



**HAL**  
open science

# Procédés de traduction humaine mis en évidence dans les ruptures lexicogrammicales avec la traduction d'un système de traduction automatique probabiliste

Éric Poirier

► **To cite this version:**

Éric Poirier. Procédés de traduction humaine mis en évidence dans les ruptures lexicogrammicales avec la traduction d'un système de traduction automatique probabiliste. 7èmes Journées Internationales de la Linguistique de Corpus, Sep 2013, Lorient, France. hal-02306443

**HAL Id: hal-02306443**

**<https://hal.science/hal-02306443>**

Submitted on 5 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Procédés de traduction humaine mis en évidence dans les ruptures lexicogrammicales avec la traduction d'un système de traduction automatique probabiliste

Éric Poirier

Université du Québec à Trois-Rivières

[Eric.Poirier@uqtr.ca](mailto:Eric.Poirier@uqtr.ca)

## Résumé :

Nous présentons une analyse des procédés de traduction humaine mis en évidence dans les ruptures lexicogrammicales des traductions humaines avec les traductions d'un système automatique probabiliste. Le repérage des ruptures lexicogrammicales et leur classement permettent de caractériser la valeur ajoutée des interventions des traducteurs humains et la créativité dont ils font preuve sur les plans discursif et cognitif. L'objectif de notre travail ne consiste pas à juger de la qualité de la traduction machine ou humaine mais de jeter les bases d'une étude empirique des faits de traduction constatés dans la comparaison des traductions d'un même texte dans deux corpus bilingues parallèles.

## Introduction

Le présent article fait état de travaux préliminaires de recherche qui ont été menés à l'été 2013. Le projet est né de la reconnaissance de l'importance des avancées techniques en traduction automatique et du rôle éventuel de cette dernière dans la formation à la traduction. La cohabitation désormais possible des deux modes de traduction, humaine et automatique, impose aux traducteurs professionnels et encore plus aux formateurs en traduction professionnelle de reconnaître et de mettre en évidence la valeur ajoutée de la traduction humaine par rapport à la traduction automatique. C'est précisément le but des travaux préliminaires que nous avons entrepris et qui consistent à étudier les différences, ce que nous appelons aussi les ruptures lexicogrammicales, dans un seul corpus de textes sources du domaine de l'économie, entre la traduction officielle publiée par un organisme bilingue du Canada et la traduction automatique du même texte source réalisée par le système de traduction automatique *Google Traduction*.

### 1. Description du corpus étudié

Nos deux corpus bilingues comportent le même texte source constitué de textes publiés dans un périodique appelé *Business Outlook Survey* publié par la Banque du Canada, un organisme officiellement bilingue dont les traductions sont réalisées par des professionnels. Dans l'ensemble des publications de la Banque du Canada, le périodique choisi présente une analyse de la situation économique et financière du pays et dans le monde. Son titre en français est *Enquête sur les perspectives des entreprises*<sup>1</sup>. Le texte source de nos corpus est constitué d'un échantillon prélevé dans cette publication périodique (quatre numéros par an) sur trois périodes d'un an, soit pour les années 2008 (vol. 5), 2010 (vol. 7) et 2012 (vol. 9). Nous avons réuni les textes des deux premières rubriques du périodique parus dans ces numéros, à savoir les rubriques *Overview* (= **Vue**

**d'ensemble**) et *Business Activity* (= **Activité économique**). Notre échantillon de 4767 mots est constitué de vingt-quatre extraits du périodique dont la longueur moyenne est de 146 mots pour la première rubrique et de 251 mots pour la deuxième rubrique. Pour chaque segment du texte source, nous avons relevé les ruptures lexicogrammaticales mises en évidence dans les constructions de la traduction humaine par comparaison avec la traduction machine obtenue par le service *Google Traduction*. Les généralisations que l'on pourra tirer à partir de ces différences permettront de caractériser les procédés de la traduction qui sont propres à la traduction professionnelle.

