boe, J. 2007. « Changement global et cycle hydrologique : Une étude de régionalisation sur la France ». <a href="http://thesesups.ups-tlse.fr/227/">http://thesesups.ups-tlse.fr/227/</a> (page consultée le 4 avril 2010).
CC_Bourg_politique_2003.fr	Bourg, D. 2003. « Le déficit climatique : les limites des politiques publiques ». <a href="http://www.sciencedirect.com/science">http://www.sciencedirect.com/science</a> (page consultée le 21 février 2010).
CC_Campos_captage-et-stockage_2007.fr	Campos, S., M. Ha-Duong et M. Merad. 2007. « Synthèse de littérature sur l'acceptabilité sociale du captage et stockage du CO <sub>2</sub> ». <a href="http://www.centre-cired.fr/spip.php?article534">http://www.centre-cired.fr/spip.php?article534</a> (page consultée le 20 février 2010).
CC_CES_changement_2006.fr	Académie. 2006. « Partager la connaissance et ouvrir le dialogue : le changement climatique ». <a href="Http://www.changement-climatique.fr/rapport/">Http://www.changement-climatique.fr/rapport/</a> (page consultée le 17 avril 2008).
CC_CNRS_lettre14_2002.fr	Ministère des affaires étrangères. 2002. « Le réchauffement climatique : historique scientifique et politique – scénarii futurs ». <a href="Http://www.cnrs.fr/cw/dossiers/dosclim/biblio/pigb14/11_rechauffement.htm">Http://www.cnrs.fr/cw/dossiers/dosclim/biblio/pigb14/11_rechauffement.htm</a> (page consultée le 17 avril 2008).
CC_Doc_Francaise_changement_2009.fr	La documentation française. 2009. « Le changement climatique ». <a href="http://www.ladocumentationfrancaise.fr/dossiers/changement-climatique/index.shtml">http://www.ladocumentationfrancaise.fr/dossiers/changement-climatique/index.shtml</a> (page consultée le 10 février 2010).

CC_Dron_homme_2002.fr	Dron, D. 2002. « L'homme et le climat ». <a href="http://www.institutmontaigne.org/lhomme-et-le-climat-116.html">http://www.institutmontaigne.org/lhomme-et-le-climat-116.html</a> (page consultée le 21 février 2010).
CC_Duplessy_ocean_1997.fr	Duplessy, J.-C. 1997. Vers un refroidissement de l'Europe ? La Recherche, <a href="http://www.larecherche.fr/content/recherche/article?id=20545">http://www.larecherche.fr/content/recherche/article?id=20545</a>
CC_ESCRIME_simulation_2007.fr	IPSL et MÉTÉO-France. 2007. « Livre Blanc Escrime: étude des simulations climatiques ». <a href="http://escrime.ipsl.jussieu.fr/telecharger-le-livre.html">http://escrime.ipsl.jussieu.fr/telecharger-le-livre.html</a> (page consultée le 13 février 2010).
CC_Grenelle_atmosphere_2008.fr	Richert, P. 2008. « Air et atmosphère ». <a href="http://www.ladocumentationfrancaise.fr/rapports-publics/074000137/index.shtml">http://www.ladocumentationfrancaise.fr/rapports-publics/074000137/index.shtml</a> (page consultée le 20 février 2010).
CC_INRA_Bonduelle_1997.fr	Bonduelle, A et H Connor. 1997. « Changements climatiques et effet de serre. Des réponses au problème existent... Reste à les mettre en œuvre ! » <a href="http://www.inra.fr/dpenv/climglob.htm">http://www.inra.fr/dpenv/climglob.htm</a> (page consultée le 4 avril 2010).
CC_INRA_Falloux_1997.fr	Falloux, F. 1997. « Relever le défi de l'effet de serre par le marché? Mythe ou réalité, éthique et efficacité ». <a href="http://www.inra.fr/dpenv/climglob.htm">http://www.inra.fr/dpenv/climglob.htm</a> (page consultée le 5 avril 2010).
CC_INRA_Germon_1998.fr	Germon, J.-C. 1998. « Le rôle régulateur du sol dans le changement climatique ». <a href="http://www.inra.fr/dpenv/climglob.htm">http://www.inra.fr/dpenv/climglob.htm</a> (page consultée le 4 avril 2010).
CC_INRA_Ramos_2000.fr	Ramos, S. 2000. « Le réchauffement planétaire : de Kyoto à La Haye ». <a href="http://www.inra.fr/dpenv/climglob.htm">http://www.inra.fr/dpenv/climglob.htm</a> (page consultée le 5 avril 2010).
CC_INRA_Seguín_2002.fr	Seguín, B. 2002. « La recherche agronomique face aux gaz à effet de serre ». <a href="http://www.inra.fr/dpenv/climglob.htm">http://www.inra.fr/dpenv/climglob.htm</a> (page consultée le 5 avril 2010).
CC_INSU_changement_2007.fr	INSU. 2007. Les recherches françaises sur le changement climatique. <a href="http://www.insu.cnrs.fr/a2059,recherches-francaises-changement-climatique-2007.html">http://www.insu.cnrs.fr/a2059,recherches-francaises-changement-climatique-2007.html</a> (page consultée le 17 avril 2008).
CC_INSU_ocean-atmos_2005.fr	INSU. 2005. « Prospective Océan-Atmosphère 2006-2011 ». <a href="http://www.insu.cnrs.fr/a2171,prospective-ocean-atmosphere-2005.html">http://www.insu.cnrs.fr/a2171,prospective-ocean-atmosphere-2005.html</a> (page consultée le 17 avril 2008).
CC_Jacques-Le-Treut_chang_2004.fr	Jacques, G. et H. Le Treut. 2004. <i>Le changement climatique</i> . Paris : Éditions Unesco.
CC_Lellievre_mediterraneen_2008.fr	Lellievre, F., J.-B. Finot et S. Satger. 2008. « Le changement climatique récent et futur sur l'arc péri-méditerranéen ». <a href="http://www.climfourrel.fr">www.climfourrel.fr</a> (page consultée le 5 avril 2010).
CC_Lettre_changement_2003.fr	La Lettre de l'Académie des Sciences. 2003. « Les impacts de l'effet de serre ». <a href="http://www.academie-sciences.fr/publications/lettre/lettre_anciens_numeros.htm">http://www.academie-sciences.fr/publications/lettre/lettre_anciens_numeros.htm</a> (page consultée le 11 avril 2010).
CC_Magnan_vulnerabilite_2009.fr	Magnan, A. 2009. « La vulnérabilité des territoires littoraux au changement climatique : Mise au point conceptuelle et facteurs d'influence ». <a href="http://www.iddri.org/Publications/Collections/Analyses/La-vulnerabilite-des-territoires-littoraux-au-changement-climatique-Mise-au-point-conceptuelle-et-facteurs-d'influence">http://www.iddri.org/Publications/Collections/Analyses/La-vulnerabilite-des-territoires-littoraux-au-changement-climatique-Mise-au-point-conceptuelle-et-facteurs-d'influence</a> (page consultée le 11 février 2010).
CC_Martin_sylviculture_2005.fr	Martin, Y. 2005. « La forêt face au changement de climat ». <a href="http://www.cgm.org/rapports/Foret-climat-Yves-Martin-050805.pdf">www.cgm.org/rapports/Foret-climat-Yves-Martin-050805.pdf</a> (page consultée le 22 février 2010).

CC_ONERC_adaptation_2005.fr	Observatoire national des effets du réchauffement climatique. 2005. « Un climat à la dérive : comment s'adapter? ». <a href="http://www.ladocumentationfrancaise.fr/rapports-publics/054000414/index.shtml">Http://www.ladocumentationfrancaise.fr/rapports-publics/054000414/index.shtml</a> (page consultée le 16 février 2010).
CC_ONERC_Alpes_2008.fr	ONERC. 2008. « Changement climatiques dans les Alpes : Impacts et risques naturels ». <a href="http://www.ecologie.gouv.fr/-ONERC-.html">http://www.ecologie.gouv.fr/-ONERC-.html</a> (page consultée le 5 avril 2010).
CC_ONERC_antarctique_2009.fr	ONERC. 2009. « Impacts du changement climatique dans les îles subantarctiques ». <a href="http://www.ecologie.gouv.fr/-ONERC-.html">http://www.ecologie.gouv.fr/-ONERC-.html</a> (page consultée le 7 avril 2010).
CC_ONERC_erosion_2005.fr	ONERC. 2005. « Impact du changement climatique sur le patrimoine du Conservatoire du littoral ». <a href="http://www.ecologie.gouv.fr/-ONERC-.html">http://www.ecologie.gouv.fr/-ONERC-.html</a> (page consultée le 7 avril 2010).
CC_ONERC_impact_2008.fr	ONERC. 2008. « Impacts du changement climatique, adaptation et coûts associés en France ». <a href="http://www.ecologie.gouv.fr/-ONERC-.html">http://www.ecologie.gouv.fr/-ONERC-.html</a> (page consultée le 7 avril 2010).
CC_ONERC_impact_2009.fr	ONERC. 2009. « Changement climatique coûts des impacts et pistes d'adaptation ». <a href="http://www.ecologie.gouv.fr/Rapport-de-l-ONERC-Changement.html">http://www.ecologie.gouv.fr/Rapport-de-l-ONERC-Changement.html</a> (page consultée le 12 février 2010).
CC_ONERC_partie1_2009.fr	ONERC. 2009. « Changement climatique coûts des impacts et pistes d'adaptation ». <a href="http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html">http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html</a> (page consultée le 22 février 2010).
CC_ONERC_partie2_2009.fr	ONERC. 2009. « Changement climatique coûts des impacts et pistes d'adaptation ». <a href="http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html">http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html</a> (page consultée le 22 février 2010).
CC_ONERC_partie3_2009.fr	ONERC. 2009. « Changement climatique coûts des impacts et pistes d'adaptation ». <a href="http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html">http://www.ecologie.gouv.fr/Rapport-Evaluation-du-cout-des.html</a> (page consultée le 22 février 2010).
CC_ONERC_viticulture_2006.fr	ONERC. 2008. « Impact du changement climatique sur les activités vitivinicoles ». <a href="http://www.ecologie.gouv.fr/-ONERC-.html">http://www.ecologie.gouv.fr/-ONERC-.html</a> (page consultée le 7 avril 2010).
CC_ONU_desertification_2007.fr	Nations Unies. 2007. « Suivi des résultats du Sommet mondial pour le développement durable qui intéressent la Convention et préparation aux seizième et dix-septième sessions de la Commission du développement durable ». <a href="http://www.un.org/depts/dhl/dhlf/unbisnet/indexf.htm">Http://www.un.org/depts/dhl/dhlf/unbisnet/indexf.htm</a> (page consultée le 30 janvier 2008).
CC_Petit_anthropique_2001.fr	Petit, M. 2001. « L'humanité face à l'effet de serre additionnel qu'elle Crée ». <a href="http://www.sciencedirect.com/science">http://www.sciencedirect.com/science</a> (page consultée le 20 février 2010).
CC_Petit_effet-de-serre_2006.fr	Petit, M. 2006. « Maîtriser l'effet de serre ». <a href="http://documents.irevues.inist.fr/handle/2042/15256">http://documents.irevues.inist.fr/handle/2042/15256</a> (page consultée le 20 février 2010).
CC_Quebec_Nunavik_2006.fr	Commission des transports et de l'environnement. 2006. « Le réchauffement climatique et d'autres problématiques environnementales au Nunavik ». <a href="http://www.assnat.qc.ca/fra/37legislature2/commissions/Cte/rapport-climatNunavik.html">http://www.assnat.qc.ca/fra/37legislature2/commissions/Cte/rapport-climatNunavik.html</a> (page consultée le 11 février 2010).
CC_Richert_air_2007.fr	Richert, P. 2007. « Qualité de l'air et changement climatique : un même défi, une même urgence ».

	<a href="http://www.ladocumentationfrancaise.fr/rapports-publics/074000137/index.shtml">http://www.ladocumentationfrancaise.fr/rapports-publics/074000137/index.shtml</a> (page consultée le 20 février 2010).
CC_RNCA_Quebec_2008.fr	RNCA. 2008. « Vivre avec les changements climatiques au Canada : édition 2007 ». <a href="http://adaptation.nrcan.gc.ca/assess/2007/index_f.php">http://adaptation.nrcan.gc.ca/assess/2007/index_f.php</a> (page consultée le 24 mars 2010).
CC_Seguin_gaz_2002.fr	Seguin, B. 2002. « La recherche agronomique face aux gaz à effet de serre ». <i>Le Courrier de l'environnement de l'INRA</i> n°46, 169–185.
CC_Senat_ampleur_2002.fr	Deneux, M. 2002. « Rapport sur l'évaluation de l'ampleur des changements climatiques, de leurs causes et de leur impact prévisible sur la géographie de la France à l'horizon 2025, 2050 et 2100 ». <i>Paris, Assemblée nationale</i> . <a href="http://www.assemblee-nationale.fr/documents/index-general-oecst.asp">http://www.assemblee-nationale.fr/documents/index-general-oecst.asp</a> (page consultée le 17 avril 2008).
CC_Senat_emissions_1999.fr	Lepeltier, S. 1999. « Maîtriser les émissions de gaz à effet de serre : quels instruments économiques? ». <a href="http://www.senat.fr/rap/r98-346/r98-346.html">http://www.senat.fr/rap/r98-346/r98-346.html</a> (page consultée le 7 avril 2010).
CC_Senat_energie_2006.fr	Bataille, C. et C. Birraux. 2006. « Les nouvelles technologies de l'énergie et la séquestration du dioxyde de carbone : aspects scientifiques et techniques ». <i>Paris, Assemblée nationale</i> . <a href="http://www.assemblee-nationale.fr/documents/index-general-oecst.asp">http://www.assemblee-nationale.fr/documents/index-general-oecst.asp</a> (page consultée le 17 avril 2008).
CC_Senat_transition_2006.fr	Laffitte, P. et C. Saunier. 2006. « Changement climatique et transition énergétique : dépasser la crise ». <a href="http://www.assemblee-nationale.fr/12/rap-off/i3197-t1.asp">http://www.assemblee-nationale.fr/12/rap-off/i3197-t1.asp</a> (page consultée le 10 avril 2010).

