

Exploitation quantitative de corpus de traductions annotés selon la typologie d'erreurs pour améliorer les méthodes d'enseignement de la traduction spécialisée

Natalie Kübler, Alexandra Mestivier, Mojca Pecman, Maria Zimina

CLILLAC- ARP EA 3967, Université Paris Diderot, Paris, France

Abstract

This study focuses on quantitative analysis of translations annotated according to an error typology developed in the MeLLANGE project. The purpose of the study is to improve our current methods of specialised translation (ST) teaching. We conduct a comparative analysis of translations carried out under different conditions, with and without the access to specialised corpora. The analysis provides the proof of the usefulness of corpora for ST. Furthermore, we analyse the most salient errors and morpho-syntactic patterns characterised by these errors, with a view of making methodological suggestions to help students avoid them. This experimentation opens up new possibilities for studying interlinguistic meaning transfer which, in the domain of LSPs and in an era of ever increasing needs in multilingual documents, remains one crucial and complex language performance.

Résumé

L'étude présente l'analyse quantitative de traductions annotées selon la typologie d'erreurs développée dans le cadre du projet MeLLANGE, en vue de l'amélioration des méthodologies d'enseignement de la traduction spécialisée (TS). En comparant les traductions réalisées dans des conditions différentes, avec et sans corpus, notre étude confirme la pertinence de l'utilisation des corpus en TS. Nous distinguons les catégories d'erreurs les plus saillantes et les schémas morphosyntaxiques qui les caractérisent afin de pouvoir émettre des propositions méthodologiques pour améliorer l'enseignement de la TS. Cette expérimentation ouvre des perspectives pour étudier le processus de transfert du sens qui demeure une des activités langagières des plus complexes, notamment face aux besoins croissants en documents multilingues touchant les discours spécialisés.

Mots clés : corpus annotés, enseignement de la Traduction Spécialisée (TS), langues de spécialité, méthode des spécificités, problèmes de traduction, textométrie, typologie d'erreurs de traduction

1. Introduction

Malgré les avancées notables dans le domaine de la traduction, le processus de transfert du sens demeure l'une des activités langagières les plus complexes, mobilisant des compétences linguistiques et cognitives très variées. Si, de nos jours, de nombreux travaux ont permis de baliser ces questions (e.g. Baker, 1998 ; Froeliger, 2011 ; Ladmiral, 1994, 2014), il reste que la traduction appelle des méthodes efficaces, ainsi que des solutions de plus en plus adaptées à la multiplication des documents multilingues, notamment en ce qui concerne les discours spécialisés. La recherche expérimentale offre la possibilité d'observer de manière systématique les stratégies employées par les traducteurs et les problèmes qui en découlent.

Traditionnellement, les travaux centrés sur la pratique de la traduction spécialisée (TS) et son enseignement explorent les aspects très divers de textes produits par les apprenants (Baker, 1998, 1999 ; Frankenberg-Garcia, 2009 ; Froeliger, 2011 ; Olohan, 2004 ; Mauranen, 2007). Les travaux sur l'utilisation de corpus lors du processus de traduction sont toutefois relativement récents. Par exemple, Bowker et Bennison (2003) exposent les résultats auxquels

ils sont parvenus en développant un outil permettant aux étudiants d'évaluer leurs propres traductions. Pearson (2003) explore l'utilisation d'un petit corpus parallèle dans les cours de TS permettant aux étudiants de comparer leurs traductions aux textes produits par des traducteurs professionnels afin d'attirer leur attention sur les stratégies traductionnelles employées. Loock *et al.* (2014) proposent d'exploiter les résultats de la recherche en traductologie de corpus à des fins pédagogiques. Dans le cadre du projet MeLLANGE¹, Castagnoli *et al.* (2011) envisagent des parcours éveillant la conscience linguistique des étudiants face aux stratégies de traduction, en s'appuyant sur des exercices de manipulation de corpus de traductions multilingues annotées.