## 2. Cadre théorique

L'étude comparée de la traduction que nous avons entreprise a pour objet la traduction comme résultat, comme un texte traduit, qui est le produit humain ou automatique d'un travail ou d'une opération de traduction qui nécessite un effort cognitif chez l'humain ou de calcul pour la machine. L'expression « faits de traduction » permet de désigner cette conception particulière de la traduction et concorde bien aussi avec les objectifs des *Journées internationales de linguistique de corpus à Lorient*, comme l'indique l'appel à communication qui précise que l'événement réunit « des chercheurs venus d'horizons divers qui s'intéressent à l'utilisation de l'informatique pour l'analyse des faits de langue ». Pour ce qui concerne les corpus bilingues, on pourrait tout aussi bien ajouter à ces faits de langue les faits de traduction mis en évidence dans les textes traduits, ce qui correspond tout à fait au champ d'étude des comparaisons de traductions défini par Williams et Chesterman (2011 : 6) qui consiste à comparer les traductions avec leur texte source, une approche centrée sur la conception de la traduction au sens d'un résultat, un texte traduit, un fait accompli. À la différence de ce type d'étude, nos travaux portent plutôt sur la comparaison de deux traductions différentes et ne tiennent pas compte des relations qu'elles entretiennent chacune séparément avec le texte source, comme nous le décrivons plus loin dans notre méthodologie.

Pour ce qui concerne la conception de la traduction comme résultat, nous postulons qu'il existe deux principaux procédés de traduction textuels ou sémiotiques que nous avons définis ailleurs (Poirier, 2003) à propos de la traduction des unités phraséologiques mais qui conviennent bien à toutes les autres unités de traduction : à savoir la traduction par référence bilingue (ce que d'autres, comme dans Delisle, Lee-Jahnke et Cormier (dir.) (1999), ont appelé la correspondance) et la traduction par calcul sémantique (ce que d'autres, comme dans Delisle, Lee-Jahnke et Cormier (dir.) (1999), ont appelé l'équivalence).

Nous avons examiné les ruptures lexicogrammaticales, c'est-à-dire les changements ou les ruptures entre la traduction machine et la traduction humaine. Dans sa plus simple expression, les ruptures lexicogrammaticales désignent toutes les différences linguistiques et textuelles qu'il est possible de relever entre les deux textes traduits. L'orientation de la comparaison vise à mettre en évidence les particularités de la traduction humaine par rapport à la traduction machine et c'est la raison pour laquelle notre comparaison fait abstraction du texte source pour ne s'intéresser qu'aux différences entre les deux modes de traduction. Le dénombrement et le classement de ces différences est décrit plus loin à la section Méthodologie.

Nos deux corpus de traductions constituent des corpus bilingues parallèles au sens de Kenny (2011). En effet, les segments traduits de chaque type de traduction (professionnelle et machine) sont alignés ou mis en correspondance avec les segments du même texte de départ. L'alignement des textes traduits a été réalisé de façon manuelle. Le segment de départ dans les deux corpus est toujours le même mais il peut y avoir une segmentation différente entre la traduction professionnelle et la traduction machine.

### 3. Hypothèse de travail et objectifs de l'étude

Compte tenu de la définition précédente des procédés de traduction qui s'applique à des phénomènes textuels ou sémiotiques, nous émettons l'hypothèse que des procédés textuels de traduction différents se manifestent par des résultats de traduction différents marqués par des traces ou des ruptures lexicogrammaticales constatées entre les deux textes traduits. Dans la mesure du possible, nous nous intéressons plus spécialement dans ces faits de traduction aux différences qui découlent de procédés textuels ou sémiotiques distincts et non pas de variations expressives facultatives dans le processus de traduction littérale. L'objectif visé par notre étude consiste à tirer des généralités sur les particularités de la traduction humaine au chapitre de la traduction par équivalence et par calcul sémantique, une méthode de traduction qui différencie a priori la traduction humaine de la traduction machine.