## Annexe B : Liste des 343 termes anglais validés

Terme	Partie du discours	Fréquence	Score	Répartition	Critère d'identification
<i>abatement</i>	Nom	69	18,31	8	IATE
<i>absorb</i>	Verbe	92	22,59	20	L'Homme b)
<i>absorption</i>	Nom	43	15,33	12	L'Homme c)
<i>acidification</i>	Nom	31	12,95	9	IATE
<i>activity</i>	Nom	1118	55,82	36	L'Homme b)
<i>adapt</i>	Verbe	508	53,35	32	L'Homme b)
<i>adaptation</i>	Nom	2238	110,6	26	UNTERM
<i>adaptive</i>	Adjectif	536	54,74	18	L'Homme c)
<i>aerosol</i>	Nom	505	52,11	19	UNTERM
<i>afforestation</i>	Nom	61	18,33	10	UNTERM
<i>agricultural</i>	Adjectif	581	51,83	30	Termium
<i>agriculture</i>	Nom	549	50,44	32	IATE
<i>air</i>	Nom	586	21,79	32	Termium
<i>albedo</i>	Nom	27	12,05	8	Jacques et Le Treut
<i>animal</i>	Nom	187	8,94	21	IATE
<i>anthropogenic</i>	Adjectif	290	40,28	25	IPCC 2007
<i>aquatic</i>	Adjectif	105	22,09	15	Termium
<i>aquifer</i>	Nom	71	19,81	14	Termium
<i>assess</i>	Verbe	342	43,75	34	L'Homme b)
<i>assessment</i>	Nom	664	53,46	34	L'Homme c)
<i>atmosphere</i>	Nom	639	48,07	32	IPCC 2007
<i>atmospheric</i>	Adjectif	589	55,73	29	L'Homme c)
<i>biodiversity</i>	Nom	467	51,15	20	GDT
<i>bioenergy</i>	Nom	31	12,95	9	UNTERM
<i>biofuel</i>	Nom	44	15,51	12	UNTERM
<i>biological</i>	Adjectif	194	27,54	24	L'Homme b)
<i>biomass</i>	Nom	186	31,88	25	Jacques et Le Treut
<i>biome</i>	Nom	51	16,73	9	UNTERM
<i>biosphere</i>	Nom	85	19,06	18	Termium
<i>biota</i>	Nom	28	12,28	8	UNTERM
<i>capture</i>	Nom	182	29,83	8	Brodhag et Breuil
<i>capture</i>	Verbe	123	26,16	9	L'Homme b)
<i>carbon</i>	Nom	1690	93,18	37	Termium
<i>carbonate</i>	Nom	31	12,95	8	IATE
<i>change</i>	Verbe	976	73,99	37	L'Homme b)
<i>change</i>	Nom	8912	199,19	38	L'Homme c)
<i>chemical</i>	Nom	139	6,53	25	IATE
<i>chemistry</i>	Nom	41	7,72	16	IATE
<i>circumpolar</i>	Adjectif	26	11,52	8	Termium

<i>climate</i>	Nom	9090	223,63	38	Termium
<i>climatic</i>	Adjectif	339	42,57	24	UNTERM
<i>cloud</i>	Nom	163	20,74	16	IATE
<i>coal</i>	Nom	155	19,14	15	IATE
<i>coast</i>	Nom	241	19,95	22	IATE
<i>coastal</i>	Adjectif	799	64,36	26	IPCC 2007
<i>coastline</i>	Nom	82	17,77	13	GDT
<i>cold</i>	Adjectif	185	10,72	20	IATE
<i>compliance</i>	Nom	102	21,34	12	IPCC 2007
<i>concentration</i>	Nom	1054	69,93	32	Termium
<i>consequence</i>	Nom	346	33,06	33	L'Homme b)
<i>contamination</i>	Nom	53	10,35	10	IATE
<i>cool</i>	Verbe	162	30,05	23	L'Homme b)
<i>cooling</i>	Nom	60	17,06	23	L'Homme d)
<i>coral</i>	Nom	52	15,53	11	IATE
<i>crop</i>	Nom	631	52,39	27	Termium
<i>cropland</i>	Nom	45	15,47	10	L'Homme c)
<i>cultivation</i>	Nom	27	10,3	12	IATE
<i>current</i>	Nom	41	8,38	12	Termium
<i>cyclone</i>	Nom	80	19,02	11	UNTERM
<i>damage</i>	Nom	362	20,48	22	IATE
<i>decomposition</i>	Nom	32	12,4	10	IATE
<i>decrease</i>	Verbe	465	51,04	29	L'Homme b)
<i>deforestation</i>	Nom	178	30,72	24	IPCC 2007
<i>degradation</i>	Nom	240	35,28	26	GDT
<i>delta</i>	Nom	45	14,63	15	IATE
<i>deposition</i>	Nom	68	17,52	13	IATE
<i>desertification</i>	Nom	101	23,68	14	GDT
<i>dieback</i>	Nom	17	9,44	8	UNTERM
<i>disaster</i>	Nom	390	30,96	22	IATE
<i>discharge</i>	Nom	52	12,25	9	Termium
<i>disease</i>	Nom	432	23,27	19	IATE
<i>diversion</i>	Nom	28	7,92	8	Termium
<i>diversity</i>	Nom	130	22,88	21	Termium
<i>drought</i>	Nom	673	56,85	26	Termium
<i>ecological</i>	Adjectif	182	26,7	23	L'Homme c)
<i>ecosystem</i>	Nom	1132	79,24	32	Jacques et Le Treut
<i>effect</i>	Nom	1662	65,39	36	IATE
<i>electricity</i>	Nom	264	24,02	17	IATE
<i>emission</i>	Nom	4483	156,36	37	IATE
<i>emit</i>	Verbe	103	23,92	22	L'Homme b)
<i>energy</i>	Nom	1446	69,34	38	IPCC 2007
<i>environment</i>	Nom	390	20,97	33	IATE

<i>environmental</i>	Adjectif	1097	61,16	33	Termium
<i>environmentally</i>	Adverbe	55	11,73	16	L'Homme c)
<i>erode</i>	Verbe	31	12,95	13	L'Homme c)
<i>erosion</i>	Nom	352	42,27	19	Termium
<i>eruption</i>	Nom	44	13,33	10	IATE
<i>eutrophication</i>	Nom	31	12,68	9	GDT
<i>evaporation</i>	Nom	137	26,73	17	Termium
<i>evapotranspiration</i>	Nom	77	20,64	10	GDT
<i>event</i>	Nom	1112	39,28	31	Termium
<i>exposure</i>	Nom	147	17,05	20	IATE
<i>extinction</i>	Nom	67	15,83	10	IPCC 2007
<i>extreme</i>	Adjectif	731	58,54	28	Termium
<i>extreme</i>	Nom	236	33,62	28	L'Homme c)
<i>fauna</i>	Nom	24	9,51	9	IATE
<i>feedback</i>	Nom	204	31,63	22	IPCC 2007
<i>fire</i>	Nom	409	12,79	25	IATE
<i>fish</i>	Nom	353	17,44	18	IATE
<i>fishery 1</i>	Nom	403	38,73	18	Termium
<i>fishery 2</i>	Nom	403	38,73	18	IATE
<i>flood</i>	Verbe	83	21,44	8	L'Homme b)
<i>flood 1</i>	Nom	462	45,25	15	Termium
<i>flood 2</i>	Nom	462	45,25	15	Termium
<i>flooding</i>	Nom	353	42,49	19	IATE
<i>flow 1</i>	Nom	660	51,91	37	Termium
<i>flow 2</i>	Nom	660	51,91	37	Termium
<i>flux</i>	Nom	71	19,1	15	Jacques et Le Treut
<i>food</i>	Nom	581	18,01	30	IATE
<i>forcing</i>	Nom	690	62,2	17	Termium
<i>forecast</i>	Nom	47	5,86	10	IPCC 2007
<i>forecast</i>	Verbe	47	16,05	10	L'Homme c)
<i>forest</i>	Nom	1418	76,82	33	Termium
<i>forested</i>	Adjectif	33	11,95	10	L'Homme c)
<i>forestry</i>	Nom	318	38,69	24	IATE
<i>fragmentation</i>	Nom	33	12,62	12	Termium
<i>freshwater</i>	Nom	267	38,17	16	IATE
<i>frost</i>	Nom	46	13,72	11	IATE
<i>fuel</i>	Nom	694	41,82	35	IATE
<i>gas</i>	Nom	1746	79,08	37	GDT
<i>glacial</i>	Adjectif	92	21,81	12	L'Homme c)
<i>glacier</i>	Nom	330	41,89	20	IPCC 2007
<i>global</i>	Adjectif	1728	91,38	37	GDT
<i>globally</i>	Adverbe	136	26,62	25	GDT
<i>grain</i>	Nom	73	8,85	8	IATE



<i>grassland</i>	Nom	133	25,82	13	IATE
<i>grazing</i>	Nom	46	14,25	10	IATE
<i>groundwater</i>	Nom	202	33,48	16	IATE
<i>habitat</i>	Nom	316	37,97	21	GDT
<i>hail</i>	Nom	22	5,62	8	IATE
<i>halocarbon</i>	Nom	62	18,49	8	IPCC 2007
<i>harmful</i>	Adjectif	34	4,62	12	GDT
<i>harvest</i>	Nom	91	18,27	11	IATE
<i>harvest</i>	Verbe	35	13,79	12	L'Homme c)
<i>hazard</i>	Nom	178	22,29	18	IATE
<i>heat</i>	Nom	458	28,53	27	Termium
<i>human</i>	Nom	71	7,54	8	L'Homme b)
<i>human</i>	Adjectif	935	43,37	36	L'Homme c)
<i>humid</i>	Adjectif	25	8,13	12	GDT
<i>humidity</i>	Nom	38	9,66	15	GDT
<i>hurricane</i>	Nom	95	22,05	10	IATE
<i>hydrocarbon</i>	Nom	77	18,31	9	Termium
<i>hydrological</i>	Adjectif	203	33,67	21	L'Homme c)
<i>hydrology</i>	Nom	67	19,23	11	Termium
<i>ice</i>	Nom	1018	42,54	22	Termium
<i>impact</i>	Verbe	215	34,65	29	L'Homme b)
<i>impact</i>	Nom	3276	125,24	38	IPCC 2007
<i>increase</i>	Nom	1788	69,06	37	L'Homme b)
<i>increase</i>	Verbe	3369	137,55	38	L'Homme d)
<i>infestation</i>	Nom	33	12,17	8	Termium
<i>infrastructure</i>	Nom	963	72,42	32	IPCC 2007
<i>insect</i>	Nom	128	18,09	12	IATE
<i>intense</i>	Adjectif	166	17,58	17	L'Homme c)
<i>intensity</i>	Nom	293	33,67	28	IATE
<i>interaction</i>	Nom	193	29,89	28	IATE
<i>inundation</i>	Nom	32	13,16	9	Termium
<i>irrigation</i>	Nom	291	39,2	21	IATE
<i>island</i>	Nom	241	14,17	20	IATE
<i>lake</i>	Nom	377	34,82	18	IATE
<i>land</i>	Nom	1263	60,48	35	IATE
<i>landfill</i>	Nom	51	12,91	12	IATE
<i>landscape</i>	Nom	126	18,74	10	IATE
<i>landslide</i>	Nom	56	13,96	11	IPCC 2007
<i>latitude</i>	Nom	212	33,18	20	IATE
<i>latitudinal</i>	Adjectif	15	8,82	8	L'Homme c)
<i>lifetime</i>	Nom	84	8,35	12	IATE
<i>livestock</i>	Nom	128	23,36	18	IATE
<i>locally</i>	Adverbe	56	11,43	16	GDT

<i>lowland</i>	Nom	16	8,78	8	Termium
<i>magnitude</i>	Nom	353	41,97	28	Termium
<i>mangrove</i>	Nom	86	21,5	11	IATE
<i>manure</i>	Nom	36	12,02	11	IATE
<i>mapping</i>	Nom	18	7,31	8	IATE
<i>marine</i>	Adjectif	310	36,84	19	Termium
<i>marsh</i>	Nom	62	16,26	10	IATE
<i>measure</i>	Nom	779	43,11	35	IATE
<i>melt</i>	Verbe	117	25,51	18	L'Homme b)
<i>melting</i>	Nom	103	22,62	20	IATE
<i>meteorological</i>	Adjectif	23	9,02	19	Termium
<i>methane</i>	Nom	179	30,49	28	IPCC 2007
<i>migrate</i>	Verbe	70	19,66	16	L'Homme c)
<i>migration</i>	Nom	196	29,63	18	IATE
<i>mining</i>	Nom	87	10,29	14	IATE
<i>mitigate</i>	Verbe	63	18,64	22	UNTERM
<i>mitigation</i>	Nom	844	68,75	28	Termium
<i>model</i>	Verbe	228	35,69	20	L'Homme b)
<i>model</i>	Nom	1740	82,93	29	Jacques et Le Treut
<i>modeling</i>	Nom	46	15,87	9	IATE
<i>moist</i>	Adjectif	19	3,78	9	IATE
<i>moisture</i>	Nom	244	30,76	18	IATE
<i>monitor</i>	Verbe	150	28,91	18	IATE
<i>monitoring</i>	Nom	193	30,08	28	Termium
<i>monsoon</i>	Nom	39	14,11	9	IPCC 2007
<i>mortality</i>	Nom	131	24,67	15	IPCC 2007
<i>mountain</i>	Nom	188	17,34	19	IATE
<i>mountainous</i>	Adjectif	21	7,15	8	L'Homme c)
<i>natural</i>	Adjectif	1140	63,33	33	L'Homme b)
<i>negative</i>	Adjectif	421	37,41	31	L'Homme b)
<i>negatively</i>	Adverbe	61	16,72	11	L'Homme c)
<i>nitrogen</i>	Nom	141	24,91	22	Termium
<i>nutrient</i>	Nom	74	15,03	18	IATE
<i>ocean</i>	Nom	848	62,26	27	IATE
<i>oceanic</i>	Adjectif	59	16,9	16	L'Homme c)
<i>offshore</i>	Adjectif	34	8,31	8	Termium
<i>organic</i>	Adjectif	120	18,16	21	Termium
<i>organism</i>	Nom	68	14,17	14	Termium
<i>oscillation</i>	Nom	28	12	12	UNTERM
<i>outbreak</i>	Nom	157	24,33	12	IATE
<i>oxidation</i>	Nom	19	8,29	8	IATE
<i>ozone</i>	Nom	436	49,42	22	IATE
<i>pasture</i>	Nom	45	12,73	12	IATE

<i>pattern</i>	Nom	507	40,42	31	L'Homme b)
<i>peat</i>	Nom	18	5,65	8	IPCC 2007
<i>permafrost</i>	Nom	352	44,39	15	IATE
<i>pest</i>	Nom	154	25,52	15	IATE
<i>photosynthesis</i>	Nom	40	14,3	11	IPCC 1995
<i>phytoplankton</i>	Nom	17	8,76	8	IPCC 2007
<i>plant</i>	Nom	717	32,68	25	Termium
<i>polar</i>	Adjectif	127	21,32	16	Termium
<i>pollutant</i>	Nom	113	20,85	18	IATE
<i>pollution</i>	Nom	230	18,85	27	IATE
<i>population</i>	Nom	984	52,36	32	Termium
<i>prairie</i>	Nom	60	9,11	9	Termium
<i>precipitation</i>	Nom	1131	79,16	23	Termium
<i>precursor</i>	Nom	42	13,47	13	Termium
<i>preservation</i>	Nom	29	4,62	14	Termium
<i>preserve</i>	Verbe	26	11,81	12	L'Homme c)
<i>project</i>	Verbe	974	73,92	33	L'Homme b)
<i>projection</i>	Nom	539	51,18	30	Brodhag et Breuil
<i>protection</i>	Nom	253	16,69	30	IATE
<i>radiation</i>	Nom	130	15,45	19	Termium
<i>radiative</i>	Adjectif	384	46,37	16	L'Homme c)
<i>rain</i>	Nom	164	10,15	22	IATE
<i>rainfall 1</i>	Nom	352	43,03	23	Termium
<i>rainfall 2</i>	Nom	352	43,03	23	Termium
<i>recharge</i>	Nom	70	19,66	9	IATE
<i>recovery</i>	Nom	131	17,38	19	IATE
<i>recycling</i>	Nom	59	9,8	12	IATE
<i>reduce</i>	Verbe	1807	100,71	36	L'Homme d)
<i>reduction</i>	Nom	1195	67,58	37	Termium
<i>reflect</i>	Verbe	285	39,92	10	L'Homme b)
<i>reforestation</i>	Nom	86	21,83	13	IPCC 2007
<i>regeneration</i>	Nom	43	14,66	9	IATE
<i>regime</i>	Nom	217	21,23	19	IATE
<i>region</i>	Nom	1863	82,09	35	Termium
<i>regional</i>	Adjectif	794	49,28	36	IATE
<i>regionally</i>	Adverbe	60	17,79	13	IATE
<i>release</i>	Nom	107	5,26	23	IATE
<i>release</i>	Verbe	127	26,58	27	L'Homme c)
<i>renewable</i>	Adjectif	116	24,04	20	UNTERM
<i>reservoir</i>	Nom	177	27,65	16	Brodhag et Breuil
<i>resilience</i>	Nom	176	30,76	15	IPCC 2007
<i>resilient</i>	Adjectif	38	13,46	11	UNTERM
<i>resource</i>	Nom	1190	70,92	34	IATE