2. Enseignement de la TS à l'aide de corpus : méthodologie expérimentée à l'Université Paris Diderot-Paris 7

Depuis 2001, une méthodologie d'enseignement de la TS à l'aide des outils de la linguistique de corpus est pratiquée à l'Université Paris Diderot-Paris 7, par l'équipe travaillant sur les langues de spécialité (Kübler, 2003, 2011), dans le cadre du Master 1 en ILTS (Industrie des Langues et Traduction Spécialisée)². Elle s'appuie sur la combinaison de l'approche de la linguistique de corpus, de l'analyse terminologique, de la création de ressources lexicales et sur la collaboration avec les experts du domaine, afin d'améliorer la compréhension et la traduction des textes spécialisés. Depuis 2013-2014, cette méthodologie comprend la collecte systématique de toutes les productions, avec analyse contrastive des traductions réalisées avec apport des corpus et sans accès aux corpus (Kübler *et al.*, 2015). Les comparaisons sont réalisées avec apport d'outils d'analyse textométrique (Fleury et Zimina, 2014).

Notre travail vise à contribuer au développement des méthodes d'enseignement de la TS en exploitant les avantages de l'analyse quantitative de corpus annotés selon une typologie d'erreurs, avec pour objectif l'optimisation de la formation en traduction. Après la présentation du cadre de réalisation de notre expérience, nous montrons les caractéristiques du corpus d'étude et la procédure d'annotation des traductions selon la typologie d'erreurs MeLLANGE (Castagnoli *et al.*, 2011). Nous proposons ensuite une procédure de comparaison des traductions produites avec et sans accès aux corpus, avec recours à la méthode des *spécificités* (Lebart et Salem, 1994). Les résultats sont étudiés afin de cerner l'apport des corpus à la pratique de la traduction, les types d'erreurs les plus saillants et les figements morpho-syntaxiques spécifiques qui posent des difficultés aux apprenants. Finalement, notre étude aboutira à des propositions méthodologiques qui visent à améliorer la pratique d'enseignement de la TS.

3. Constitution du corpus

3.1 Cadre méthodologique et ressources de travail

Les traductions analysées dans cette étude sont effectuées par les étudiants du Master 1 ILTS de l'UFR d'Études Interculturelles de Langues Appliquées (EILA), Université Paris Diderot, dans le cadre des cours en collaboration avec les étudiants et les enseignants de l'UFR des Sciences de la Terre, de l'Environnement et des Planètes (STEP) de la même université.

¹ MeLLANGE (Multilingual eLearning in Language Engineering) : <http://mellange.eila.univ-paris-diderot.fr>

² Master Langues étrangères appliquées (LEA), Spécialité ILTS : <http://www.eila.univ-paris-diderot.fr/enseignement/lea/master/index>

Chaque année, entre 10 et 20 articles scientifiques en sciences de la Terre, rédigés en anglais, sont sélectionnés pour la traduction vers le français. Pour préparer la traduction, les étudiants réalisent des analyses terminologiques et construisent chacun plusieurs entrées dictionnairiques en compilant les ressources dans la base terminologique et phraséologique ARTES (Aide à la Rédaction de TExtes Scientifiques)³ (Pecman et Kübler, 2011). Cette base multilingue multi-domaine est disponible en ligne et permet une exploitation collaborative des ressources traductionnelles. En parallèle, les apprentis traducteurs construisent ensemble un corpus comparable du domaine qui est compilé sur *IMS Open Corpus Workbench (CWB)*⁴ afin de permettre des recherches contextuelles avancées. En 2014-2015, par exemple, le corpus anglais comptait environ 12 millions de mots et le corpus français près de 60 millions. La disparité de taille entre les deux corpus s'explique par la difficulté à collecter des articles en français (due à l'hégémonie de l'anglais dans les publications scientifiques), qui se voit compensée par la collecte de thèses.

3.2 Annotation des traductions selon la typologie d'erreurs MeLLANGE

Le cadre méthodologique d'enseignement de la TS a été conçu afin de permettre une évaluation systématique des stratégies de traduction employées par les apprenants. Il laisse apparaître les principales difficultés afin de mieux orienter l'enseignement et permet d'obtenir une mesure quantitative de l'efficacité de l'utilisation des corpus dans la TS. Chaque texte à traduire (extrait d'un article sélectionné pour le test comportant 500 mots) est d'abord segmenté en deux parties (200 et 300 mots) afin que les traductions soient réalisées en deux temps dans des conditions différentes : dans un premier temps, en utilisant un ensemble limité de ressources dictionnairiques, sans corpus et, dans un deuxième temps, avec l'aide du corpus comparable constitué par les étudiants (cf. *Figure 1*).