### 4. Méthodologie

L'ensemble des échantillons a été réparti en deux groupes d'extraits présentés séquentiellement (Partie 1 et Partie 2) pour les besoins de l'analyse qui a été effectuée respectivement par deux étudiantes (Étudiant 1 et Étudiant 2) du programme de baccalauréat en traduction de l'Université du Québec à Trois-Rivières (UQTR). Les données ont été analysées dans un tableau de quatre colonnes comprenant le segment de départ, la traduction machine de ce segment de départ, la traduction professionnelle du même segment et la description des ruptures lexicogrammaticales entre les deux traductions. La figure ci-dessous présente un extrait de l'analyse faite par l'étudiant 2 à partir des deux premiers segments qui étaient à analyser dans la Partie 2 des extraits. On peut remarquer dans la dernière colonne l'utilisation d'indices numérotés qui servent de repères pour faciliter la délimitation de l'analyse lexicogrammaticale.

**Figure 1. Extrait des données recueillies dans un tableau**

No	Texte source (extraits à segmenter par phrase)	Traduction automatique (Google)	Traduction professionnelle	Procédés de traduction humaine mis en évidence par les ruptures lexicogrammaticales
	Document 7 (octobre 2010)			
7.1	Responses to the autumn survey suggest that the economic recovery is progressing.	Les réponses à l'enquête de l'automne donnent à penser que la reprise économique se poursuit.	Les résultats de l'enquête de l'automne indiquent que la reprise économique se poursuit.	1. Variante fonctionnelle; traduction-rédaction non littérale [résultat n'est pas une des solutions proposées à l'article réponse du R&C]; touche l'ensemble du syntagme y compris la préposition; 2. Variante fonctionnelle : traduction-rédaction non littérale; [indiquer n'est pas une des solutions proposées à l'article suggest du R&C]; le sujet (non-humain) contribue sans doute à ce choix
7.2	Firms remain positive about the outlook over the next 12 months, but they generally expect growth to be modest, owing in part to a weaker outlook for the U.S. economy.	Les entreprises restent optimistes quant aux perspectives pour les 12 prochains mois, mais ils s'attendent généralement à ce que la croissance soit modeste, en partie à cause d'un affaiblissement des perspectives pour l'économie américaine.	Les entreprises continuent de se montrer optimistes à l'égard des perspectives pour les 12 mois à venir, mais elles s'attendent en général à ce que la croissance soit modeste, en raison notamment de l'affaiblissement des perspectives d'évolution de l'économie américaine.	1. Expansion [par périphrase] 2. Variante codée (quant aux-à l'égard des) 3. Transposition Adj-Prép V (prochains – à venir) 4. Interprétation juste du genre et du référent – ils/elles [type de base de la personnalisation des référents du texte] 5. Transposition Adv-Prép N 6. Variante codée; à cause-en raison (changement de registre) 7. Transposition PrépN- Adv en partie-notamment 8. Expansion 9. Variante fonctionnelle (relation sémantique entre les éléments unis par la préposition)

L'analyse proposée à la cinquième et dernière colonne a porté uniquement sur les ruptures constatées entre les segments des colonnes quatre et trois. La comparaison ne portait pas sur les

ruptures traductologiques entre les segments des colonnes trois et une, ou deux et une, ce qui correspond à la comparaison des traductions avec leur texte source, une méthode qui est utilisée le plus souvent dans l'analyse des corpus parallèles de textes traduits.

Différents critères, qui ne sont pas propres aux langues en présence, ont été utilisés pour classer les ruptures lexicogrammaticales. Le premier critère est celui de l'erreur localisée de traduction ou d'expression en langue d'arrivée. En partant du principe que la traduction humaine publiée ne contient plus d'erreurs, nous avons conçu trois principaux types de ruptures mettant en évidence les erreurs de la traduction automatique ou les écarts incorrects de celle-ci avec la traduction humaine : la **post-édition obligatoire**, qui comprend les fautes d'orthographe et d'accord, l'emploi de mots outils incorrects, les constructions impropres, etc.; les **interprétations justes** (de la traduction humaine) qui correspondent le plus souvent à la correction d'une traduction littérale qui crée un faux sens; et la **modalité**, qui regroupe des erreurs sur des éléments accessoires seulement comme le genre, le nombre, les prépositions, le temps verbal, etc. Ces trois types d'erreur ont été traités en priorité sur tous les autres types de rupture. Ce qui explique qu'en cas de double classement possible des différences entre les deux traductions, seul le classement parmi les erreurs de traduction a été retenu.