<i>retention</i>	Nom	16	5,9	8	IATE
<i>rise</i>	Verbe	367	45,33	25	L'Homme b)
<i>rise</i>	Nom	669	47,88	34	L'Homme c)
<i>risk</i>	Nom	1326	66,37	33	IATE
<i>river</i>	Nom	452	34,87	22	IATE
<i>road</i>	Nom	326	10,98	23	IATE
<i>runoff</i>	Nom	293	38,27	15	IPCC 2007
<i>salinity</i>	Nom	60	17,98	14	Termium
<i>scarcity</i>	Nom	64	17,87	16	IATE
<i>scenario</i>	Nom	1602	92,19	33	IPCC 2007
<i>sea</i>	Nom	972	61,03	30	Termium
<i>seasonal</i>	Adjectif	259	33,72	20	L'Homme b)
<i>seasonality</i>	Nom	33	13,11	8	L'Homme c)
<i>seasonally</i>	Adverbe	28	7,09	10	L'Homme c)
<i>seawater</i>	Nom	23	9,46	12	Termium
<i>sediment</i>	Nom	107	22,98	16	IATE
<i>sensitivity</i>	Nom	377	42,76	22	IPCC 2007
<i>sequester</i>	Verbe	50	16,56	8	L'Homme c)
<i>sequestration</i>	Nom	149	28,81	15	L'Homme d)
<i>shoreline</i>	Nom	81	15,05	11	Termium
<i>simulate</i>	Verbe	87	21,96	17	L'Homme b)
<i>simulated</i>	Adjectif	76	17,24	11	L'Homme c)
<i>simulation</i>	Nom	171	29,75	20	IATE
<i>sink</i>	Nom	147	25,06	27	Jacques et Le Treut
<i>snow</i>	Nom	340	28,46	20	IATE
<i>snowfall</i>	Nom	54	14	10	IATE
<i>snowmelt 1</i>	Nom	28	12,28	8	IATE
<i>snowmelt 2</i>	Nom	28	12,28	8	IATE
<i>soil</i>	Nom	673	52,31	30	Termium
<i>solar</i>	Adjectif	179	23,64	20	Jacques et Le Treut
<i>source</i>	Nom	1085	46,01	29	IPCC 2007
<i>specie</i>	Nom	982	69,42	26	UNTERM
<i>stabilization</i>	Nom	350	43,37	16	IPCC 2007
<i>stabilize</i>	Verbe	93	22,71	19	L'Homme b)
<i>storage</i>	Nom	505	46,42	24	IATE
<i>store</i>	Verbe	110	24,72	20	L'Homme b)
<i>storm</i>	Nom	521	44,23	21	IATE
<i>stratosphere</i>	Nom	56	16,77	9	UNTERM
<i>stratospheric</i>	Adjectif	97	22,74	15	GDT
<i>stream</i>	Nom	143	18,57	18	IATE
<i>streamflow</i>	Nom	67	19,23	8	IPCC 2007
<i>stress</i>	Nom	440	38,01	22	UNTERM
<i>subsidence</i>	Nom	28	12,28	10	Termium

<i>sulphur</i>	Nom	59	12,96	14	GDT
<i>sustainability</i>	Nom	140	27,92	24	IATE
<i>sustainable</i>	Adjectif	475	49,87	30	Termium
<i>temperate</i>	Adjectif	104	23,16	12	Termium
<i>temperature</i>	Nom	2185	100,48	32	Termium
<i>terrestrial</i>	Adjectif	218	34,69	21	Termium
<i>thaw</i>	Nom	74	16,81	12	Termium
<i>thermohaline</i>	Adjectif	39	14,58	9	Termium
<i>thunderstorm</i>	Nom	38	11,12	8	IATE
<i>tidal</i>	Adjectif	37	10,75	9	L'Homme c)
<i>tide</i>	Nom	27	3,76	9	IATE
<i>tornado</i>	Nom	31	8,8	8	IATE
<i>tree</i>	Nom	290	19,65	22	IATE
<i>troposphere</i>	Nom	48	16,22	8	IPCC 1995
<i>tropospheric</i>	Adjectif	75	20,36	11	L'Homme c)
<i>tundra</i>	Nom	62	16,72	10	IPCC 2007
<i>uncertainty</i>	Nom	725	60,08	29	IPCC 2007
<i>unsustainable</i>	Adjectif	23	10,77	9	Termium
<i>upstream</i>	Adjectif	38	13,25	9	IATE
<i>uptake</i>	Nom	111	24,13	16	UNTERM
<i>urbanization</i>	Nom	36	13,99	11	IPCC 2007
<i>variability</i>	Nom	646	60,12	29	L'Homme c)
<i>variable</i>	Adjectif	72	17,7	8	L'Homme c)
<i>variable</i>	Nom	133	26,06	23	L'Homme c)
<i>variation</i>	Nom	308	35,31	27	L'Homme b)
<i>vary</i>	Verbe	418	48,38	29	L'Homme b)
<i>vegetation</i>	Nom	236	34,59	24	IATE
<i>volcanic</i>	Adjectif	66	17,85	15	IATE
<i>vulnerability</i>	Nom	753	64,36	20	Termium
<i>vulnerable</i>	Adjectif	482	45,18	18	L'Homme c)
<i>warm</i>	Adjectif	188	14,48	18	L'Homme c)
<i>warm</i>	Verbe	173	31,07	30	L'Homme b)
<i>warming</i>	Nom	873	66,59	28	UNTERM
<i>waste</i>	Nom	227	13,23	24	IATE
<i>water</i>	Nom	3757	112	33	IATE
<i>watershed</i>	Nom	68	18,32	11	Termium
<i>weather</i>	Nom	541	36,12	26	Termium
<i>weed</i>	Nom	28	4,93	10	IATE
<i>wetland</i>	Nom	241	36,11	17	Termium
<i>wildfire</i>	Nom	45	15,25	9	UNTERM
<i>wildlife</i>	Nom	102	12,87	14	Termium
<i>wind</i>	Nom	256	18,82	23	IATE
<i>wood</i>	Nom	165	10,92	19	Termium

<i>woodland</i>	Nom	23	6,99	10	IATE
<i>world</i>	Nom	586	5,59	35	IATE
<i>worldwide</i>	Adverbe	60	10,11	16	L'Homme c)

## Annexe C : Liste des ressources terminologiques et autres dictionnaires

### A Ressources terminologiques utilisées pour la validation des termes anglais

Brodhag, C. et F. Breuil. 2009. *Glossaire du climat : traduction anglais/français et définitions*. Organisation internationale de la francophonie.

<http://www.iepf.org/ressources/ressources-pub-desc.php?id=344> (page consultée le 21 juin 2010).

GDT. *Grand dictionnaire terminologique de l'Office de la langue français*.

<http://www.olf.gouv.qc.ca/ressources/gdt.html> (page consultée le 21 décembre 2009).

IATE. *Inter Active Terminology for Europe*.

<http://iate.europa.eu/iatediff/SearchByQueryLoad.do?method=load> (page consultée le 21 juin 2010).

IPCC. 1995. *Glossary by the Intergovernmental Panel on Climate Change*.

[http://www.ipcc.ch/publications\\_and\\_data/publications\\_and\\_data\\_glossary.htm](http://www.ipcc.ch/publications_and_data/publications_and_data_glossary.htm)  
(page consultées le 21 juin 2010).

IPCC. 2007. « Glossary ». Dans *Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. 76–89.

[http://www.ipcc.ch/publications\\_and\\_data/ar4/syr/en/annexes.html](http://www.ipcc.ch/publications_and_data/ar4/syr/en/annexes.html) (page consultée le 21 juin 2010).

Jacques, G. et H. Le Treut. 2004. *Le changement climatique*. Paris : Éditions Unesco.

Termium Plus. *La base de données terminologiques et linguistiques du gouvernement du Canada*. <http://www.termiumplus.com/> (page consultée le 21 juin 2010).

UNTERM. *United Nations Multilingual Terminology Database*. <http://unterm.un.org/> (page consultée le 21 juin 2010).

**B Autres ressources terminologiques consultées pour trouver des indices d'équivalence et/ou pour valider des équivalents (en plus des ressources terminologiques précédentes)**

- Bioret, F., R. Estève et A. Sturbois. 2009. *Dictionnaire de la protection de la nature*. Rennes : Presses Universitaire de Rennes.
- Castany, G. et J. Margat. 1977. *Dictionnaire français d'hydrogéologie*. Orléans : Bureau de recherches géologiques et minières, Service géologique national.
- Commission générale de terminologie et de néologie. *FranceTerme*.  
<http://franceterme.culture.fr/FranceTerme/recherche.html> (page consultée le 5 août 2011).
- Comité National Français des Sciences Hydrologiques (C.N.F.S.H.). 2000. *Dictinnaire français d'hydrologie*.  
<http://webworld.unesco.org/water/ihp/db/glossary/glu/indexdic.htm> (page consultée le 11 juin 2011).
- Conseil international de la langue française. 1978. *Vocabulaire de l'hydrologie et de la météorologie (VHM)*. Paris : La maison du dictionnaire.
- Côté, M. 2003. *Dictionnaire de la foresterie – Dictionary of forestry*. Sainte-Foy (Québec) : Ordre des ingénieurs du Québec.
- Dutuit, P. et R. Gorenflot. 2008. *Glossaire pour le développement durable –Des mots pour les maux de la planète*. Paris : Éditions des archives contemporaines.
- Groupe d'experts OMM/Unesco sur la terminologie. 1992. *International glossary of hydrology = Glossaire international d'hydrologie*, 2<sup>e</sup> édition. Paris : Jouve.
- Moureau, M. et G. Brace. 2000. *Dictionnaire des Sciences de la Terre*. Paris : Éditions Technip.
- Reeves, H. et F. Lenoir. 2005. *Mal de Terre*. Paris : Éditions du Seuil.
- Villeneuve, C. et F. Richard. 2005. *Vivre les changements climatiques – Quoi de neuf?* Sainte-Foy (Québec) : Les Éditions MultiMondes.
- Water Treatment Solutions – Lenntech. *Glossaire du changement climatique*.  
<http://www.lenntech.fr/effet-de-serre/glossaire-changement-climatique.htm#ixzz1MW2KOZD2> (page consultée le 2 juillet 2011).



## **C Dictionnaires de langues générales et autres ouvrages consultés pour trouver des indices d'équivalence et attester des faits de langue**

Antidote. *Logiciel d'aide à la rédaction du français – correcteur, dictionnaires, guides.*

<http://www.druide.com/>

CRISCO (Centre de Recherche Inter-langues sur la Signification en Contexte). *Le dictionnaire des synonymes.* <http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi> (page consultée le 6 novembre 2011).

Google. Moteur de recherche. <http://www.google.ca/>

Le nouveau petit Robert. 2002. *Dictionnaire de la langue française.* Paris : Dictionnaire le Robert.

OneLook Dictionary Search. <http://www.onelook.com/> (page consultée le 6 juillet 2011).

Le grand Robert & Collins électronique. 2003. *Dictionnaire anglais français et français anglais.* Version 1.0. VUEF, Paris.

Le trésor de la langue française informatisé.

<http://atilf.atilf.fr/dendien/scripts/tlfiv4/showps.exe?p=combi.htm;java=no>; (page consultée le 29 juillet 2011).

Meertens, R. 2004. *Guide anglais, français de la traduction.* Paris : Chiron.

Van Roey, J., S. Granger et H. Swallow. 1991. *Dictionnaire des faux amis – Dictionary of faux amis.* Paris, Bruxelles : De Boeck & Larcier.

Wikipedia. *L'encyclopédie libre.*

[http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil\\_principal](http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal) (page consultée le 10 novembre 2011).

## Annexe D : Liste des équivalents complexes du corpus parallèle et du corpus comparable

- les équivalents typiques sont identifiés par l'abréviation ÉqT
- les équivalents non typiques par ÉqNT

Tableau D.1. Équivalents complexes du corpus parallèle

Terme anglais	Équivalent complexe du CP	ÉqT/ÉqNT	Équivalents typique
<i>biodiversity</i> -NOM	<i>diversité biologique</i>	ÉqT	<i>biodiversité</i> <i>diversité biologique</i>
<i>chemical</i> -NOM	<i>produit chimique</i>	ÉqT	<i>substance chimique</i> <i>produit chimique</i>
<i>chemical</i> -NOM	<i>substance chimique</i>	ÉqT	<i>produit chimique</i>
<i>cropland</i> -NOM	<i>terre cultivée</i>	ÉqT	s.o.
<i>cropland</i> -NOM	<i>terre agricole</i>	ÉqNT	<i>terre cultivée</i>
<i>cropland</i> -NOM	<i>terre arable</i>	ÉqNT	<i>terre cultivée</i>
<i>drought</i> -NOM	<i>période de sécheresse</i>	ÉqNT	<i>sécheresse</i>
<i>freshwater</i> -NOM	<i>eau douce</i>	ÉqT	s.o.
<i>groundwater</i> -NOM	<i>eau souterraine</i>	ÉqT	s.o.
<i>groundwater</i> -NOM	<i>nappe phréatique</i>	ÉqNT	<i>eau souterraine</i>
<i>groundwater</i> -NOM	<i>nappe souterraine</i>	ÉqNT	<i>eau souterraine</i>
<i>halocarbon</i> -NOM	<i>hydrocarbure halogéné</i>	ÉqT	<i>halocarbure</i> <i>hydrocarbure halogéné</i>
<i>halocarbon</i> -NOM	<i>gaz halocarboné</i>	ÉqNT	<i>halocarbure</i>
<i>landslide</i> -NOM	<i>glissement de terrain</i>	ÉqT	s.o.
<i>lifetime</i> -NOM	<i>durée de vie</i>	ÉqT	s.o.
<i>lowland</i> -NOM	<i>basse terre</i>	ÉqT	s.o.
<i>lowland</i> -NOM	<i>terre basse</i>	ÉqNT	<i>basse terre</i>
<i>magnitude</i> -NOM	<i>ordre de grandeur</i>	ÉqNT	<i>ampleur</i>
<i>nutrient</i> -NOM	<i>élément nutritif</i>	ÉqT	s.o.
<i>nutrient</i> -NOM	<i>substance nutritive</i>	ÉqT	<i>élément nutritif</i> <i>substance nutritive</i>
<i>offshore</i> -ADJ	<i>en mer</i>	ÉqT	<i>au/du large</i> <i>en mer</i>
<i>offshore</i> -ADJ	<i>au/du large</i>	ÉqT	<i>en mer</i> <i>au/du large</i>
<i>pest</i> -NOM	<i>animal nuisible</i>	ÉqT	<i>organisme nuisible</i> <i>animal nuisible, etc.</i>
<i>project</i> -VER	<i>à venir</i>	ÉqNT	<i>prévoir</i>
<i>resilience</i> -NOM	<i>capacité d'adaptation</i>	ÉqNT	<i>résilience</i> <i>capacité de récupération</i>
<i>river</i> -NOM	<i>cours d'eau</i>	ÉqT	<i>rivière</i> <i>fleuve</i>

<i>seawater</i> -NOM	<i>eau de mer</i>	ÉqT	s.o.
<i>snowfall</i> -NOM	<i>chute de neige</i>	ÉqT	s.o.
<i>snowmelt</i> -NOM	<i>eau de fonte</i>	ÉqT	s.o.
<i>snowmelt</i> -NOM	<i>fonte de la/des/de neige</i>	ÉqT	s.o.
<i>stream</i> -NOM	<i>cours d'eau</i>	ÉqNT	<i>ruisseau</i>
<i>subsidence</i> -NOM	<i>affaissement de terrain</i>	ÉqT	<i>subsidence</i> <i>affaissement de sol</i>
<i>unsustainable</i> -ADJ	<i>non durable</i>	ÉqT	s.o.
<i>watershed</i> -NOM	<i>bassin versant</i>	ÉqT	s.o.
<i>watershed</i> -NOM	<i>bassin hydrographique</i>	ÉqNT	<i>bassin versant</i>
<i>weather</i> -NOM	<i>condition atmosphérique</i>	ÉqNT	<i>temps</i>
<i>weed</i> -NOM	<i>mauvaise herbe</i>	ÉqT	s.o.
<i>wetland</i> -NOM	<i>milieu humide</i>	ÉqT	<i>zone/milieu/terre</i> <i>humide</i>
<i>wetland</i> -NOM	<i>terre humide</i>	ÉqT	<i>zone/milieu/terre humide</i>
<i>wetland</i> -NOM	<i>zone humide</i>	ÉqT	<i>zone/milieu/terre humide</i>
<i>wildfire</i> -NOM	<i>feu de friche</i>	ÉqT	<i>feu/incendie de forêt</i> <i>feu de friche</i>
<i>wildfire</i> -NOM	<i>feu de forêt</i>	ÉqT	<i>feu/incendie de forêt</i>
<i>wildfire</i> -NOM	<i>incendie de forêt</i>	ÉqT	<i>feu/incendie de forêt</i>
<i>wildlife</i> -NOM	<i>espèce sauvage</i>	ÉqNT	<i>faune</i> <i>animal sauvage</i>
<i>wildlife</i> -NOM	<i>faune sauvage</i>	ÉqNT	<i>faune</i> <i>animal sauvage</i>
<i>wildlife</i> -NOM	<i>animal sauvage</i>	ÉqT	<i>faune</i> <i>animal sauvage</i>
<i>woodland</i> -NOM	<i>région boisée</i>	ÉqT	s.o.