Figure 1 : Les étapes de production et d'annotation de traductions

Les traductions ainsi réalisées sont ensuite déposées sur un serveur *Brat*⁵ et annotées selon la typologie d'erreurs développée dans le cadre du projet MeLLANGE (Castagnoli et al., 2011)⁶. Le schéma d'annotation est construit sur la distinction fondamentale entre les erreurs liées au transfert de contenu et les erreurs de langue (cf. *Figure 2*). Ces deux catégories principales sont divisées en sous-catégories telles que « Intrusion de la langue source » ou « Terminologie et lexique », regroupant des erreurs plus spécifiques, comme : « Trop littéral »

³ Interface de consultation de la base ARTES : <https://artes.eila.univ-paris-diderot.fr>

⁴ *IMS Open Corpus Workbench (CWB)* : <http://cwb.sourceforge.net>

⁵ Serveur d'annotation *Brat* : <http://brat.nlplab.org>

⁶ La typologie MeLLANGE a été créée dans le but de décrire les difficultés de traduction et non pas pour transmettre des jugements de valeur : elle ne permet pas d'encoder la « gravité » d'une erreur.

ou encore « Collocation incorrecte ». Cette arborescence inclut également, à plusieurs niveaux, des catégories de type « Défini par l'utilisateur ». Ce type de marquage peut être employé soit pour signaler un type d'erreur qui n'a pas été prévu dans le schéma d'annotation, et donc mener à une révision du schéma, soit pour annoter des solutions de traduction particulièrement réussies.

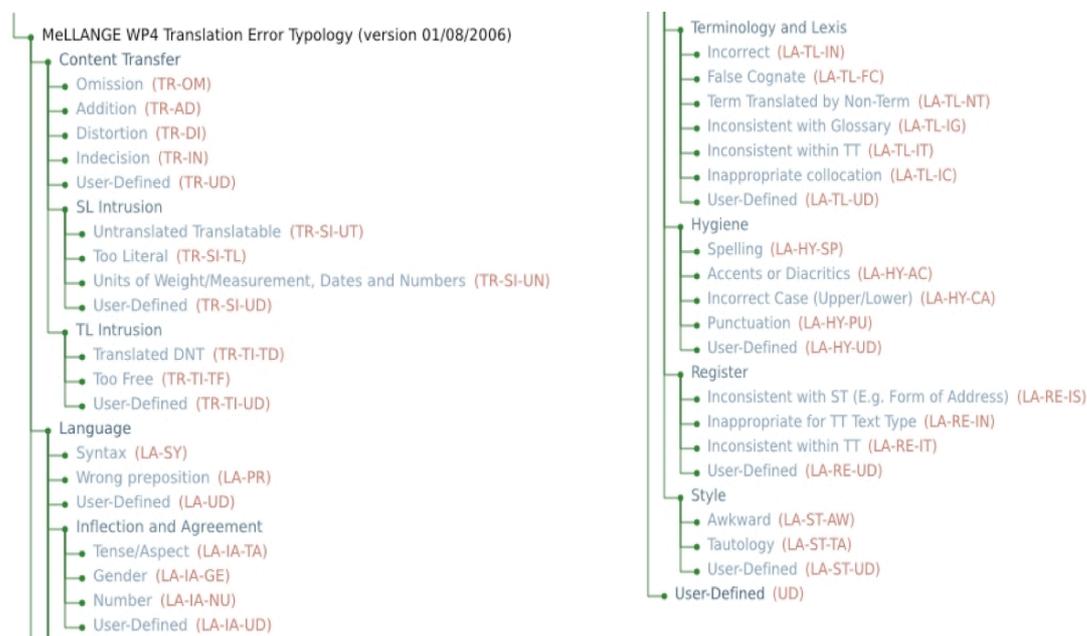


Figure 2 : Typologie d'erreurs MeLLANGE utilisée pour l'annotation des traductions

Au cours du processus d'annotation, chaque erreur est associée à des mots, syntagmes ou phrases correspondants et un commentaire est éventuellement ajouté à l'attention des étudiants (cf. Figures 3-4).

3.3 Corpus de travail *ER-TRAD-SP1* et *ER-TRAD-SP2*

La présente étude exploite les traductions réalisées en 2014-2015. Il s'agit de 55 extraits de 14 articles scientifiques (37 324 occurrences au total). Ce corpus est subdivisé en deux sous-corpus : *ER-TRAD-SP1* (15 311 occurrences) constitué de traductions réalisées sans accès aux corpus, et *ER-TRAD-SP2* (22 013 occurrences) constitué de traductions réalisées avec l'aide du corpus.