Un deuxième critère de classement est celui du nombre d'éléments de sens qui sont touchés par la rupture lexicogrammaticale. Pour ce qui est du décompte des éléments de sens, nous avons pris en compte les classes de mots riches que sont les noms, les verbes, les adjectifs et les adverbes lorsqu'ils sont employés dans des expressions libres non figées. Dans le cas des expressions figées, elles ont été considérées sur ce plan comme des éléments de sens simples. C'est le cas par exemple de la locution prépositionnelle « compte tenu de », de la locution adverbiale « tout de même ». Lorsque la variation entraîne un élément supplémentaire dans la traduction humaine, nous avons classé cette rupture dans la catégorie des **expansions** (ce que d'autres ont appelé l'étoffement) tandis que lorsque la variation entraîne un élément en moins dans la traduction humaine, nous avons classé cette rupture dans la catégorie des **réductions** (ou une simplification textuelle). Lorsque la rupture met en relation des expressions qui ont le même nombre d'éléments de sens, nous avons créé deux autres types : la **variante codée** qui désigne une variante synonymique reconnue comme telle en langue d'arrivée ainsi que la **variante fonctionnelle** qui n'est pas liée sur le plan du code ou de la langue mais uniquement sur le plan du discours ou en raison d'un contexte particulier.

Enfin, nous avons utilisé un troisième critère de classement des ruptures qui se rapporte à la nature différentielle des éléments de sens que contient la rupture. Ce critère réunit des procédés traditionnels de traduction qui s'applique à des unités de même niveau linguistique ou qui sont comparables sur le plan textuel (de l'ordre de l'unité simple ou du syntagme) : la **transposition** (changement de catégorie grammaticale des éléments de sens) et la **modulation** (changement de point de vue dans le passage d'une expression à une autre), la première caractérise les différences de catégories de mots tandis que la deuxième caractérise les nuances de sens entre les deux formulations.

Toutes les ruptures n'ont pas été prises en compte. C'est le cas des accords grammaticaux qui découlent du genre ou du nombre d'un mot. Seule la rupture constatée dans le choix du mot lui-même a été consignée, et non pas, par exemple, les changements qui découlent de la variation en genre des deux mots et qui entraîne des accords différents (masculin et féminin) du même adjectif qui se rapporte à ce mot. À titre d'exemple, dans la rupture lexicogrammaticale constatée pour la traduction du segment « the most frequently cited target » = a) la cible la plus souvent citée et b) l'élément le plus souvent cité, seule la rupture de « cible » avec « élément » a été consignée et non pas la variation en genre du participe adjectif « cité », laquelle découle d'un choix fait ailleurs. Par contre, certains accords mal faits dans la traduction automatique ont été consignés comme des post-éditions obligatoires lorsqu'ils ne sont pas associés à un choix lexical particulier.

Le dénombrement des ruptures est brut. Les mêmes ruptures qui sont attribuables à une répétition interne du texte source ont été prises en compte. Suivant ce choix méthodologique, une même rupture comme celle du « solde d'opinion » et « solde des opinions » a compté pour de nombreux cas de post-éditions obligatoires, soit autant de fois que l'on retrouve ce segment dans les extraits choisis. Dans l'ensemble, nous avons cherché à attribué une seule catégorie pour chaque rupture lexico-grammaticale. Dans de très rares cas où les ruptures peuvent figurer dans deux catégories, nous avons privilégié l'ordre des trois critères en accordant la priorité aux erreurs et post-éditions obligatoires. En deuxième lieu, nous avons privilégié le dénombrement différent des unités de sens (expansion et réduction). En troisième lieu, lorsqu'il y avait transposition et modulation, nous avons privilégié la transposition. Dans la catégorisation des variantes (codées ou fonctionnelles), il n'y a pas de double catégorisation puisque les deux catégories sont mutuellement exclusives. La principale difficulté dans leur cas a été de déterminer si certaines unités lexicales composées (contenant plusieurs éléments intégrés) appartenaient aux variantes en raison de leur caractère figé, ou si elles pouvaient être rangées parmi les expansions ou les réductions, en partant du principe que l'expansion n'est pas figée.