Tableau D.2. Équivalents complexes du corpus comparable

<b>Terme anglais</b>	<b>Équivalent complexe du CC</b>	<b>ÉqT/ÉqNT</b>	<b>Équivalents typiques</b>
<i>biodiversity</i> -NOM	<i>diversité biologique</i>	ÉqT	<i>biodiversité</i> <i>diversité biologique</i>
<i>chemical</i> -NOM	<i>produit chimique</i>	ÉqT	<i>substance chimique</i> <i>produit chimique</i>
<i>chemical</i> -NOM	<i>substance chimique</i>	ÉqT	<i>produit chimique</i>
<i>coastline</i> -NOM	<i>trait de côte</i>	ÉqNT	<i>littoral</i>
<i>cropland</i> -NOM	<i>terre agricole</i>	ÉqNT	<i>terre cultivée</i>
<i>cropland</i> -NOM	<i>terre arable</i>	ÉqNT	<i>terre cultivée</i>
<i>cropland</i> -NOM	<i>terre cultivée</i>	ÉqT	s.o.
<i>drought</i> -NOM	<i>période de sécheresse</i>	ÉqNT	<i>sécheresse</i>
<i>freshwater</i> -NOM	<i>eau douce</i>	ÉqT	s.o.
<i>groundwater</i> -NOM	<i>nappe phréatique</i>	ÉqNT	<i>eau souterraine</i>
<i>groundwater</i> -NOM	<i>nappe souterraine</i>	ÉqNT	<i>eau souterraine</i>
<i>groundwater</i> -NOM	<i>eau souterraine</i>	ÉqT	s.o.
<i>halocarbon</i> -NOM	<i>hydrocarbure halogéné</i>	ÉqT	<i>halocarbure</i>

			<i>hydrocarbure halogéné</i>
<i>landfill-NOM</i>	<i>centre d'enfouissement</i>	ÉqT	<i>décharge</i>
<i>landslide-NOM</i>	<i>glissement de terrain</i>	ÉqT	s.o.
<i>lifetime-NOM</i>	<i>temps de résidence</i>	ÉqNT	<i>durée de vie</i>
<i>lifetime-NOM</i>	<i>durée de vie</i>	ÉqT	s.o.
<i>lowland-NOM</i>	<i>basse terre</i>	ÉqT	s.o.
<i>magnitude-NOM</i>	<i>ordre de grandeur</i>	ÉqNT	<i>ampleur</i>
<i>nutrient-NOM</i>	<i>élément nutritif</i>	ÉqT	s.o.
<i>offshore-ADJ</i>	<i>en mer</i>	ÉqT	<i>au/du large</i> <i>en mer</i>
<i>offshore-ADJ</i>	<i>au/du large</i>	ÉqT	<i>en mer</i> <i>au/du large</i>
<i>project-VER</i>	<i>à venir</i>	ÉqNT	<i>prévoir</i>
<i>region-NOM</i>	<i>aire géographique</i>	ÉqNT	<i>région</i>
<i>resilience-NOM</i>	<i>capacité d'adaptation</i>	ÉqNT	<i>résilience</i> <i>capacité de récupération</i>
<i>river-NOM</i>	<i>cours d'eau</i>	ÉqT	<i>rivière</i> <i>fleuve</i>
<i>seawater-NOM</i>	<i>eau de mer</i>	ÉqT	s.o.
<i>shoreline-NOM</i>	<i>trait de côte</i>	ÉqNT	<i>littoral</i>
<i>snowfall-NOM</i>	<i>chute de neige</i>	ÉqT	s.o.
<i>snowmelt-NOM</i>	<i>fonte nivale</i>	ÉqNT	<i>fonte de la/des/de neige</i>
<i>snowmelt-NOM</i>	<i>eau de fonte</i>	ÉqT	s.o.
<i>snowmelt-NOM</i>	<i>fonte de la/des/de neige</i>	ÉqT	s.o.
<i>stream-NOM</i>	<i>cours d'eau</i>	ÉqNT	<i>ruisseau</i>
<i>watershed-NOM</i>	<i>bassin hydrographique</i>	ÉqNT	<i>bassin versant</i>
<i>watershed-NOM</i>	<i>bassin versant</i>	ÉqT	s.o.
<i>weather-NOM</i>	<i>condition atmosphérique</i>	ÉqNT	<i>temps</i>
<i>weed-NOM</i>	<i>mauvaise herbe</i>	ÉqT	s.o.
<i>wetland-NOM</i>	<i>terre humide</i>	ÉqT	<i>zone/milieu/terre humide</i>
<i>wetland-NOM</i>	<i>zone humide</i>	ÉqT	<i>zone/milieu/terre humide</i>
<i>wildfire-NOM</i>	<i>feu de forêt</i>	ÉqT	<i>feu/incendie de forêt</i>
<i>wildfire-NOM</i>	<i>incendie de forêt</i>	ÉqT	<i>feu/incendie de forêt</i>
<i>wildlife-NOM</i>	<i>faune sauvage</i>	ÉqNT	<i>faune</i> <i>animal sauvage</i>
<i>woodland-NOM</i>	<i>surface boisée</i>	ÉqNT	<i>terre boisée</i>

## Annexe E : Liste des 272 termes anglais ayant des équivalents identiques dans les deux corpus

Les termes *tidal*-NOM et *upstream*-NOM associés à des équivalents appartenant à une autre partie du discours sont identifiés en gras et dans des cases en gris clair.

Terme anglais	Partie du discours	Équivalents français + partie du discours	ÉqT/ÉqNT	#
<i>abatement</i>	NOM	<i>réduction</i> -NOM	ÉqT	1
<i>absorption</i>	NOM	<i>absorption</i> -NOM	ÉqT	1
<i>acidification</i>	NOM	<i>acidification</i> -NOM	ÉqT	1
<i>activity</i>	NOM	<i>activité</i> -NOM	ÉqT	1
<i>afforestation</i>	NOM	<i>boisement</i> -NOM	ÉqT	1
<i>albedo</i>	NOM	<i>albédo</i> -NOM	ÉqT	1
<i>aquatic</i>	ADJ	<i>aquatique</i> -ADJ	ÉqT	1
<i>biological</i>	ADJ	<i>biologique</i> -ADJ	ÉqT	1
<i>biomass</i>	NOM	<i>biomasse</i> -NOM	ÉqT	1
<i>biosphere</i>	NOM	<i>biosphère</i> -NOM	ÉqT	1
<i>carbon</i>	NOM	<i>carbone</i> -NOM	ÉqT	1
<i>carbonate</i>	NOM	<i>carbonate</i> -NOM	ÉqT	1
<i>circumpolar</i>	ADJ	<i>circumpolaire</i> -ADJ	ÉqT	1
<i>cold</i>	ADJ	<i>froid</i> -ADJ	ÉqT	1
<i>current</i>	NOM	<i>courant</i> -NOM	ÉqT	1
<i>decomposition</i>	NOM	<i>décomposition</i> -NOM	ÉqT	1
<i>deposition</i>	NOM	<i>dépôt</i> -NOM	ÉqT	1
<i>desertification</i>	NOM	<i>désertification</i> -NOM	ÉqT	1
<i>dieback</i>	NOM	<i>dépérissement</i> -NOM	ÉqT	1
<i>discharge</i>	NOM	<i>débit</i> -NOM	ÉqT	1
<i>diversity</i>	NOM	<i>diversité</i> -NOM	ÉqT	1
<i>emission</i>	NOM	<i>émission</i> -NOM	ÉqT	1
<i>erosion</i>	NOM	<i>érosion</i> -NOM	ÉqT	1
<i>eruption</i>	NOM	<i>éruption</i> -NOM	ÉqT	1
<i>eutrophication</i>	NOM	<i>eutrophisation</i> -NOM	ÉqT	1
<i>evaporation</i>	NOM	<i>évaporation</i> -NOM	ÉqT	1
<i>evapotranspiration</i>	NOM	<i>évapotranspiration</i> -NOM	ÉqT	1
<i>fauna</i>	NOM	<i>faune</i> -NOM	ÉqT	1
<i>feedback</i>	NOM	<i>rétroaction</i> -NOM	ÉqT	1
<i>fishery1</i>	NOM	<i>pêche</i> -NOM	ÉqT	1
<i>flood2</i>	NOM	<i>crue</i> -NOM	ÉqT	1
<i>flow1</i>	NOM	<i>débit</i> -NOM	ÉqT	1
<i>flow2</i>	NOM	<i>écoulement</i> -NOM	ÉqT	1

<i>flux</i>	NOM	<i>flux</i> -NOM	ÉqT	1
<i>forcing</i>	NOM	<i>forçage</i> -NOM	ÉqT	1
<i>forecast</i>	NOM	<i>prévision</i> -NOM	ÉqT	1
<i>forecast</i>	VER	<i>prévoir</i> -VER	ÉqT	1
<i>grazing</i>	NOM	<i>pâturage</i> -NOM	ÉqT	1
<i>habitat</i>	NOM	<i>habitat</i> -NOM	ÉqT	1
<i>hail</i>	NOM	<i>grêle</i> -NOM	ÉqT	1
<i>humid</i>	ADJ	<i>humide</i> -ADJ	ÉqT	1
<i>humidity</i>	NOM	<i>humidité</i> -NOM	ÉqT	1
<i>hurricane</i>	NOM	<i>ouragan</i> -NOM	ÉqT	1
<i>hydrocarbon</i>	NOM	<i>hydrocarbure</i> -NOM	ÉqT	1
<i>infrastructure</i>	NOM	<i>infrastructure</i> -NOM	ÉqT	1
<i>insect</i>	NOM	<i>insecte</i> -NOM	ÉqT	1
<i>interaction</i>	NOM	<i>interaction</i> -NOM	ÉqT	1
<i>inundation</i>	NOM	<i>inondation</i> -NOM	ÉqT	1
<i>irrigation</i>	NOM	<i>irrigation</i> -NOM	ÉqT	1
<i>landscape</i>	NOM	<i>paysage</i> -NOM	ÉqT	1
<i>latitude</i>	NOM	<i>latitude</i> -NOM	ÉqT	1
<i>mangrove</i>	NOM	<i>mangrove</i> -NOM	ÉqT	1
<i>manure</i>	NOM	<i>fumier</i> -NOM	ÉqT	1
<i>measure</i>	NOM	<i>mesure</i> -NOM	ÉqT	1
<i>moist</i>	ADJ	<i>humide</i> -ADJ	ÉqT	1
<i>monsoon</i>	NOM	<i>mousson</i> -NOM	ÉqT	1
<i>natural</i>	ADJ	<i>naturel</i> -ADJ	ÉqT	1
<i>organic</i>	ADJ	<i>organique</i> -ADJ	ÉqT	1
<i>organism</i>	NOM	<i>organisme</i> -NOM	ÉqT	1
<i>oscillation</i>	NOM	<i>oscillation</i> -NOM	ÉqT	1
<i>oxidation</i>	NOM	<i>oxydation</i> -NOM	ÉqT	1
<i>ozone</i>	NOM	<i>ozone</i> -NOM	ÉqT	1
<i>pasture</i>	NOM	<i>pâturage</i> -NOM	ÉqT	1
<i>photosynthesis</i>	NOM	<i>photosynthèse</i> -NOM	ÉqT	1
<i>phytoplankton</i>	NOM	<i>phytoplancton</i> -NOM	ÉqT	1
<i>polar</i>	ADJ	<i>polaire</i> -ADJ	ÉqT	1
<i>pollutant</i>	NOM	<i>polluant</i> -NOM	ÉqT	1
<i>prairie</i>	NOM	<i>prairie</i> -NOM	ÉqT	1
<i>precursor</i>	NOM	<i>précurseur</i> -NOM	ÉqT	1
<i>preserve</i>	VER	<i>préserver</i> -VER	ÉqT	1
<i>radiative</i>	ADJ	<i>radiatif</i> -ADJ	ÉqT	1
<i>recycling</i>	NOM	<i>recyclage</i> -NOM	ÉqT	1
<i>regeneration</i>	NOM	<i>régénération</i> -NOM	ÉqT	1
<i>regime</i>	NOM	<i>régime</i> -NOM	ÉqT	1
<i>renewable</i>	ADJ	<i>renouvelable</i> -ADJ	ÉqT	1
<i>reservoir</i>	NOM	<i>réservoir</i> -NOM	ÉqT	1

<i>resource</i>	NOM	<i>ressource</i> -NOM	ÉqT	1
<i>retention</i>	NOM	<i>réretention</i> -NOM	ÉqT	1
<i>salinity</i>	NOM	<i>salinité</i> -NOM	ÉqT	1
<i>seawater</i>	NOM	<i>eau de mer</i> -NOM	ÉqT	1
<i>sink</i>	NOM	<i>puits</i> -NOM	ÉqT	1
<i>snowfall</i>	NOM	<i>chute de neige</i> -NOM	ÉqT	1
<i>snowmelt1</i>	NOM	<i>eau de fonte</i> -NOM	ÉqT	1
<i>source</i>	NOM	<i>source</i> -NOM	ÉqT	1
<i>storm</i>	NOM	<i>tempête</i> -NOM	ÉqT	1
<i>stratosphere</i>	NOM	<i>stratosphère</i> -NOM	ÉqT	1
<i>sulphur</i>	NOM	<i>soufre</i> -NOM	ÉqT	1
<i>temperate</i>	ADJ	<i>tempéré</i> -ADJ	ÉqT	1
<i>temperature</i>	NOM	<i>température</i> -NOM	ÉqT	1
<i>thermohaline</i>	ADJ	<i>thermohalin</i> -ADJ	ÉqT	1
<i>tidal</i>	ADJ	<i>marée</i> -NOM	ÉqT	1
<i>tide</i>	NOM	<i>marée</i> -NOM	ÉqT	1
<i>tornado</i>	NOM	<i>tornade</i> -NOM	ÉqT	1
<i>troposphere</i>	NOM	<i>troposphère</i> -NOM	ÉqT	1
<i>tundra</i>	NOM	<i>toundra</i> -NOM	ÉqT	1
<i>upstream</i>	ADJ	<i>amont</i> -NOM	ÉqT	1
<i>urbanization</i>	NOM	<i>urbanisation</i> -NOM	ÉqT	1
<i>variable</i>	NOM	<i>variable</i> -NOM	ÉqT	1
<i>volcanic</i>	ADJ	<i>volcanique</i> -ADJ	ÉqT	1
<i>absorb</i>	VER	<i>absorber</i> -VER <i>absorption</i> -NOM	ÉqT ÉqNT	2
<i>adapt</i>	VER	<i>adapter</i> -VER <i>adaptation</i> -NOM	ÉqT ÉqNT	2
<i>adaptation</i>	NOM	<i>adaptation</i> -NOM <i>adapter</i> -VER	ÉqT ÉqNT	2
<i>animal</i>	NOM	<i>animal</i> -NOM <i>faune</i> -NOM	ÉqT ÉqNT	2
<i>aquifer</i>	NOM	<i>aquifère</i> -NOM <i>aquifère</i> -ADJ	ÉqT ÉqNT	2
<i>atmosphere</i>	NOM	<i>atmosphère</i> -NOM <i>atmosphérique</i> -ADJ	ÉqT ÉqNT	2
<i>atmospheric</i>	ADJ	<i>atmosphérique</i> -ADJ <i>atmosphère</i> -NOM	ÉqT ÉqNT	2
<i>biodiversity</i>	NOM	<i>biodiversité</i> -NOM <i>diversité biologique</i> -NOM	ÉqT ÉqT	2
<i>change</i>	VER	<i>changement</i> -NOM <i>changer</i> -VER	ÉqT ÉqNT	2
<i>chemical</i>	NOM	<i>produit chimique</i> -NOM <i>substance chimique</i> -NOM	ÉqT ÉqT	2
<i>chemistry</i>	NOM	<i>chimie</i> -NOM <i>chimique</i> -ADJ	ÉqT ÉqNT	2