Les traductions portent sur des extraits d'articles scientifiques avec une très haute densité terminologique et le registre de langue caractéristique du discours scientifique. Les problèmes de traduction qu'affrontent les apprentis traducteurs sont nombreux et variés. L'annotation des erreurs selon la typologie MeLLANGE permet toutefois une catégorisation fine des différents problèmes rencontrés dans ce type de discours.⁷ Au total, le sous-corpus *ER-TRAD-SP1* compte 886 annotations ; le sous-corpus *ER-TRAD-SP2* compte 893 annotations.

⁷ L'annotation est effectuée par des enseignants-chercheurs qui appliquent rigoureusement une méthodologie commune de marquage d'erreurs (l'équipe étant aussi à l'origine du projet MeLLANGE).

4.2 Analyse des surfaces d'erreurs dans les deux types de productions

4.2.1 Longueurs moyennes d'empans d'annotations

L'analyse de la variabilité des empans d'annotations dans *ER-TRAD-SP1* et *ER-TRAD-SP2* n'a pas révélé de différence significative entre les deux parties (l'annotation étant réalisée par les mêmes annotateurs dans des conditions similaires). Les résultats présentés sur le *Tableau 1* pour quelques types d'erreurs montrent que l'écart ne dépasse pas une occurrence.

Type d'erreur	Longueur d'empans d'annotations (moyenne)	
	<i>ER-TRAD-SP1</i>	<i>ER-TRAD-SP2</i>
Choix incorrect	1,73 occ.	1,76 occ.
Collocation incorrecte	2,27 occ.	1,37 occ.
Mauvaise préposition	1,48 occ.	1,28 occ.
Formulation maladroite	4,97 occ.	3,96 occ.
Distorsion	3,95 occ.	3,61 occ.

Tableau 1 : Longueur moyenne des empans d'annotations dans les deux sous-corpus (exemples)

4.2.2 Fréquences et spécificités d'erreurs de traduction

Le calcul des *spécificités* montre que les erreurs sont nettement sous-représentées dans *ER-TRAD-SP2* qui correspond aux productions réalisées avec recours au corpus spécialisé (cf. *Tableau 2*).⁹

	<i>ER-TRAD-SP1</i>		<i>ER-TRAD-SP2</i>	
	Fréquence	Spécificité	Fréquence	Spécificité
Surface d'erreurs	2 277 occ.	48	2 150 occ.	-49

Tableau 2 : Surface d'erreurs dans ER-TRAD-SP1 et ER-TRAD-SP2

La *Figure 5* montre que l'apport du corpus se révèle particulièrement pertinent pour surmonter les erreurs de terminologie (type « Terme traduit par non terme »). Comme les dictionnaires spécialisés présentent des lacunes, les corpus sont souvent la seule ressource qui permette d'identifier les équivalents des termes, notamment lorsque les textes spécialisés à traduire sont récents¹⁰. La *Figure 5* montre par ailleurs que les corpus aident à résoudre la majorité des problèmes bien connus en traduction : les erreurs liées au choix des temps et des aspects verbaux, à l'utilisation des prépositions, au non-respect du style, au choix des collocations, etc. Les traductions effectuées à l'aide des corpus sont par ailleurs moins littérales, elles respectent mieux le style requis et sont plus adaptées au type de texte cible.

⁹ Pour conduire cette analyse, une annotation supplémentaire a été rajoutée à la base textométrique pour indiquer la présence/absence d'erreur (sans préciser son type). Cette fonction est disponible dans *Le Trameur*.

¹⁰ Les textes récents reflètent un discours et une terminologie caractéristique des axes de travail émergents. Ils regorgent de termes récents ou de néologismes, pour lesquels on ne trouve pas d'entrées dans les dictionnaires.