Ces différents critères ont permis de dégager une grille d'analyse réunissant neuf cas de rupture types qui sont définis ci-après.

## 5. Grille d'analyse des ruptures lexicogrammicales

L'analyse des ruptures a permis de dégager neuf situations récurrentes de rupture qui mettent en évidence des conditions textuelles ou communicatives qui peuvent être invoquées pour justifier, décrire ou expliquer l'écart constaté entre la traduction machine et la traduction professionnelle. Nous avons cherché à établir des critères qui peuvent s'appliquer systématiquement à la comparaison de segments traduits. Même si la nature des ruptures et leur étendue peuvent varier quelque peu d'un traducteur à l'autre, nous avons cherché à utiliser des critères simples dont l'application reste dans la mesure du possible la plus objective. Comme pour ce qui est de toute variation en traduction, nous avons cherché à dégager l'analyse la plus courante qui serait faite sur ces ruptures plutôt que de chercher à dégager une analyse qui ferait l'unanimité. De notre point de vue, cette orientation de l'analyse est la seule qui permette une approche scientifique des faits de traduction.

Le premier type de ruptures réunit des segments traduits qui comportent des erreurs et des fautes manifestes de langue ou de traduction en traduction machine. Le principal critère retenu est le caractère incontestable de l'erreur ou de la faute. Pour cette raison, ce type de rupture a été désigné comme une post-édition obligatoire dans le contexte où, par exemple, un traducteur professionnel utiliserait la traduction machine comme première ébauche de sa traduction. Bien entendu, il s'agit d'une situation entièrement hypothétique et nous ne prenons pas en compte bien d'autres facteurs que ce mode de travail soulève, notamment sa rentabilité. Nous cherchons simplement à décrire ici une réalité qui pourrait être envisagée dans certains contextes particuliers. La figure ci-dessous présente trois exemples de ce type de rupture tirés de notre corpus.

**Figure 2. Trois exemples de rupture dans un même segment**

The balance of opinion on employment edged up in the spring survey (Chart 4).	Le solde d'opinion sur l'emploi a légèrement up dans l'enquête du printemps (Chart 4).	Le solde des opinions concernant le niveau de l'emploi s'inscrit en légère hausse dans l'enquête du printemps (Graphique 4).
---	--	--

Dans l'exemple ci-dessus, les blocs d'éléments en noir correspondent à trois ruptures qui correspondent chacune à une post-édition obligatoire. Les trois ruptures illustrent deux types d'erreur qui sont à l'origine des post-éditions obligatoires : l'inacceptabilité grammaticale attribuable à un mauvais usage en langue d'arrivée dans le premier cas et des mots (un verbe et un nom) non traduits dans les deux autres cas. Dans le relevé que nous avons fait des cas de post-édition obligatoire, nous n'avons pas recensé toutes les explications qui sont à l'origine de l'erreur puisque nous reportons cette analyse à des travaux ultérieurs. Notre analyse se limite pour l'instant à un dénombrement du phénomène à l'intérieur des segments pour dégager un portrait d'ensemble, mais il y aurait lieu d'analyser plus en profondeur la nature et la portée de ces écarts.

La deuxième catégorie de rupture réunit les erreurs de traduction qui reposent sur une interprétation littérale (codée) mais inadéquate d'une unité lexicale, qui est remplacée en traduction humaine par une correspondance textuelle (fonctionnelle). Le phénomène que nous avons voulu isoler ici est celui de l'emploi d'un mot qui se rapproche de la traduction adéquate mais qui ne convient pas pour des raisons contextuelles (collocation, acception, domaine d'emploi, etc.) qui restent à préciser. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.