<i>climate</i>	NOM	<i>climat</i> -NOM <i>climatique</i> -ADJ	ÉqT ÉqNT	2
<i>climatic</i>	ADJ	<i>climatique</i> -ADJ <i>climat</i> -NOM	ÉqT ÉqNT	2
<i>cloud</i>	NOM	<i>nuage</i> -NOM <i>nuageux</i> -ADJ	ÉqT ÉqNT	2
<i>concentration</i>	NOM	<i>concentration</i> -NOM <i>teneur</i> -NOM	ÉqT ÉqNT	2
<i>contamination</i>	NOM	<i>contamination</i> -NOM <i>pollution</i> -NOM	ÉqT ÉqNT	2
<i>cool</i>	VER	<i>refroidir</i> -VER <i>refroidissement</i> -NOM	ÉqT ÉqNT	2
<i>cooling</i>	NOM	<i>refroidissement</i> -NOM <i>refroidir</i> -VER	ÉqT ÉqNT	2
<i>coral</i>	NOM	<i>corail</i> -NOM <i>corallien</i> -ADJ	ÉqT ÉqNT	2
<i>cyclone</i>	NOM	<i>cyclone</i> -NOM <i>cyclonique</i> -ADJ	ÉqT ÉqNT	2
<i>degradation</i>	NOM	<i>dégradation</i> -NOM <i>détérioration</i> -NOM	ÉqT ÉqT	2
<i>disaster</i>	NOM	<i>catastrophe</i> -NOM <i>désastre</i> -NOM	ÉqT ÉqT	2
<i>drought</i>	NOM	<i>sécheresse</i> -NOM <i>période de sécheresse</i> -NOM	ÉqT ÉqT	2
<i>ecological</i>	ADJ	<i>écologique</i> -ADJ <i>écologie</i> -NOM	ÉqT ÉqNT	2
<i>ecosystem</i>	NOM	<i>écosystème</i> -NOM <i>écosystémique</i> -ADJ	ÉqT ÉqNT	2
<i>erode</i>	VER	<i>éroder</i> -VER <i>érosion</i> -NOM	ÉqT ÉqNT	2
<i>exposure</i>	NOM	<i>exposition</i> -NOM <i>exposer</i> -VER	ÉqT ÉqNT	2
<i>extreme</i>	NOM	<i>extrême</i> -NOM <i>extrême</i> -ADJ	ÉqT ÉqNT	2
<i>fire</i>	NOM	<i>feu</i> -NOM <i>incendie</i> -NOM	ÉqT ÉqT	2
<i>fishery2</i>	NOM	<i>pêche</i> -NOM <i>halieutique</i> -ADJ	ÉqT ÉqNT	2
<i>flood1</i>	NOM	<i>inondation</i> -NOM <i>débordement</i> -NOM	ÉqT ÉqNT	2
<i>forested</i>	ADJ	<i>boisé</i> -ADJ <i>forestier</i> -ADJ	ÉqT ÉqNT	2
<i>frost</i>	NOM	<i>gel</i> -NOM <i>gelée</i> -NOM	ÉqT ÉqT	2
<i>fuel</i>	NOM	<i>combustible</i> -NOM <i>carburant</i> -NOM	ÉqT ÉqT	2



<i>gas</i>	NOM	<i>gaz</i> -NOM <i>gazeux</i> -ADJ	ÉqT ÉqNT	2
<i>glacier</i>	NOM	<i>glacier</i> -NOM <i>glaciaire</i> -ADJ	ÉqT ÉqNT	2
<i>grassland</i>	NOM	<i>prairie</i> -NOM <i>pâturage</i> -NOM	ÉqT ÉqNT	2
<i>harvest</i>	VER	<i>récolter</i> -VER <i>récolte</i> -NOM	ÉqT ÉqNT	2
<i>heat</i>	NOM	<i>chaleur</i> -NOM <i>thermique</i> -ADJ	ÉqT ÉqNT	2
<i>hydrological</i>	ADJ	<i>hydrologique</i> -ADJ <i>hydrologie</i> -NOM	ÉqT ÉqNT	2
<i>hydrology</i>	NOM	<i>hydrologie</i> -NOM <i>hydrologique</i> -ADJ	ÉqT ÉqNT	2
<i>island</i>	NOM	<i>île</i> -NOM <i>insulaire</i> -ADJ	ÉqT ÉqNT	2
<i>lake</i>	NOM	<i>lac</i> -NOM <i>lacustre</i> -ADJ	ÉqT ÉqNT	2
<i>land</i>	NOM	<i>terre</i> -NOM <i>terrestre</i> -ADJ	ÉqT ÉqNT	2
<i>latitudinal</i>	ADJ	<i>latitudinal</i> -ADJ <i>latitude</i> -NOM	ÉqT ÉqNT	2
<i>locally</i>	ADV	<i>localement</i> -ADV (au niveau, à l'échelle) <i>local</i> -ADJ	ÉqT ÉqNT	2
<i>mapping</i>	NOM	<i>cartographie</i> -NOM <i>carte</i> -NOM	ÉqT ÉqNT	2
<i>marine</i>	ADJ	<i>marin</i> -ADJ <i>mer</i> -NOM	ÉqT ÉqNT	2
<i>melt</i>	VER	<i>fondre</i> -VER <i>fonte</i> -NOM	ÉqT ÉqNT	2
<i>meteorological</i>	ADJ	<i>météorologique</i> -ADJ <i>météorologie</i> -NOM	ÉqT ÉqNT	2
<i>methane</i>	NOM	<i>méthane</i> -NOM <i>CH4</i> -SYM	ÉqT ÉqT	2
<i>model</i>	NOM	<i>modèle</i> -NOM <i>modélisation</i> -NOM	ÉqT ÉqNT	2
<i>modeling</i>	NOM	<i>modélisation</i> -NOM <i>modèle</i> -NOM	ÉqT ÉqNT	2
<i>mountainous</i>	ADJ	<i>montagneux</i> -ADJ <i>montagne</i> -NOM	ÉqT ÉqNT	2
<i>nitrogen</i>	NOM	<i>azote</i> -NOM <i>azoté</i> -ADJ	ÉqT ÉqNT	2
<i>oceanic</i>	ADJ	<i>océanique</i> -ADJ <i>océan</i> -NOM	ÉqT ÉqNT	2
<i>pollution</i>	NOM	<i>pollution</i> -NOM <i>polluant</i> -NOM	ÉqT ÉqNT	2

<i>protection</i>	NOM	<i>protection</i> -NOM <i>protéger</i> -VER	ÉqT ÉqNT	2
<i>rain</i>	NOM	<i>pluie</i> -NOM <i>précipitation</i> -NOM	ÉqT ÉqNT	2
<i>rainfall1</i>	NOM	<i>pluie</i> -NOM <i>précipitation</i> -NOM	ÉqT ÉqNT	2
<i>rainfall2</i>	NOM	<i>pluviosité</i> -NOM <i>pluviométrie</i> -NOM	ÉqT ÉqNT	2
<i>recharge</i>	NOM	<i>recharge</i> -NOM <i>alimentation</i> -NOM	ÉqT ÉqNT	2
<i>regional</i>	ADJ	<i>régional</i> -ADJ (à l'échelle de la, dans la, pour une) <i>région</i> -NOM	ÉqT ÉqNT	2
<i>runoff</i>	NOM	<i>ruissellement</i> -NOM <i>écoulement</i> -NOM	ÉqT ÉqNT	2
<i>scenario</i>	NOM	<i>scénario</i> -NOM <i>hypothèse</i> -NOM	ÉqT ÉqNT	2
<i>seasonal</i>	ADJ	<i>saisonnier</i> -ADJ <i>saison</i> -NOM	ÉqT ÉqNT	2
<i>seasonality</i>	NOM	<i>saisonnalité</i> -NOM <i>saisonnier</i> -ADJ	ÉqT ÉqNT	2
<i>seasonally</i>	ADV	<i>saisonnier</i> -ADJ <i>saison</i> -NOM	ÉqT ÉqNT	2
<i>sediment</i>	NOM	<i>sédiment</i> -NOM <i>sédimentaire</i> -ADJ	ÉqT ÉqNT	2
<i>sensitivity</i>	NOM	<i>sensibilité</i> -NOM <i>sensible</i> -ADJ	ÉqT ÉqNT	2
<i>simulation</i>	NOM	<i>simulation</i> -NOM <i>simuler</i> -VER	ÉqT ÉqNT	2
<i>solar</i>	ADJ	<i>solaire</i> -ADJ <i>soleil</i> -NOM	ÉqT ÉqNT	2
<i>stabilize</i>	VER	<i>stabiliser</i> -VER <i>stabilisation</i> -NOM	ÉqT ÉqNT	2
<i>store</i>	VER	<i>stocker</i> -VER <i>stockage</i> -NOM	ÉqT ÉqNT	2
<i>stratospheric</i>	ADJ	<i>stratosphérique</i> -ADJ <i>stratosphère</i> -NOM	ÉqT ÉqNT	2
<i>streamflow</i>	NOM	<i>écoulement</i> -NOM <i>débit</i> -NOM	ÉqT ÉqT	2
<i>terrestrial</i>	ADJ	<i>terrestre</i> -ADJ <i>terre</i> -NOM	ÉqT ÉqNT	2
<i>thaw</i>	NOM	<i>dégel</i> -NOM <i>fonte</i> -NOM	ÉqT ÉqT	2
<i>thunderstorm</i>	NOM	<i>orage</i> -NOM <i>orageux</i> -ADJ	ÉqT ÉqNT	2
<i>tropospheric</i>	ADJ	<i>troposphérique</i> -ADJ <i>troposphère</i> -NOM	ÉqT ÉqNT	2

<i>uncertainty</i>	NOM	<i>incertitude</i> -NOM <i>incertain</i> -ADJ	ÉqT ÉqNT	2
<i>variability</i>	NOM	<i>variabilité</i> -NOM <i>varier</i> -VER	ÉqT ÉqNT	2
<i>vegetation</i>	NOM	<i>végétation</i> -NOM <i>végétal</i> -ADJ	ÉqT ÉqNT	2
<i>warm</i>	ADJ	<i>chaud</i> -ADJ <i>réchauffement</i> -NOM	ÉqT ÉqNT	2
<i>warm</i>	VER	<i>réchauffer</i> -VER <i>réchauffement</i> -NOM	ÉqT ÉqNT	2
<i>warming</i>	NOM	<i>réchauffement</i> -NOM <i>réchauffer</i> -VER	ÉqT ÉqNT	2
<i>water</i>	NOM	<i>eau</i> -NOM <i>hydrique</i> -ADJ	ÉqT ÉqNT	2
<i>wind</i>	NOM	<i>vent</i> -NOM <i>éolien</i> -ADJ	ÉqT ÉqNT	2
<i>wood</i>	NOM	<i>bois</i> -NOM <i>ligneux</i> -ADJ	ÉqT ÉqNT	2
<i>adaptive</i>	ADJ	<i>adaptatif</i> -ADJ <i>adaptation</i> -NOM <i>adapter</i> -VER	ÉqNT ÉqT ÉqNT	3
<i>agricultural</i>	ADJ	<i>agricole</i> -ADJ <i>agriculture</i> -NOM <i>culture</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>agriculture</i>	NOM	<i>agriculture</i> -NOM <i>agricole</i> -ADJ <i>culture</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>air</i>	NOM	<i>air</i> -NOM <i>atmosphérique</i> -ADJ <i>aérien</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>crop</i>	NOM	<i>récolte</i> -NOM <i>culture</i> -NOM <i>agricole</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>cropland</i>	NOM	<i>terre cultivée</i> -NOM <i>terre agricole</i> -NOM <i>terre arable</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>disease</i>	NOM	<i>maladie</i> -NOM <i>affection</i> -NOM <i>épidémie</i> -NOM	ÉqT ÉqT ÉqNT	3
<i>electricity</i>	NOM	<i>électricité</i> -NOM <i>électrique</i> -ADJ <i>énergie</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>emit</i>	VER	<i>émettre</i> -VER <i>émission</i> -NOM <i>rejeter</i> -VER	ÉqT ÉqNT ÉqNT	3
<i>energy</i>	NOM	<i>énergie</i> -NOM <i>énergétique</i> -ADJ	ÉqT ÉqNT	3

		<i>électricité</i> -NOM	ÉqNT	
<i>environment</i>	NOM	<i>environnement</i> -NOM <i>milieu</i> -NOM <i>environnemental</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>environmentally</i>	ADV	<i>environnemental</i> -ADJ <i>environnement</i> -NOM <i>écologiquement</i> -ADV	ÉqT ÉqNT ÉqNT	3
<i>event</i>	NOM	<i>événement</i> -NOM <i>phénomène</i> -NOM <i>épisode</i> -NOM	ÉqT ÉqT ÉqNT	3
<i>extinction</i>	NOM	<i>extinction</i> -NOM <i>disparition</i> -NOM <i>disparaître</i> -VER	ÉqT ÉqT ÉqNT	3
<i>fish</i>	NOM	<i>poisson</i> -NOM <i>halieutique</i> -ADJ <i>piscicole</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>flood</i>	VER	<i>inonder</i> -VER <i>inondation</i> -NOM <i>crue</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>glacial</i>	ADJ	<i>glaciaire</i> -ADJ <i>glacier</i> -NOM <i>glace</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>groundwater</i>	NOM	<i>eau souterraine</i> -NOM <i>nappe souterraine</i> -NOM <i>nappe phréatique</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>harmful</i>	ADJ	<i>nocif</i> -ADJ <i>nuisible</i> -ADJ <i>néfaste</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>intensity</i>	NOM	<i>intensité</i> -NOM <i>intensification</i> -NOM <i>intense</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>mining</i>	NOM	<i>minier</i> -ADJ <i>mine</i> -NOM <i>extraction</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>model</i>	VER	<i>modéliser</i> -VER <i>modélisation</i> -NOM <i>modèle</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>moisture</i>	NOM	<i>humidité</i> -NOM <i>hydrique</i> -ADJ <i>eau</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>monitor</i>	VER	<i>surveiller</i> -VER <i>sivre</i> -VER <i>surveillance</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>mountain</i>	NOM	<i>montagne</i> -NOM <i>montagneux</i> -ADJ <i>montagnard</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>projection</i>	NOM	<i>projection</i> -NOM	ÉqT	3

		prévision-NOM prévoir-VER	ÉqT ÉqNT	
<i>regionally</i>	ADV	<i>régionalement</i> -ADV (à l'échelle, au niveau, sur le plan) <i>régional</i> -ADJ (d'une ... à l'autre, selon les) région-NOM	ÉqT ÉqNT ÉqNT	3
<i>resilience</i>	NOM	<i>résilience</i> -NOM <i>résistance</i> -NOM <i>capacité d'adaptation</i> -NOM	ÉqT ÉqNT ÉqT	3
<i>road</i>	NOM	<i>route</i> -NOM <i>chemin</i> -NOM <i>routier</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>scarcity</i>	NOM	<i>pénurie</i> -NOM <i>rareté</i> -NOM <i>raréfaction</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>sea</i>	NOM	<i>mer</i> -NOM <i>marin</i> -ADJ <i>océan</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>simulate</i>	VER	<i>simuler</i> -VER <i>simulé</i> -ADJ <i>simulation</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>simulated</i>	ADJ	<i>simulé</i> -ADJ <i>simuler</i> -VER <i>simulation</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>snow</i>	NOM	<i>neige</i> -NOM <i>neigeux</i> -ADJ <i>nival</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>stream</i>	NOM	<i>rivière</i> -NOM <i>fluvial</i> -ADJ <i>cours d'eau</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>sustainable</i>	ADJ	<i>durable</i> -ADJ <i>durabilité</i> -NOM <i>viable</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>tree</i>	NOM	<i>arbre</i> -NOM <i>forestier</i> -ADJ <i>arbustif</i> -ADJ	ÉqT ÉqNT ÉqNT	3
<i>watershed</i>	NOM	<i>bassin versant</i> -NOM <i>bassin</i> -NOM <i>bassin hydrographique</i> -NOM	ÉqT ÉqT ÉqNT	3
<i>worldwide</i>	ADV	<i>mondial</i> -ADJ <i>monde</i> -NOM <i>planète</i> -NOM	ÉqT ÉqNT ÉqNT	3
<i>assessment</i>	NOM	<i>évaluation</i> -NOM <i>évaluer</i> -VER <i>analyse</i> -NOM <i>étude</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4