Item (Fq > 10)	Fq.	Sp.	Fq.	Sp.
	ER-TRAD-SP1	R-TRAD-SP1	ER-TRAD-SP2	ER-TRAD-SP2 ▲
Terme-traduit-par-non-terme_LA-TL-NT	150	33	25	-35
Collocation-incorrecte_LA-TL-IC	93	11	36	-14
Temps-aspect_LA-IA-TA	40	12	4	-12
Style	33	13	1	-12
Rajout_TR-AD	42	12	5	-12
Type-annotateur_UD	187	11	134	-10
Trop-litterale_TR-SI-TL	286	7	277	-7
Mauvaise-preposition_LA-PR	54	8	23	-7
Inadapte-au-type-texte-cible_LA-RE-IT	100	5	83	-4
Formulation-maladroite_LA-ST-AW	366	4	425	-3
Nombre_LA-IA-NU	38	3	31	-3
Choix-incorrecte_LA-TL-IN	128	0	152	-2
Type-annotateur_TR-UD	16	2	11	-2
Distorsion_TR-DI	395	2	506	-2
Trop-libre_TR-TI-TF	26	2	24	-2
Omission_TR-OM	98	2	111	-2
Orthographe_LA-HY-SP	19	0	31	0
Ponctuation_LA-HY-PU	7	0	11	0
Genre_LA-IA-GE	16	0	18	0
Transfert-contenu	51	0	84	0
Terminologie-lexique	14	0	15	0
Syntaxe_LA-SY	70	0	86	0
Type-annotateur_LA-TL-UD	1	-3	11	3
Indecision_TR-IN	4	-4	28	5

Figure 5 : Fréquences et spécificités des erreurs de traduction dans ER-TRAD-SP1 et ER-TRAD-SP2

Il est toutefois intéressant de constater que certains types d'erreurs augmentent légèrement dans les traductions réalisées avec l'aide de corpus. Toutefois, cette diminution de qualité ne touche qu'une partie d'erreurs et reste très modérée. Les indécisions sont ainsi plus récurrentes dans les productions réalisées avec corpus. Les raisons exactes de l'augmentation de ce type d'erreurs sont difficiles à expliquer dans les conditions de réalisation de nos tests. Certaines hypothèses de travail nous amènent à envisager que les corpus induisent un sentiment de trop grande confiance qui aboutit à des interprétations trop libres du texte source. Une autre explication est liée au manque de temps : les deux tests de traduction ont la même durée (2 heures), ce qui pourrait être à l'origine d'une légère augmentation de certaines erreurs dans le cas d'utilisation intensive d'outils d'interrogation de corpus.

5. Analyse des catégories grammaticales les plus marquées par les erreurs

Les catégories morpho-syntaxiques qui reviennent le plus souvent dans les annotations d'erreurs sont les noms, suivis des prépositions, des adjectifs, des verbes au présent et des déterminants. Les fréquences relatives représentées sur la *Figure 6* montrent que les noms posent des problèmes récurrents en TS. En effet, l'une des principales difficultés de la TS est liée à la forte concentration de termes. Bien que les corpus permettent de réduire le nombre d'erreurs, cette catégorie reste néanmoins la plus marquée par les erreurs dans les productions réalisées à l'aide de corpus.