**Figure 3. Rupture classée comme une interprétation juste**

Indicators of pressures on production capacity are above the levels recorded during the recession, (...).	Indicateurs de tensions sur les capacités de production sont supérieurs aux niveaux enregistrés durant la récession, (...).	Les indicateurs des pressions s'exerçant sur la capacité de production sont supérieurs aux niveaux enregistrés durant la récession, (...).
---	---	--

L'erreur de traduction repose ici sur l'absence d'interprétation de la part de la traduction machine qui associe une traduction courante (*tensions*) d'un mot (*pressures*) sans tenir compte des contextes textuels ou thématiques du correspondant en langue d'arrivée et de ses acceptions particulières. Ce type d'erreur se produit systématiquement lorsque la traduction s'appuie sur des associations de mots et non pas sur des associations d'acceptions, lesquelles sont plus sensibles aux contextes textuels et thématiques d'utilisation des mots. Dans l'exemple ci-dessus, la traduction la plus courante du mot *pressures* est utilisée en français sans tenir compte de son inadéquation avec le contexte discursif ou thématique qui crée un faux sens ou une erreur d'usage en langue d'arrivée. L'exclusion de la traduction usuelle impose l'emploi d'un autre mot de même sens (un synonyme) qui ne crée pas de faux sens ou d'erreur d'usage en langue d'arrivée. Encore une fois, notre analyse se limite à une appréciation globale du phénomène à l'intérieur des segments, mais il y aurait lieu d'analyser plus en profondeur les déterminismes de ces ruptures sur les plans textuels et thématiques.

La troisième catégorie de ruptures porte sur la reformulation par expansion et est mise en évidence par des étoffements et des explicitations qui ne se retrouvent pas dans la traduction machine. Comme nous l'avons indiqué précédemment, l'analyse que nous avons faite ne consiste pas à comparer le segment traduit avec le segment source mais à comparer entre eux les deux segments traduits. Comme indiqué précédemment, nous avons défini l'expansion par une augmentation systématique du nombre d'unités lexicales présentes dans la formulation, en faisant abstraction des unités lexicales fonctionnelles (que l'on appelle aussi les mots outils) comme les déterminants, les prépositions et les conjonctions. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.



**Figure 4. Rupture classée comme une expansion**

Indicators of pressures on production capacity are above the levels recorded during the recession, (...).	Indicateurs de tensions <b>sur</b> les capacités de production sont supérieurs aux niveaux enregistrés durant la récession, (...).	Les indicateurs des pressions <b>s'exerçant sur</b> la capacité de production sont supérieurs aux niveaux enregistrés durant la récession, (...).
---	--	---

Dans l'exemple ci-dessus, on considère qu'il y a expansion du fait qu'il y a ajout d'une unité lexicale pleine (le verbe s'exercer) à partir d'un texte où il n'y a qu'une unité lexicale fonctionnelle (la préposition sur). Comme on peut le constater, la rupture lexicogrammaticale par expansion caractérise aussi bien les variations de l'énonciation dans une même langue comme c'est le cas ici (entre les éléments en noir des troisième et deuxième colonnes) que les traductions interlinguistiques (entre la locution prépositionnelle de la troisième colonne en noir et la préposition *on* de la première colonne à laquelle elle correspond). On pourrait dire que le phénomène n'est pas propre à la traduction mais qu'il est caractéristique de la démarche de reformulation et d'expression linguistique dans le processus de rédaction.