<i>change</i>	NOM	<i>changement</i> -NOM <i>changer</i> -VER <i>variation</i> -NOM <i>varier</i> -VER	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>compliance</i>	NOM	<i>conformité</i> -NOM <i>respecter</i> -VER <i>respect</i> -NOM <i>conformer</i> -VER	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>consequence</i>	NOM	<i>conséquence</i> -NOM <i>répercussion</i> -NOM <i>incidence</i> -NOM <i>effet</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>damage</i>	NOM	<i>dommage</i> -NOM <i>dégât</i> -NOM <i>endommager</i> -VER <i>dégradation</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>environmental</i>	ADJ	<i>environnemental</i> -ADJ <i>environnement</i> -NOM <i>milieu</i> -NOM <i>écologique</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>extreme</i>	ADJ	<i>extrême</i> -ADJ <i>extrême</i> -NOM <i>intense</i> -ADJ <i>extrêmement</i> -ADV	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>flooding</i>	NOM	<i>inondation</i> -NOM <i>crue</i> -NOM <i>inonder</i> -VER <i>débordement</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>food</i>	NOM	<i>aliment</i> -NOM <i>nourriture</i> -NOM <i>alimentaire</i> -ADJ <i>alimentation</i> -NOM	ÉqT ÉqT ÉqNT ÉqNT	4
<i>forest</i>	NOM	<i>forêt</i> -NOM <i>forestier</i> -ADJ <i>peuplement</i> -NOM <i>boisé</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>intense</i>	ADJ	<i>intense</i> -ADJ <i>intensité</i> -NOM <i>intensification</i> -NOM <i>grave</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>negative</i>	ADJ	<i>néгатif</i> -ADJ <i>néfaste</i> -ADJ <i>défavorable</i> -ADJ <i>fâcheux</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>negatively</i>	ADV	<i>néгатivement</i> -ADV (de manière, de façon) <i>néгатif</i> -ADJ <i>néfaste</i> -ADJ	ÉqT ÉqNT ÉqNT	4

		<i>nuire</i> -VER	ÉqNT	
<i>ocean</i>	NOM	<i>océan</i> -NOM <i>océanique</i> -ADJ <i>mer</i> -NOM <i>marin</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>plant</i>	NOM	<i>plante</i> -NOM <i>végétal</i> -ADJ <i>végétal</i> -NOM <i>flore</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>population</i>	NOM	<i>population</i> -NOM <i>démographique</i> -ADJ <i>habitant</i> -NOM <i>peuplement</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>precipitation</i>	NOM	<i>précipitation</i> -NOM <i>pluie</i> -NOM <i>pluviosité</i> -NOM <i>pluviométrie</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>recovery</i>	NOM	<i>régénération</i> -NOM <i>rétablissement</i> -NOM <i>reconstitution</i> -NOM <i>restauration</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>reduction</i>	NOM	<i>réduction</i> -NOM <i>diminution</i> -NOM <i>baisse</i> -NOM <i>réduire</i> -VER	ÉqT ÉqT ÉqT ÉqNT	4
<i>release</i>	NOM	<i>libération</i> -NOM <i>rejet</i> -NOM <i>émission</i> -NOM <i>libérer</i> -VER	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>sustainability</i>	NOM	<i>durabilité</i> -NOM <i>viabilité</i> -NOM <i>durable</i> -ADJ <i>pérennité</i> -NOM	ÉqT ÉqT ÉqNT ÉqNT	4
<i>uptake</i>	NOM	<i>absorption</i> -NOM <i>absorber</i> -VER <i>fixation</i> -NOM <i>fixer</i> -VER	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>variable</i>	ADJ	<i>variable</i> -ADJ <i>variation</i> -NOM <i>varier</i> -VER <i>variabilité</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>variation</i>	NOM	<i>variation</i> -NOM <i>fluctuation</i> -NOM <i>varier</i> -VER <i>variabilité</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>vary</i>	VER	<i>varier</i> -VER <i>variation</i> -NOM	ÉqT ÉqNT	4

		<i>fluctuer</i> -VER <i>variable</i> -ADJ	ÉqNT ÉqNT	
<i>weather</i>	NOM	<i>temps</i> -NOM <i>météorologique</i> -ADJ <i>condition atmosphérique</i> -NOM <i>climatique</i> -ADJ	ÉqT ÉqNT ÉqNT ÉqNT	4
<i>world</i>	NOM	<i>monde</i> -NOM <i>planète</i> -NOM <i>globe</i> -NOM <i>mondial</i> -ADJ	ÉqT ÉqT ÉqT ÉqNT	4
<i>effect</i>	NOM	<i>effet</i> -NOM <i>conséquence</i> -NOM <i>impact</i> -NOM <i>répercussion</i> -NOM <i>incidence</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT ÉqNT	5
<i>hazard</i>	NOM	<i>danger</i> -NOM <i>risque</i> -NOM <i>catastrophe</i> -NOM <i>phénomène</i> -NOM <i>aléa</i> -NOM	ÉqT ÉqT ÉqNT ÉqNT ÉqNT	5
<i>outbreak</i>	NOM	<i>épidémie</i> -NOM <i>prolifération</i> -NOM <i>infestation</i> -NOM <i>invasion</i> -NOM <i>pullulation</i> -NOM	ÉqT ÉqNT ÉqNT ÉqNT ÉqNT	5
<i>release</i>	VER	<i>libérer</i> -VER <i>libération</i> -NOM <i>rejeter</i> -VER <i>rejet</i> -NOM <i>dégager</i> -VER	ÉqT ÉqNT ÉqNT ÉqNT ÉqNT	5
<i>rise</i>	NOM	<i>élévation</i> -NOM <i>hausse</i> -NOM <i>augmentation</i> -NOM <i>montée</i> -NOM <i>monter</i> -VER	ÉqT ÉqNT ÉqT ÉqNT ÉqNT	5
<i>river</i>	NOM	<i>rivière</i> -NOM <i>fleuve</i> -NOM <i>cours d'eau</i> -NOM <i>fluvial</i> -ADJ <i>hydrographique</i> -ADJ	ÉqT ÉqT ÉqNT ÉqNT ÉqNT	5
<i>assess</i>	VER	<i>évaluer</i> -VER <i>estimer</i> -VER <i>déterminer</i> -VER <i>évaluation</i> -NOM <i>mesurer</i> -VER <i>analyser</i> -VER	ÉqT ÉqT ÉqNT ÉqNT ÉqT ÉqNT	6
<i>decrease</i>	VER	<i>diminuer</i> -VER	ÉqT	6



		<i>baisser-VER</i> <i>réduire-VER</i> <i>diminution-NOM</i> <i>baisse-NOM</i> <i>réduction-NOM</i>	ÉqNT ÉqNT ÉqNT ÉqNT ÉqNT	
<i>human</i>	NOM	<i>humain-NOM</i> <i>homme-NOM</i> <i>humain-ADJ</i> <i>personne-NOM</i> <i>population-NOM</i> <i>anthropique-ADJ</i>	ÉqT ÉqT ÉqNT ÉqNT ÉqNT ÉqNT	6
<i>reduce</i>	VER	<i>réduire-VER</i> <i>diminuer-VER</i> <i>baisser-VER</i> <i>réduction-NOM</i> <i>diminution-NOM</i> <i>baisse-NOM</i>	ÉqT ÉqT ÉqT ÉqNT ÉqNT ÉqNT	6
<i>rise</i>	VER	<i>élever-VER</i> <i>monter-VER</i> <i>augmenter-VER</i> <i>élévation-NOM</i> <i>hausse-NOM</i> <i>montée-NOM</i>	ÉqT ÉqNT ÉqT ÉqNT ÉqNT ÉqNT	6
<i>vulnerable</i>	ADJ	<i>vulnérable-ADJ</i> <i>vulnérabilité-NOM</i> <i>fragile-ADJ</i> <i>exposé-ADJ</i> <i>touché-ADJ</i> <i>menacé-ADJ</i>	ÉqT ÉqNT ÉqNT ÉqNT ÉqNT ÉqNT	6
<i>increase</i>	NOM	<i>augmentation-NOM</i> <i>augmenter-VER</i> <i>hausse-NOM</i> <i>accroissement-NOM</i> <i>intensification-NOM</i> <i>intensifier-VER</i> <i>accentuer-VER</i>	ÉqT ÉqNT ÉqNT ÉqT ÉqT ÉqNT ÉqNT	7
<i>pattern</i>	NOM	<i>modèle-NOM</i> <i>schéma-NOM</i> <i>tendance-NOM</i> <i>configuration-NOM</i> <i>mode-NOM</i> <i>régime-NOM</i> <i>habitude-NOM</i>	ÉqNT ÉqNT ÉqT ÉqNT ÉqNT ÉqNT ÉqNT	7
<i>increase</i>	VER	<i>augmenter-VER</i> <i>augmentation-NOM</i> <i>accroître-VER</i> <i>accroissement-NOM</i>	ÉqT ÉqNT ÉqT ÉqNT	8



## Annexe F : Analyse des 67 termes anglais avec des différences d'équivalence

– Liste des abréviations utilisées dans cette Annexe (voir *Liste des symboles et abréviations* pour une description complète).

CC	corpus comparable
CP	corpus parallèle
ÉqNT	équivalent non typique
ÉqT	équivalent typique
ÉqF	équivalent fréquent
ÉqPF	équivalent peu fréquent

- Lorsque utiles, certaines fréquences sont indiquées entre parenthèses.
- Les équivalents différents analysés sont en gras dans le tableau.

Tableau F.1. Terme anglais avec différence d'équivalence peu marquée ( $\pm$ )

Terme anglais	Équivalent CP	Équivalent CC	ÉqT/ÉqNT	ÉqF/ÉqPF	Commentaire
<i>aerosol</i> -NOM	<i>aérosol</i> -NOM (ÉqT)	<i>aérosol</i> -NOM (ÉqT) <b><i>particule</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	Dans le CC, <i>particule</i> -NOM remplace <i>aérosol</i> -NOM dans certains contextes (ex. <i>réduire les émissions d'aérosols, de particules</i> ).
<i>bioenergy</i> -NOM	<i>bioénergie</i> -NOM (ÉqT) <b><i>bioénergétique</i></b> -ADJ (ÉqNT) (9)	<i>bioénergie</i> -NOM (ÉqT) (8)	ÉqNT	ÉqPF	<i>bioénergétique</i> -ADJ est un équivalent transcatégoriel
<i>biofuel</i> -NOM	<i>biocarburant</i> -NOM (ÉqT) <b><i>biocombustible</i></b> -NOM (ÉqT) (9)	<i>biocarburant</i> -NOM (ÉqT)	ÉqT	ÉqPF	<i>biocombustible</i> -NOM n'est pas synonyme de <i>biocarburant</i> -NOM. Le

					terme anglais couvre une réalité plus grande qu'en français (IATE).
<i>coal</i> -NOM	<i>charbon</i> -NOM (ÉqT) <i>houille</i> -NOM (ÉqT) <b><i>houiller</i></b> -ADJ (ÉqNT) (9)	<i>charbon</i> -NOM (ÉqT) <i>houille</i> -NOM (ÉqT) <b><i>charbonnier</i></b> -ADJ (ÉqNT) (25)	ÉqNT ÉqNT	ÉqPF 25 occ.	<i>houiller</i> -ADJ et <i>charbonnier</i> -ADJ sont des équivalents transcatégoriels
<i>coastal</i> -ADJ	<i>côtier</i> -ADJ (ÉqT) <i>côte</i> -NOM (ÉqNT) <i>littoral</i> -NOM (ÉqNT)	<i>côtier</i> -ADJ (ÉqT) <i>côte</i> -NOM (ÉqNT) <i>littoral</i> -NOM (ÉqNT) <b><i>littoral</i></b> -ADJ (ÉqT)	ÉqT	s.o.	<i>côtier</i> -ADJ et <i>littoral</i> -ADJ sont synonymes (GDT, Le grand Robert & Collins)
<i>coastline</i> -NOM	<i>littoral</i> -NOM (ÉqT) <i>côte</i> -NOM (ÉqT)	<i>littoral</i> -NOM (ÉqT) <i>côte</i> -NOM (ÉqT) <b><i>trait de côte</i></b> -NOM (ÉqNT) <b><i>rivage</i></b> -NOM (ÉqNT)	ÉqNT ÉqNT	s.o. s.o.	Dans le CC, <i>trait de côte</i> -NOM et <i>rivage</i> -NOM sont employés dans des contextes semblables au CP (ex. <i>recul du trait de côte</i> , <i>recul du rivage</i> , <i>recul des côtes</i> )
<i>coast</i> -NOM	<i>côte</i> -NOM (ÉqT) <i>littoral</i> -NOM (ÉqT) <i>côtier</i> -ADJ (ÉqNT)	<i>côte</i> -NOM (ÉqT) <i>littoral</i> -NOM (ÉqT) <i>côtier</i> -ADJ (ÉqNT) <b><i>littoral</i></b> -ADJ (ÉqNT) <b><i>rivage</i></b> -NOM (ÉqNT)	ÉqNT ÉqNT	s.o. s.o.	<i>littoral</i> -ADJ est un équivalent transcatégoriel. <i>rivage</i> -NOM est un quasi-synonyme de <i>côte</i> -NOM (Antidote).
<i>cultivation</i> -NOM	<i>culture</i> -NOM (ÉqT)	<i>culture</i> -NOM (ÉqT) <b><i>cultural</i></b> -ADJ (ÉqNT) (31)	ÉqNT	s.o.	<i>culture</i> -ADJ est un équivalent transcatégoriel
<i>delta</i> -NOM	<i>delta</i> -NOM (ÉqT)	<i>delta</i> -NOM (ÉqT) <b><i>deltaïque</i></b> -ADJ (ÉqNT) (6)	ÉqNT	ÉqPF	<i>deltaïque</i> -ADJ est un équivalent transcatégoriel
<i>forestry</i> -NOM	<i>foresterie</i> -NOM (ÉqT) <i>forestier</i> -ADJ (ÉqNT)	<i>foresterie</i> -NOM (ÉqT) <i>forestier</i> -ADJ (ÉqNT)	ÉqNT	s.o.	<i>sylvicole</i> -ADJ est un équivalent transcatégoriel