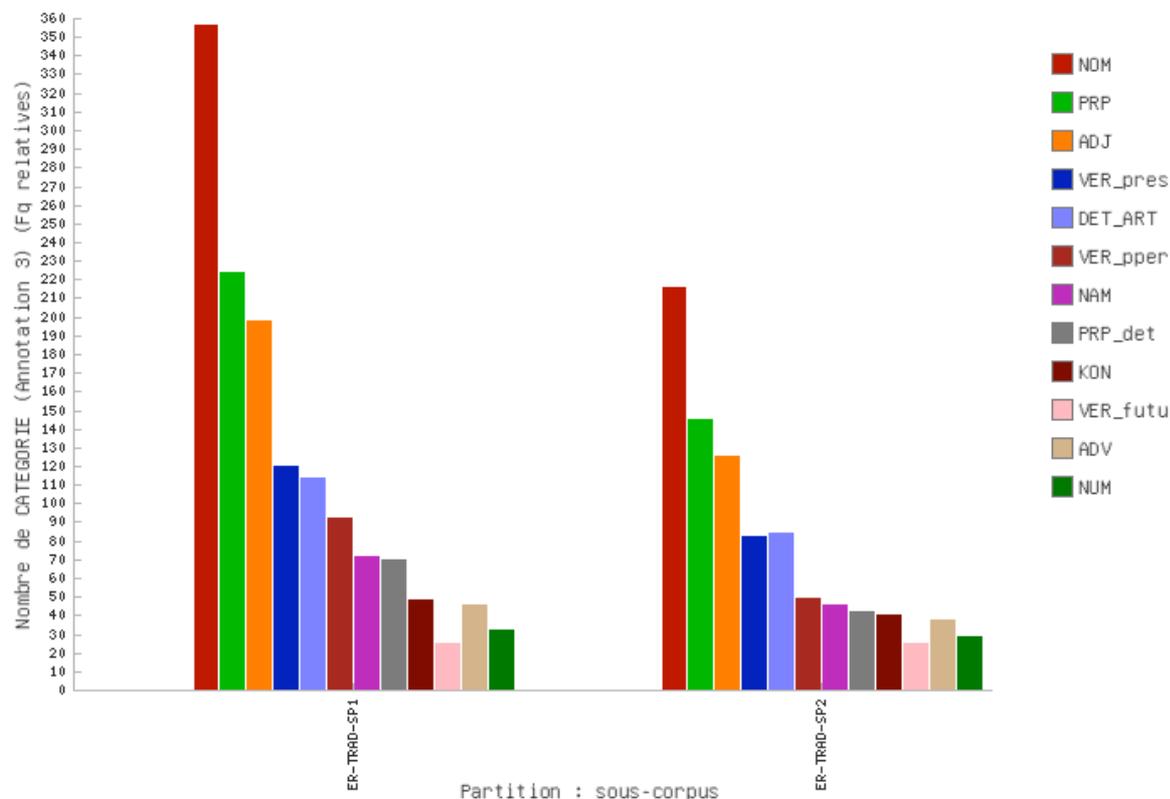


Figure 6 : Les fréquences relatives des catégories grammaticales le plus marquées par les erreurs : productions ER-TRAD-SP1 et ER-TRAD-SP2

Afin de comprendre les difficultés des étudiants face aux choix traductionnels pour les noms anglais, nous avons relevé les erreurs de traduction les plus fréquentes affectant les noms (Freq >10) qui persistent après l'introduction des corpus dans les tests (cf. *Tableau 3*). On constate que les erreurs liées à des problèmes de formulation et de sens sont en tête de liste, suivies par des erreurs de terminologie, de collocation incorrecte associée à un mot ou terme, et des erreurs de contresens.

Type d'erreur affectant la catégorie NOM	Fréquence totale ER-TRAD-SP2
Formulation-maladroite LA-ST-AW	52
Distorsion TR-DI	52
Terme-traduit-par-non-terme LA-TL-NT	50
Trop-littérale TR-SI-TL	31
Type-annotateur UD	28
Choix-incorrect LA-TL-IN	25
Inadapte-au-type-texte-cible LA-RE-IT	19
Omission TR-OM	17
Collocation-incorrecte LA-TL-IC	13
Transfert-contenu	11

Tableau 3 : Les erreurs de traduction les plus fréquentes affectant les noms (ER-TRAD-SP2)

Cette analyse nous aide à comprendre que les noms anglais en langue de spécialité soulèvent de grandes difficultés dans la traduction du point de vue de leur structure, sens, ou encore de leur statut de terme. Les retours aux contextes permettent de relever les noms qui ont été fréquemment annotés comme erreurs :

- le terme « fonte » (annoté comme erreur 12 fois dans *ER-TRAD-SP1* et 6 fois dans *ER-TRAD-SP2*) qui a été proposé comme traduction du terme anglais *melt* au lieu de « fusion », et
- le terme « ratio » (annoté 18 fois comme erreur dans *ER-TRAD-SP1* et 30 fois dans *ER-TRAD-SP2*) qui constitue une traduction par emprunt à l'anglais, alors que le terme « rapport » est disponible en français et utilisé par les experts francophones.

6. Perspectives

L'analyse quantitative des erreurs de traduction nous permet d'envisager une amélioration de notre méthode d'enseignement de la TS par un retour à la langue source pour des catégories morpho-syntaxiques les plus problématiques et la création d'exercices en amont pour les affronter. Par exemple, afin de comprendre l'augmentation des annotations d'erreurs dans le sous-corpus *ER-TRAD-SP2* pour le terme « ratio », nous pouvons relever les contextes d'emploi de ce terme et de son équivalent dans la LS, comme :

- a. Exemple de phrase source : *The identical 182W/184W ratios of the lunar and terrestrial mantles provide a key constraint on the formation of the Moon.* (Source : Touboul *et al.*, 2007. Late formation and prolonged differentiation of the Moon inferred from W isotopes in lunar metals. *Nature*, 450: 1206-1209.)
- b. Exemple de traduction par l'étudiant : « Les **ratios** identiques de 182W/184W des manteaux lunaires et terrestres sont des contraintes déterminantes pour l'évaluation de la formation de la Lune. »

Ce bi-contexte permet d'interpréter les difficultés sous-jacentes pour les étudiants lorsqu'il s'agit de traduire correctement l'unité *ratio*. Dans un premier temps, on peut supposer que les étudiants ont l'impression de maîtriser les termes comme *melt* et *ratio* suffisamment bien pour ne pas chercher à les vérifier dans le corpus comparable en langue cible afin de les traduire correctement. Une interrogation de celui-ci permet assez facilement de découvrir que les équivalents français de ces termes sont respectivement « fusion »¹¹ et « rapport »¹². Dans un deuxième temps, on peut également noter que ces termes peuvent apparaître dans des syntagmes nominaux très complexes.