Le départage entre les variantes et les ruptures par expansion est parfois purement formel. Le plus souvent, l'expansion induit un changement de sens par l'ajout d'un élément de sens à l'énoncé. Mais ce n'est pas toujours le cas, comme le montre l'exemple de l'expansion suivante dans laquelle on constate qu'il n'y a pas de changement de sens malgré la longueur différente des syntagmes en présence (selon notre critère définitionnel de l'expansion) : *sur* (Prép) et *s'exerçant sur* (Vpprésent Prép). D'un point de vue sémantique et fonctionnel, force est de constater que ces deux syntagmes sont unis par une relation de synonymie, malgré leur nature très différente. On peut faire la même analyse avec *beaucoup de* (Adv Prép) et *bon nombre de* (Adj N Prép). Curieusement, certaines expansions formelles représentent donc, sur le plan sémantique, des variantes. Cette situation s'explique très certainement par la lexicalisation des unités constitutives des expressions en cause. Nous renvoyons l'étude de ce phénomène à des travaux ultérieurs. Dans l'analyse du corpus, nous avons écarté l'évaluation du caractère phraséologique des segments de traduction pour des raisons de simplification. Il en découle qu'une partie des ruptures lexicogrammaticales par expansion pourront figurer dans une deuxième analyse parmi les variantes.

Dans certains cas qui ont été observés, les ruptures par expansion caractérisent même des formes d'hypercorrection comme dans l'exemple suivant : « Les opinions *sur* l'emploi », qui a été traduit en traduction humaine par « Les opinions *concernant le niveau de* l'emploi ». Dans ces cas-là, on constate que la traduction humaine est la version qui présente des erreurs ou des fautes par rapport à la traduction automatique. Ce phénomène, que nous avons relevé à quelques reprises, mérite d'être étudié plus en profondeur mais nous reportons son étude à des travaux ultérieurs.

La quatrième catégorie de ruptures lexicogrammaticales désigne au contraire de la précédente les situations où on constate une réduction de la formulation proposée par la traduction machine. La traduction professionnelle peut être décrite dans ce cas comme une simplification de la formulation que l'on trouve dans la traduction machine. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.



**Figure 5. Deux ruptures classées comme une réduction**

<p>This positive sentiment, supported by firms' expectations of improving demand and plans for Expansion, is widespread across all regions and most sectors.</p>	<p>Ce <b>sentiment positif</b>, soutenu par les attentes d'amélioration de la demande et des projets d'expansion des entreprises, <b>est très répandue</b> dans toutes les régions et la plupart des secteurs.</p>	<p>Cet <b>optimisme</b>, que révèlent les attentes des entreprises quant à un redressement de la demande ainsi que les projets d'expansion, <b>s'est observé</b> dans toutes les régions et la plupart des secteurs.</p>
--	--	--

La cinquième catégorie de ruptures lexicogrammaticales porte sur les modalités d'énonciation en langue d'arrivée. Ces ruptures s'appliquent à des éléments secondaires de l'énonciation qui contribuent au sens global mais qui y jouent un rôle accessoire. Le critère qui a été retenu est l'appartenance des unités lexicales en cause au groupe des mots fonctionnels, un groupe de mots qui a déjà servi à la définition des expansions et des réductions. De façon générale, une mauvaise traduction de ces éléments nuit moins à la compréhension du sens de l'énoncé que les erreurs qui touchent les autres unités lexicales. La figure suivante présente deux exemples de ce type de rupture tiré de notre corpus.

**Figure 6. Ruptures classées comme des modalités**

<p>Indicators of pressures on production capacity are above the levels recorded during the recession, (...).</p>	<p>(∅) Indicateurs de tensions sur <b>les</b> capacités de production sont supérieurs aux niveaux enregistrés durant la récession, (...).</p>	<p><b>Les</b> indicateurs des pressions s'exerçant sur <b>la</b> capacité de production sont supérieurs aux niveaux enregistrés durant la récession, (...).</p>
--	---	---

On remarque ici l'élément omis ou manquant qui est signalé par le symbole de l'ensemble vide. Contrairement aux ruptures classées comme une expansion, la première rupture dans l'exemple ci-dessus est classée parmi les ruptures de modalité parce qu'elle porte sur un élément lexical qui fait partie des mots outils comme les déterminants, les prépositions, les conjonctions, les auxiliaires, les adjectifs possessifs, etc. La deuxième rupture est marquée par l'élément modal du nombre qui passe du pluriel au singulier.