	<i>forêt</i> -NOM (ÉqNT) <i>sylviculture</i> -NOM (ÉqNT)	<i>forêt</i> -NOM (ÉqNT) <i>sylviculture</i> -NOM (ÉqNT) <b><i>sylvicole</i></b> -ADJ (ÉqNT)			
<i>fragmentation</i> -NOM	<i>fragmentation</i> -NOM (ÉqT) <b><i>morcellement</i></b> -NOM (ÉqNT) (6)	<i>fragmentation</i> -NOM (ÉqT)	ÉqNT	ÉqPF	La <i>fragmentation</i> est un processus de <i>morcellement</i> (Bioret <i>et al.</i> 2009).
<i>freshwater</i> -NOM	<i>eau douce</i> -NOM (ÉqT) <b><i>dulcicole</i></b> -ADJ (ÉqNT) (15)	<i>eau douce</i> -NOM (ÉqT)	ÉqNT	ÉqPF	<i>dulcicole</i> -ADJ est un équivalent transcatégoriel
<i>globally</i> -ADV	<i>mondial</i> -ADJ (ÉqNT) <i>monde</i> -NOM (ÉqNT) <i>globe</i> -NOM (ÉqNT) <i>planétaire</i> -ADJ (ÉqNT) <i>planète</i> -NOM (10) (ÉqNT) <b><i>mondialement</i></b> -ADV (ÉqT) (9)	<i>mondial</i> -ADJ (ÉqNT) <i>monde</i> -NOM (ÉqNT) <i>globe</i> -NOM (ÉqNT) <i>planétaire</i> -ADJ (ÉqNT) <i>planète</i> -NOM (ÉqNT) <b><i>international</i></b> -ADJ (ÉqNT)	ÉqT ÉqNT	ÉqPF s.o.	<i>mondialement</i> -ADV est absent du CC. <i>international</i> -ADJ est un équivalent transcatégoriel.
<i>grain</i> -NOM	<i>céréale</i> -NOM (ÉqT) <i>grain</i> -NOM (ÉqNT) <b><i>céréaliier</i></b> -ADJ (12) (ÉqNT)	<i>céréale</i> -NOM (ÉqT) <i>grain</i> -NOM (ÉqNT)	ÉqNT	ÉqPF	<i>céréaliier</i> -ADJ est un équivalent transcatégoriel.
<i>halocarbon</i> -NOM	<i>halocarbure</i> -NOM (ÉqT) <i>hydrocarbure halogéné</i> -NOM (ÉqT) <b><i>gaz halocarboné</i></b> -NOM (ÉqNT) (4)	<i>halocarbure</i> -NOM (ÉqT) <i>hydrocarbure halogéné</i> -NOM (ÉqT)	ÉqNT	ÉqPF	<i>gaz halocarboné</i> -NOM est moins spécialisé que <i>halocarbure</i> -NOM et que <i>hydrocarbure halogéné</i> -NOM.
<i>harvest</i> -NOM	<i>récolte</i> -NOM (ÉqT) <i>récolter</i> -VER (ÉqNT)	<i>récolte</i> -NOM (ÉqT) <i>récolter</i> -VER (ÉqNT) <b><i>cueillette</i></b> -NOM (ÉqNT) (11)	ÉqNT	ÉqPF	<i>cueillette</i> -NOM est un quasi-synonyme de <i>récolte</i> -NOM (Petit Robert).
<i>human</i> -ADJ	<i>humain</i> -ADJ (ÉqT) <i>homme</i> -NOM (ÉqNT)	<i>humain</i> -ADJ (ÉqT) <i>homme</i> -NOM (ÉqNT)	ÉqNT	s.o.	<i>humain</i> -NOM est un équivalent transcatégoriel

	<i>personne</i> -NOM (ÉqNT) <i>population</i> -NOM (ÉqNT) <i>anthropique</i> -ADJ (ÉqNT)	<i>personne</i> -NOM (ÉqNT) <i>population</i> -NOM (ÉqNT) <i>anthropique</i> -ADJ (ÉqNT) <b><i>humain</i></b> -NOM (ÉqNT)			
<i>ice</i> -NOM	<i>glace</i> -NOM (ÉqT) <i>glaciaire</i> -ADJ (ÉqNT) <b><i>glaciel</i></b> -ADJ (ÉqNT) (38)	<i>glace</i> -NOM (ÉqT) <i>glaciaire</i> -ADJ (ÉqNT)	ÉqNT	38 occ.	<i>glaciel</i> -ADJ est un équivalent transcatégoriel
<i>infestation</i> -NOM	<i>infestation</i> -NOM (ÉqT) <i>prolifération</i> -NOM (ÉqNT)	<i>infestation</i> -NOM (ÉqT) <i>prolifération</i> -NOM (ÉqNT) <b><i>invasion</i></b> -NOM (ÉqNT) (4)	ÉqNT	ÉqPF	Autre point de vue (ex. <i>invasion des espèces, infestation par les insectes</i> ).
<i>landfill</i> -NOM	<i>décharge</i> -NOM (ÉqT)	<i>décharge</i> -NOM (ÉqT) <b><i>centre d'enfouissement</i></b> -NOM (ÉqNT) (5)	ÉqNT	ÉqPF	<i>centre d'enfouissement</i> -NOM est un quasi-synonyme de <i>décharge</i> -NOM (Wikipedia)
<i>landslide</i> -NOM	<i>glissement de terrain</i> -NOM (ÉqT)	<i>glissement de terrain</i> -NOM (ÉqT) <b><i>éboulement</i></b> -NOM (ÉqT)(6)	ÉqT	ÉqPF	<i>éboulement</i> -NOM est synonyme de <i>glissement de terrain</i> -NOM (GDT, eau et géologie)
<i>lifetime</i> -NOM	<i>durée de vie</i> -NOM (ÉqT) <i>durée</i> -NOM (ÉqT)	<i>durée de vie</i> -NOM (ÉqT) <i>durée</i> -NOM (ÉqT) <b><i>temps de résidence</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	<i>temps de résidence</i> -NOM est synonyme selon Termium et IATE.
<i>livestock</i> -NOM	<i>bétail</i> -NOM (ÉqT) <i>élevage</i> -NOM (ÉqNT) <i>animal</i> -ADJ (ÉqNT) <i>animal</i> -NOM (ÉqNT)	<i>bétail</i> -NOM (ÉqT) <i>élevage</i> -NOM (ÉqNT) <i>animal</i> -ADJ (ÉqNT) <i>animal</i> -NOM (ÉqNT) <b><i>cheptel</i></b> -NOM (ÉqNT) (10) <b><i>troupeau</i></b> -NOM (ÉqNT) (12)	ÉqNT ÉqNT	ÉqPF ÉqPF	<i>cheptel</i> -NOM et <i>troupeau</i> -NOM sont des quasi-synonymes (Antidote).
<i>lowland</i> -NOM (16)	<i>basse terre</i> -NOM (ÉqT) <b><i>terre basse</i></b> -NOM (ÉqNT) (4)	<i>basse terre</i> -NOM (ÉqT) (7)	ÉqNT	ÉqPF	<i>terre basse</i> -NOM est attesté dans le Petit

					Robert.
<i>magnitude</i> -NOM	<i>ampleur</i> -NOM (ÉqT) <i>ordre de grandeur</i> -NOM (ÉqNT) <i>amplitude</i> -NOM (ÉqNT) <i>importance</i> -NOM (ÉqNT) <i>intensité</i> -NOM (ÉqNT) <i>envergure</i> -NOM (ÉqNT) <b><i>magnitude</i></b> -NOM (ÉqNT) (4)	<i>ampleur</i> -NOM (ÉqT) <i>ordre de grandeur</i> -NOM (ÉqNT) <i>amplitude</i> -NOM (ÉqNT) <i>importance</i> -NOM (ÉqNT) <i>intensité</i> -NOM (ÉqNT) <i>envergure</i> -NOM (ÉqNT)	ÉqNT	ÉqPF	<i>magnitude</i> -NOM est un quasi-synonyme d' <i>ampleur</i> -NOM (Antidote)
<i>marsh</i> -NOM	<i>marais</i> -NOM (ÉqT) <b><i>marécage</i></b> -NOM (ÉqNT) (5)	<i>marais</i> -NOM (ÉqT)	ÉqNT	ÉqPF	<i>marécage</i> -NOM est un synonyme selon IATE (agriculture, sylviculture et pêche), mais un quasi-synonyme selon le GDT (géomorphologie)
<i>melting</i> -NOM	<i>fonte</i> -NOM (ÉqT) <b><i>fusion</i></b> -NOM (ÉqNT) (4)	<i>fonte</i> -NOM (ÉqT)	ÉqNT	ÉqPF	<i>fusion</i> -NOM est un quasi-synonyme (GDT, eau)
<i>migrate</i> -VER	<i>migrer</i> -VER (ÉqT) <i>migration</i> -NOM (ÉqNT) <i>déplacer</i> -VER (se) (ÉqNT)	<i>migrer</i> -VER (ÉqT) <i>migration</i> -NOM (ÉqNT) <i>déplacer</i> -VER (se) (ÉqNT) <b><i>migratoire</i></b> -ADJ (ÉqNT)	ÉqNT	s.o.	<i>migratoire</i> -ADJ équivalent transcategoriel (ex. de contexte : <i>capacité de migrer</i> , <i>capacité migratoire</i> )
<i>migration</i> -NOM	<i>migration</i> -NOM (ÉqT) <i>migratoire</i> -ADJ (ÉqNT) <i>migrer</i> -VER (ÉqNT) <i>déplacement</i> -NOM (ÉqNT) <b><i>exode</i></b> -NOM (ÉqNT) (5)	<i>migration</i> -NOM (ÉqT) <i>migratoire</i> -ADJ (ÉqNT) <i>migrer</i> -VER (ÉqNT) <i>déplacement</i> -NOM (ÉqNT)	ÉqNT	ÉqPF	Autre point de vue, <i>exode</i> -NOM est un départ en masse (ex. <i>population migration =&gt; exode des populations</i> ).
<i>mortality</i> -NOM	<i>mortalité</i> -NOM (ÉqT) <i>décès</i> -NOM (ÉqNT)	<i>mortalité</i> -NOM (ÉqT) <i>décès</i> -NOM (ÉqNT) <b><i>mort</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	Autre point de vue selon le contexte (ex. <i>causing widespread mortality, qui a causé la mort d'un très</i>

					<i>grand nombre</i> ).
<i>nutrient</i> -NOM	<i>élément nutritif</i> -NOM (ÉqT) <i>nutriment</i> -NOM (ÉqT) <b><i>substance nutritive</i></b> -NOM (ÉqT) (9) <b><i>nutriant</i></b> -NOM (ÉqT) (5)	<i>élément nutritif</i> -NOM (ÉqT) <i>nutriment</i> -NOM (ÉqT)	ÉqT ÉqT	ÉqPF ÉqPF	<i>substance nutritive</i> -NOM et <i>nutriant</i> -NOM sont des synonymes de <i>nutriment</i> -NOM (Termium)
<i>peat</i> -NOM (18)	<i>tourbe</i> -NOM (ÉqT) <b><i>tourbeux</i></b> -ADJ (ÉqNT) (6)	<i>tourbe</i> -NOM (ÉqT)	ÉqNT	ÉqPF	<i>tourbeux</i> -ADJ est un équivalent transcatégoriel
<i>pest</i> -NOM	<i>parasite</i> -NOM (ÉqT) <i>ravageur</i> -NOM (ÉqNT) <i>nuisible</i> -ADJ (ÉqNT) <b><i>animal nuisible</i></b> -NOM (ÉqNT) (7)	<i>parasite</i> -NOM (ÉqT) <i>ravageur</i> -NOM (ÉqNT) <i>nuisible</i> -ADJ (ÉqNT)	ÉqNT	ÉqPF	<i>animal nuisible</i> -NOM est un quasi-synonyme de <i>ravageur</i> -NOM. Son sens est plus restreint, car <i>ravageur</i> inclut les insectes, les acariens, etc.
<i>preservation</i> -NOM (29)	<i>préservation</i> -NOM (ÉqT)	<i>préservation</i> -NOM (ÉqT) <b><i>préserver</i></b> -VER (ÉqNT)	ÉqNT	s.o.	<i>préserver</i> -VER équivalent transcatégoriel
<i>radiation</i> -NOM	<i>rayonnement</i> -NOM (ÉqT) <i>radiatif</i> -ADJ (ÉqNT) <i>rayon</i> -NOM (ÉqNT)	<i>rayonnement</i> -NOM (ÉqT) <i>radiatif</i> -ADJ (ÉqNT) <i>rayon</i> -NOM (ÉqNT) <b><i>radiation</i></b> -NOM (ÉqT) (7)	ÉqT	ÉqPF	<i>radiation</i> -NOM est synonyme de <i>rayonnement</i> -NOM selon Termium.
<i>region</i> -NOM	<i>région</i> -NOM (ÉqT) <i>zone</i> -NOM (ÉqNT)	<i>région</i> -NOM (ÉqT) <i>zone</i> -NOM (ÉqNT) <b><i>aire géographique</i></b> -NOM (ÉqT) (5)	ÉqT	ÉqPF	<i>aire géographique</i> -NOM est synonyme de <i>région</i> -NOM selon Termium.
<i>resilient</i> -ADJ	<i>résistant</i> -ADJ (ÉqNT) <i>résister</i> -VER (ÉqNT) <b><i>résilient</i></b> -ADJ (ÉqT)	<i>résistant</i> -ADJ (11) (ÉqNT) <i>résister</i> -VER (ÉqNT)	ÉqT	17 occ. ÉqPF	La notion de <i>résilient</i> s'apparente à celle de <i>résistant</i> (en anglais <i>resistant</i> ). Cependant, un système résilient se régénèrera plus rapidement qu'un système



					résistant, et un système résistant se maintiendra plus facilement face à une perturbation qu'un système résilient (Bioret <i>et al.</i> 2009).
<i>risk</i> -NOM	<i>risque</i> -NOM (ÉqT) <i>risquer</i> -VER (ÉqNT) <i>menace</i> -NOM (ÉqNT) <i>menacé</i> -ADJ (ÉqNT) <i>danger</i> -NOM (ÉqNT) <i>exposer</i> -VER (ÉqNT)	<i>risque</i> -NOM (ÉqT) <i>risquer</i> -VER (ÉqNT) <i>menace</i> -NOM (ÉqNT) <i>menacé</i> -ADJ (ÉqNT) <i>danger</i> -NOM (ÉqNT) <i>exposer</i> -VER (ÉqNT) <b><i>aléa</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	<i>aléa</i> -NOM est un quasi-synonyme de <i>risque</i> -NOM (Antidote).
<i>sequester</i> -VER	<i>piéger</i> -VER (39) (ÉqT) <i>piégeage</i> -NOM (5) (ÉqNT) <i>fixer</i> -VER (7) (ÉqNT) <b><i>séquestrer</i></b> -VER (7) (ÉqT)	<i>piéger</i> -VER (ÉqT) <i>piégeage</i> -NOM (ÉqNT) <i>fixer</i> -VER (ÉqNT) <b><i>capturer</i></b> -VER (ÉqT) <b><i>fixation</i></b> -NOM (ÉqNT)	ÉqT ÉqT ÉqNT	ÉqPF ÉqPF	<i>fixation</i> -NOM est un équivalent transcatégoriel. <i>séquestrer</i> -VER est un équivalent typique. Ce verbe n'est pas présent dans le CC, mais nous y trouvons comme équivalent le nom <i>séquestration</i> -NOM (voir <i>sequestration</i> -NOM)
<i>sequestration</i> -NOM	<i>piégeage</i> -NOM (ÉqT) <i>séquestration</i> -NOM (ÉqT) <i>fixation</i> -NOM (ÉqNT)	<i>piégeage</i> -NOM (ÉqT) <i>séquestration</i> -NOM (ÉqT) <i>fixation</i> -NOM (ÉqNT) <b><i>fixer</i></b> -VER (4) (ÉqNT)	ÉqNT	ÉqPF	<i>fixer</i> -VER est un équivalent transcatégoriel
<i>shoreline</i> -NOM	<i>littoral</i> -NOM (ÉqT) <i>rivage</i> -NOM (ÉqNT) <i>côte</i> -NOM (ÉqT)	<i>littoral</i> -NOM (ÉqT) <i>rivage</i> -NOM (ÉqNT) <i>côte</i> -NOM (ÉqT)	ÉqNT	s.o.	Dans le CC, <i>trait de côte</i> -NOM est employé dans des contextes semblables