Dans l'exemple *a*, on peut observer la capacité de la langue anglaise à faire figurer en position sujet des syntagmes nominaux comportant des expansions nominales étendues (cf. *the identical 182W/184W ratios of the lunar and terrestrial mantles*). Or le français n'admet que difficilement une telle structure syntaxique (cf. « les ratios identiques de 182W/184W des manteaux lunaires et terrestres »). Une solution possible face à ce problème de traduction de l'anglais vers le français serait de faire passer l'information essentielle contenue dans le thème de la phrase source en rhème dans la phrase cible, ou de la mettre en tête du GN :

¹¹ Il s'agit de l'équivalent en tectonique des plaques ou volcanologie ; en glaciologie, *melt* se traduira par *fonte*.

¹² L'emprunt à l'anglais *ratio* est certes très souvent utilisé en sciences, mais en sciences de la terre, le terme « rapport » est disponible et privilégié par les experts.

- c. Suggestions de traduction : « **Les rapports 182W/184W des manteaux terrestre et lunaire étant identiques**, ils fournissent une contrainte essentielle pour l'évaluation de la formation de la Lune. » Ou « **La similitude des rapports 182W/184W pour les manteaux terrestre et lunaire** apporte une contrainte essentielle dans la compréhension de la formation de la Lune. »

On constate que l'anglais possède une grande plasticité dans la formation des noms composés complexes, ce qui aboutit souvent dans les traductions des apprenants à des paraphrases longues et des tournures lourdes. En langues de spécialités, les noms et les groupes nominaux sont porteurs de connaissances, ils renvoient à des concepts spécialisés et leur structure interne en anglais est très différente de celle du français. Les résultats des analyses quantitatives sur corpus nous ont amenées à intégrer dans nos enseignements des exercices qui ciblent les problématiques que pose la traduction des termes composés et complexes.

7. Conclusion

Dans cette étude, nous avons procédé à l'exploration quantitative de corpus de traductions annotés selon la typologie d'erreurs MeLLANGE. Les analyses quantitatives nous ont permis d'identifier les erreurs saillantes des apprenants en traduction spécialisée (TS) et d'évaluer la méthodologie d'enseignement mise en place à l'Université Paris Diderot dans le cadre du Master ILTS. Les résultats de nos expérimentations montrent que la linguistique de corpus apporte incontestablement une approche, des ressources et des outils qui permettent d'aboutir à des traductions de plus grande qualité. Les corpus de traductions annotés se révèlent aussi très utiles pour évaluer les méthodologies de l'enseignement de la TS et pour développer des modules spécifiques qui ciblent les difficultés récurrentes des apprenants.

Par ailleurs, l'analyse textométrique nous a permis de distinguer les catégories d'erreurs les plus saillantes et les schémas morphosyntaxiques qui les caractérisent en contexte, à savoir les problèmes touchant les termes composés et complexes, mais également les collocations, les prépositions, les verbes, etc. Les analyses faites sur les productions des apprenants de ces deux dernières années nous ont aussi amenées à intégrer à l'enseignement de la TS une approche plus ciblée des problèmes posés par la traduction des termes spécialisés et des GN complexes. Nous envisageons d'approfondir cette problématique afin de trouver des solutions concrètes pour nos futurs traducteurs. Sur ce plan, une analyse des corpus comparables permettrait de relever les moyens dont dispose le français pour la transmission de ces éléments typiques des discours scientifiques, souvent condensés en anglais dans le GN.