	<i>côtier</i> -ADJ (ÉqNT) <i>rive</i> -NOM (ÉqNT)	<i>côtier</i> -ADJ (ÉqNT) <i>rive</i> -NOM (ÉqNT) <b><i>trait de côte</i></b> -NOM (ÉqNT)			au CP (ex. <i>érosion du littoral, des côtes</i> ou du <i>trait de côte</i> ).
<i>snowmelt</i> <sub>2</sub> -NOM	<i>fonte de la/des/de neige</i> -NOM (ÉqT)	<i>fonte de la/des/de neige</i> -NOM (ÉqT) <b><i>fonte nivale</i></b> -NOM (ÉqT) (8)	ÉqT	ÉqPF	<i>fonte nivale</i> -NOM est un synonyme de <i>fonte de la neige</i> -NOM (Termium)
<i>soil</i> -NOM	<i>sol</i> -NOM (ÉqT) <i>terre</i> -NOM (ÉqNT)	<i>sol</i> -NOM (ÉqT) <i>terre</i> -NOM (ÉqNT) <b><i>pédologique</i></b> -ADJ (ÉqNT) (4)	ÉqNT	ÉqPF	<i>pédologique</i> -NOM est un équivalent transcatégoriel
<i>specie</i> -NOM	<i>espèce</i> -NOM (ÉqT) <i>essence</i> -NOM (ÉqNT) <b><i>taxinomique</i></b> -ADJ (ÉqNT) (6)	<i>espèce</i> -NOM (ÉqT) <i>essence</i> -NOM (ÉqNT)	ÉqNT	ÉqPF	<i>taxinomique</i> -ADJ est un équivalent transcatégoriel
<i>stabilization</i> -NOM	<i>stabilisation</i> -NOM (ÉqT) <i>stabiliser</i> -VER (ÉqNT)	<i>stabilisation</i> -NOM (ÉqT) <i>stabiliser</i> -VER (ÉqNT) <b><i>maîtrise</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	dans le CC, <i>maîtrise</i> -NOM est employé dans des contextes similaires à ceux de <i>stabilisation</i> -NOM (ex. <i>maîtrise, stabilisation des émissions</i> ), mais ce terme ne peut être utilisé dans tous les contextes.
<i>storage</i> -NOM	<i>stockage</i> -NOM (ÉqT) <i>stocker</i> -VER (ÉqNT) <b><i>entreposage</i></b> -NOM (ÉqNT) (7)	<i>stockage</i> -NOM (ÉqT) <i>stocker</i> -VER (ÉqNT) <b><i>séquestration</i></b> -NOM (ÉqT)	ÉqNT ÉqT	ÉqPF s.o.	pour certains <i>séquestration</i> -NOM est synonyme de <i>stockage</i> -NOM (Brodhag et Breuils 2009). Dans le CP, <i>entreposage</i> -NOM est employé dans le sens de <i>stockage</i> (ex. <i>stockage du CO<sub>2</sub>, entreposage du CO<sub>2</sub></i> ).

<i>stress</i> -NOM	<i>stress</i> -NOM (ÉqT) <i>contrainte</i> -NOM (ÉqNT) <i>pression</i> -NOM (ÉqNT) <i>menace</i> -NOM (ÉqNT)	<i>stress</i> -NOM (ÉqT) <i>contrainte</i> -NOM (ÉqNT) <i>pression</i> -NOM (ÉqNT) <i>menace</i> -NOM (ÉqNT) <b><i>tension</i></b> -NOM (ÉqNT) <b><i>agression</i></b> -NOM (ÉqNT) <b><i>nuisance</i></b> -NOM (ÉqNT)	ÉqNT ÉqNT ÉqNT	s.o. s.o. s.o.	Autres points de vue
<i>subsidence</i> -NOM	<i>subsidence</i> -NOM (ÉqT) <b><i>affaissement de terrain</i></b> -NOM (ÉqT) (6)	<i>subsidence</i> -NOM (ÉqT)	ÉqT	ÉqPF	<i>affaissement de terrain</i> -NOM est un synonyme de <i>subsidence</i> -NOM (Wikipédia)
<i>vulnerability</i> -NOM	<i>vulnérabilité</i> -NOM (ÉqT) <i>vulnérable</i> -ADJ (ÉqNT)	<i>vulnérabilité</i> -NOM (ÉqT) <i>vulnérable</i> -ADJ (ÉqNT) <b><i>fragilité</i></b> -NOM (ÉqNT)	ÉqNT	s.o.	<i>fragilité</i> -NOM apparaît dans des contextes semblables à <i>vulnérabilité</i> -NOM (ex. <i>fragilité du territoire</i> , <i>vulnérabilité du territoire</i> ).
<i>waste</i> -NOM	<i>déchet</i> -NOM (ÉqT)	<i>déchet</i> -NOM (ÉqT) <b><i>ordure</i></b> -NOM (ÉqNT) (7)	ÉqNT	ÉqPF	<i>ordure</i> -NOM est un quasi-synonyme de <i>déchet</i> -NOM (Antidote).
<i>weed</i> -NOM	<i>mauvaise herbe</i> -NOM (ÉqT)	<i>mauvaise herbe</i> -NOM (ÉqT) <b><i>adventice</i></b> -NOM (11) (ÉqT)	ÉqT	ÉqPF	<i>adventice</i> -NOM est un synonyme de <i>mauvaise herbe</i> -NOM (GDT, agriculture, botanique, écologie)
<i>wetland</i> -NOM	<i>zone humide</i> -NOM (ÉqNT) <i>terre humide</i> -NOM (ÉqT) <i>marais</i> -NOM (ÉqNT) <b><i>milieu humide</i></b> -NOM (ÉqT) (10)	<i>zone humide</i> -NOM (ÉqNT) <i>terre humide</i> -NOM (ÉqT) <i>marais</i> -NOM (ÉqNT)	ÉqT	ÉqPF	<i>milieu humide</i> -NOM est synonyme de <i>zone humide</i> -NOM selon Termium et le GDT.
<i>wildfire</i> -NOM	<i>feu de forêt</i> -NOM (ÉqT)	<i>feu de forêt</i> -NOM (ÉqT)	ÉqT	26 occ.	<i>feu de friche</i> -NOM est

	<i>incendie de forêt</i> -NOM (ÉqT) <b>feu de friche</b> -NOM (ÉqT) (26)	<i>incendie de forêt</i> -NOM (ÉqT)			synonyme d' <i>incendie de forêt</i> -NOM selon UNTERM (biodiversité, désastres).
<i>wildlife</i> -NOM	<i>faune</i> -NOM (ÉqT) <i>faunique</i> -ADJ (ÉqNT) <i>faune sauvage</i> -NOM (ÉqNT) <b>espèce sauvage</b> -NOM (ÉqNT) (29) <b>animal sauvage</b> -NOM (ÉqT) (18)	<i>faune</i> -NOM (ÉqT) <i>faunique</i> -ADJ (ÉqNT) <i>faune sauvage</i> -NOM (ÉqNT)	ÉqNT ÉqT	29 occ. 18 occ.	Dans le sens 'ensemble des animaux qui vivent à l'état sauvage', c'est-à-dire en excluant les plantes, l'équivalent <i>animal sauvage</i> -NOM est un synonyme de <i>faune</i> -NOM.
<i>woodland</i> -NOM (23)	<b>région boisée</b> -NOM (ÉqT) (5)	<b>surface boisée</b> -NOM (ÉqT) (6)	ÉqT ÉqT	ÉqPF ÉqPF	<i>région boisée</i> -NOM et <i>surface boisée</i> -NOM sont synonymes selon IATE (environnement)

Tableau F.2. Équivalents dont la différence est marquée dans le CP (≠CP)

– Les équivalents différents analysés sont en gras dans le tableau.

Terme anglais	Équivalent CP	Équivalent CC	ÉqT/ÉqNT	ÉqF/ÉqPF	Commentaire
<i>anthropogenic</i> -ADJ	<i>anthropique</i> -ADJ (ÉqT) <i>humain</i> -ADJ (ÉqT) <b>anthropogène</b> -ADJ (ÉqNT) (11) <b>anthropogénique</b> -ADJ (ÉqNT) (4)	<i>anthropique</i> -ADJ (ÉqT) <i>humain</i> -ADJ (ÉqT) <i>homme</i> -NOM (ÉqNT)	ÉqNT ÉqNT	ÉqPF ÉqPF	<i>anthropogène</i> -ADJ est à éviter selon Termium. <i>anthropogénique</i> -ADJ est un anglicisme selon Termium. Ces termes ne figurent pas dans un grand nombre de sources européennes.
<i>reflect</i> -VER	<i>réfléchir</i> -VER (ÉqT) <b>refléter</b> -VER (ÉqNT) (6)	<i>réfléchir</i> -VER (ÉqT) <i>renvoyer</i> -VER (ÉqNT) (6)	ÉqNT	ÉqPF	<i>refléter</i> -VER est un faux ami (Van Roey <i>et al.</i> 1991).

Tableau F.3. Équivalents dont la différence est marquée dans le CC ( $\neq$ CC)

– Les équivalents différents analysés sont en gras dans le tableau

Terme anglais	Équivalent CP	Équivalent CC	ÉqT/ÉqNT	ÉqF/ÉqPF	Commentaire
<i>capture</i> -NOM	<i>piégeage</i> -NOM (ÉqT) (145) <i>piéger</i> -VER (ÉqNT) (14)	<b><i>capture</i></b> -NOM (ÉqT) (120) <i>captation</i> -NOM (ÉqNT) <i>captage</i> -NOM (ÉqT) <i>capter</i> -VER (ÉqNT) <i>capturer</i> -VER (ÉqNT) <i>piégeage</i> -NOM (ÉqT) <i>piéger</i> -VER (ÉqNT) <i>fixation</i> -NOM (ÉqNT) <i>fixer</i> -VER (ÉqNT)	ÉqT	ÉqF	<i>capture</i> -NOM est un cognat du terme anglais. Très fréquent dans le CC, mais absent du CP. <i>Captage</i> -NOM est le terme recommandé dans le <i>Journal officiel de la République française</i> .
<i>capture</i> -VER	<i>piéger</i> -VER (ÉqT) (30) <i>piégeage</i> -NOM (ÉqT) (10)	<b><i>capturer</i></b> -VER (ÉqT) (8) <i>capture</i> -NOM (ÉqNT) <i>capter</i> -VER (ÉqT) (6) <i>captage</i> -NOM (ÉqNT) <i>captation</i> -NOM (ÉqNT) <i>piéger</i> -VER (ÉqT) (6) <i>piégeage</i> -NOM (ÉqNT)	ÉqT	ÉqPF	<i>capturer</i> -VER est un cognat du terme anglais.
<i>deforestation</i> -NOM	<i>déboisement</i> -NOM (ÉqT) (145) <i>déforestation</i> -NOM (ÉqT) (13)	<b><i>déforestation</i></b> -NOM (ÉqT) (42) <i>déboisement</i> -NOM (ÉqT) (6)	ÉqT	42 occurrences	Les chiffres montrent que le cognat <i>deforestation</i> -NOM est plus fréquent que <i>déboisement</i> -NOM dans le CC, alors que c'est l'inverse dans le CP. Selon Termium

					et IATE ces deux équivalents sont synonymes. Le <i>Dictionnaire de la foresterie</i> (Côté 2003), quant à lui, n'inclut pas dans sa nomenclature les équivalents <i>déforestation</i> et <i>reforestation</i> , seuls <i>déboisement</i> et <i>reboisement</i> y figurent. Au dire de l'auteur, ces deux derniers termes sont ceux qui correspondent le mieux aux termes anglais <sup>119</sup> .
<i>reforestation</i> -NOM (85)	<i>reboisement</i> -NOM (ÉqT)	<i>reboisement</i> -NOM (ÉqT) (19) <b><i>reforestation</i></b> -NOM (ÉqT) (6)	ÉqT	ÉqPF	<i>reforestation</i> -NOM est un cognat du terme anglais. Selon (UNTERM, Termium, IATE et le GDT), <i>reforestation</i> -NOM et <i>reboisement</i> -NOM sont synonymes.
<i>global</i> -ADJ	<i>mondial</i> -ADJ (ÉqT) (953) (TleCP : 1505) <i>planétaire</i> -ADJ	<b><i>global</i></b> -ADJ (ÉqNT) (844) <i>mondial</i> -ADJ (ÉqT) (836) <i>planétaire</i> -ADJ (ÉqT) (136)	ÉqNT	ÉqF	<i>global</i> -ADJ ('relatif au globe terrestre') est un anglicisme selon Van Roey <i>et al.</i>

<sup>119</sup> « Lorsque nous avons préparé l'édition 2003 du *Dictionnaire de la foresterie*, nous avons consulté des centaines de références et avons retenu celles qui correspondaient le mieux aux termes généralement utilisés en sciences forestières. À notre avis, les termes "déboisement" et "reboisement" sont ceux qui correspondent le mieux aux termes anglais "deforestation" et "reforestation". Ce sont d'ailleurs ceux qui sont utilisés dans le langage courant (Marc Côté, ing. f., M.Sc., responsable du *Dictionnaire de la foresterie*) ».

	(ÉqNT) (231) <i>global</i> -ADJ (ÉqNT) (154) (TleCP : 415) <i>globe</i> -NOM (ÉqNT) (88) <i>planète</i> -NOM (ÉqNT) (54) <i>monde</i> -NOM (ÉqNT) (41) <i>international</i> -ADJ (ÉqNT) (22)	<i>globe</i> -NOM (ÉqNT) (101) <i>planète</i> -NOM (ÉqNT) <i>international</i> -ADJ (ÉqNT) <i>monde</i> -NOM (ÉqNT)			(1991) et Termium.
<i>mitigate</i> -VER	<i>atténuer</i> -VER (ÉqT) <i>atténuation</i> -NOM (ÉqNT)	<i>atténuer</i> -VER (ÉqT) <b><i>mitiger</i></b> -VER (ÉqNT) (4) <i>atténuation</i> -NOM (ÉqNT) <i>mitigation</i> -NOM (ÉqNT)	ÉqNT	ÉqPF	<i>mitiger</i> -VER est un anglicisme, car c'est un équivalent morphologiquement relié à <i>mitiger</i> -NOM
<i>mitigation</i> -NOM	<i>atténuation</i> -NOM (ÉqT) (684) <i>atténuer</i> -VER (ÉqNT)	<i>atténuation</i> -NOM (ÉqT) <b><i>mitigation</i></b> -NOM (ÉqNT) (35) <i>atténuer</i> -VER (ÉqNT) <i>mitiger</i> -VER (ÉqNT)	ÉqNT	35 occurrences	<i>mitigation</i> -NOM est un anglicisme (Termium)
<i>monitoring</i> -NOM	<i>surveillance</i> -NOM (ÉqT) <i>suivi</i> -NOM (ÉqNT) <i>contrôle</i> -NOM (ÉqNT) <i>surveiller</i> -VER (ÉqNT) <i>observation</i> -NOM (ÉqNT) <i>évaluation</i> -NOM (ÉqNT)	<i>surveillance</i> -NOM (ÉqT) <i>suivi</i> -NOM (ÉqNT) <i>contrôle</i> -NOM (ÉqNT) <i>surveiller</i> -VER (ÉqNT) <b><i>monitoring</i></b> -NOM (ÉqNT) (4)	ÉqNT	ÉqPF	<i>monitoring</i> -NOM est un anglicisme (Termium et FranceTerme).

<i>offshore</i> -ADJ (40)	<i>en mer</i> -LOC (ÉqT) <i>au/du large</i> -LOC (ÉqT)	<i>en mer</i> -LOC (ÉqT) <i>au/du large</i> -LOC (ÉqT) <b><i>offshore</i></b> -ADJ (ÉqNT) (31)	ÉqNT	31 occurrences	<i>offshore</i> -ADJ est un anglicisme (Termium et FranceTerme). recommandation officielle <i>en mer</i> (Petit Robert).
<i>permafrost</i> -NOM	<i>pergélisol</i> -NOM (ÉqT) (308) <i>permafrost</i> -NOM (ÉqT) (16) <i>pergélisolé</i> -ADJ (ÉqNT) (6)	<i>pergélisol</i> -NOM (ÉqT) (96) <b><i>permafrost</i></b> -NOM (ÉqT) (87)	ÉqT	ÉqF	<i>permafrost</i> -NOM est un anglicisme selon Termium et un équivalent admis selon FranceTerme. Nous considérons cet équivalent comme un cognat.