Au fil de nos expériences sur corpus de traductions, nous avons constaté la pertinence de l'utilisation du schéma d'annotation d'erreurs MeLLANGE. Cependant, en procédant à l'analyse des commentaires associés aux annotations, nous avons constaté que ce schéma pourrait être enrichi par des catégories d'erreurs à visée pédagogique. Par exemple, l'analyse des catégories grammaticales les plus marquées par les erreurs a permis de relever de nombreux cas de « Distorsions » dus aux erreurs dans l'analyse des GN particulièrement complexes en anglais. L'effet d'une telle erreur d'analyse est évidemment une distorsion du sens, mais l'identification erronée de la tête d'un groupe nominal est un type d'erreur récurrent que nous avons décidé d'aborder dans le cadre de nos enseignements. On pourrait ainsi être amené à mettre en place une typologie d'annotation plus fine pour en tenir compte. L'annotation des productions des étudiants et l'analyse quantitative des erreurs offrent l'avantage de permettre un retour efficace sur notre propre méthodologie d'enseignement et d'envisager une amélioration de celle-ci de manière incrémentale.

Références bibliographiques

- Baker M. (1998). Réexplorer la langue de la traduction : une approche par corpus. *Meta : journal des traducteurs / Meta: Translators' Journal*, 43(4) : 480-485.
- Baker M. (1999). The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators. *International Journal of Corpus Linguistics*, 4(2): 281-298.
- Bowker L. and Bennison P. (2003). Student Translation Archive and Student Translation Tracking System. Design, Development and Application. In Zanettin F., Bernardini S. and Stewart D. editors, *Corpora in translator education*. Manchester: St. Jerome Publishing.
- Castagnoli, S., Ciobanu D., Kübler N., Kunz K. and Volanschi A. (2011). Designing a Learner Translator Corpus for Training Purposes. In Kübler N. editor, *Corpora, Language, Teaching, and Resources : From Theory to Practice*. Bern: Peter Lang.
- Fleury S. and Zimina M. (2014). Trameur: A Framework for Annotated Text Corpora Exploration. *Proc. of COLING 2014 (the 25th International Conference on Computational Linguistics: System Demonstrations)*, August 2014, Dublin, Ireland, pages 57-61.
- Frankenberg-Garcia A. (2009). Are translations longer than source texts? A corpus-based study of explicitation. In Beeby A., Rodríguez P., Sánchez-Gijón P. editors, *Corpus use and translating*, Amsterdam and Philadelphia: John Benjamins.
- Froeliger N. (2013). *Les Noces de l'analogique et du numérique - De la traduction pragmatique*. Paris : Les Belles lettres (collection Traductologiques).
- Kübler N. (2003). Corpora and LSP translation. In Zanettin F., Bernardini S. and Stewart D. editors, *Corpora in Translator Education*. Manchester: St Jerome Publishing.
- Kübler, N. (2011). Working with different corpora in translation teaching. In Frankenberg-Garcia A., Flowerdew L., and Aston G. editors, *New Trends in Corpora and Language Learning*. London: Continuum.
- Kübler N., Mestivier A. et Pecman M. (2015). Etude sur l'utilisation des corpus dans l'enseignement de la terminologie et de la traduction spécialisée. *Terrains de recherche en linguistique appliquée (TRELA)*. Conférence internationale, CLILLAC-ARP, Université Paris Diderot, Paris 8-10 juillet.
- Ladmiral J.-R. (1994). *Traduire : théorèmes pour la traduction*. Paris : Payot.
- Ladmiral, J.-R. (2014). *Sourcier ou cibliste*. Paris : Les Belles lettres (collection Traductologiques).
- Lebart L. et Salem A. (1994). *Statistique textuelle*. Dunod.
- Looock R., Mariaule M. et Oster C. (2014). Traductologie de corpus et qualité : étude de cas, Actes du colloque Tralogy II, Paris, CNRS, 17-18 janvier 2013.
- Mauranen A. (2007). Universal tendencies in translation. In Rogers M. and Anderman G. editors, *Incorporating corpora: The linguist and the translator*. Clevedon: Multilingual Matters.
- Olohan M. (2004). *Introducing corpora in translation studies*. London and New York: Routledge.
- Pecman M. and Kübler N. (2011). ARTES: an online lexical database for research and teaching in specialized translation and communication. *Proc. from International Workshop on Lexical Resources (WoLeR) 2011 at ESSLLI*. August 1-5, Ljubljana, Slovenia, pages 87-93.
- Pearson J. (2003). Using parallel texts in the Translator Training Environment. In Zanettin F., Bernardini S. and Stewart D. editors, *Corpora in Translator Education*. Manchester: St Jerome Publishing.
- Zimina M. et Fleury S. (2015). Ressources textuelles incrémentales pour la modélisation des interactions linguistiques multiples. *Terrains de recherche en linguistique appliquée (TRELA)*. Conférence internationale, CLILLAC-ARP, Université Paris Diderot, Paris 8-10 juillet.