La sixième catégorie de ruptures lexicogrammaticales réunit les transpositions, une catégorie de ruptures qui caractérise traditionnellement des procédés formels de traduction. Les ruptures que nous avons classées dans cette catégorie correspondent à un réaménagement des éléments à l'intérieur des syntagmes qui comporte très souvent un changement de catégorie grammaticale. L'élément de sens reste le même mais change de catégorie grammaticale. Le critère que nous avons retenu ici est que la transposition n'entraîne pas de changement significatif de sens ou de point de vue. Les deux segments des textes traduits expriment donc sensiblement le même sens. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.

**Figure 7. Deux ruptures classées comme une transposition**

(...) following a period during which prices were falling or frozen.	après une période pendant laquelle les prix étaient <b>en baisse</b> ou <b>gelés</b> .	après une période de <b>rabais</b> ou de <b>gel</b> .
--	--	---

Dans la figure 7, la première transposition est aussi « précédée » textuellement de la réduction de la proposition relative par un complément du nom. La relation entre une proposition relative et un complément du nom est un procédé de traduction qui a été observé dans le corpus, dans le sens de l'expansion ou de la réduction. Comme le procédé est fréquent et qu'il caractérise une transformation syntaxique, on pourrait aussi l'analyser comme une transposition. Ce n'est cependant pas ce que nous avons fait puisque les phénomènes d'expansion et de réduction ont été classés de façon prioritaire par rapport aux transpositions. Rien n'empêche, dans une deuxième analyse, de relever les ruptures classées comme expansion ou réduction en tant que transposition particulière, comme celle qui unit les relatives et les compléments du nom (prépositionnels ou adjectivaux). Un autre exemple intéressant de transposition dans notre corpus est fourni par les relations entre les adjectifs et les compléments de nom utilisés pour les noms toponymiques : « des États-Unis » = « américains » ou « étatsuniens ». Ce type de transposition, à l'instar de la relation complément du nom - proposition relative, est biunivoque.

La septième catégorie de ruptures lexicogrammaticales réunit les modulations et regroupe des ruptures qui sont caractéristiques d'un changement de point de vue sur le plan sémantique. Le critère que nous avons retenu pour les modulations c'est la présence d'un lien analogique sémantique de la version de la traduction humaine avec la traduction machine. Lorsque la rupture était marquée à la fois par une différence formelle et sémantique, nous avons priorisé la première qui est caractéristique des transpositions. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.

**Figure 8. Rupture classée comme une modulation**

This indicator continues to point to an increase in investment over the next 12 months.	Cet indicateur continue de <b>pointer vers</b> une augmentation de l'investissement au cours des 12 prochains mois.	Cet indicateur continue de <b>laisser entrevoir</b> une hausse des investissements au cours des 12 mois qui viennent.
---	---	---

Cet exemple de modulation pourrait être analysé comme une expansion s'il n'avait pas été fait abstraction des mots outils comme les prépositions et les auxiliaires. Dans cet exemple, la traduction humaine ne compte pas deux verbes différents puisque le verbe *laisser* représente en fait un verbe auxiliaire. Pour cette raison, et contrairement à ce qu'une analyse superficielle pourrait laisser croire, il n'y a pas ici d'expansion mais plutôt une modulation.

Une huitième catégorie de ruptures grammaticales réunit les variantes codées qui sont principalement définies comme des variations acceptables (ou non fautives) de l'énonciation d'un élément de sens en langue d'arrivée (unité lexicale simple ou complexe ou syntagme). Le critère qui a été retenu ici est le caractère codé ou lexicalisé de la variation qui pourrait être décrite dans un dictionnaire de synonymes ou dans un dictionnaire bilingue qui décrirait les liens analogiques de sens entre les deux variantes. En ce sens, la variation est en quelque sorte prévisible jusqu'à un certain point puisqu'elle est relativement indépendante du contexte. La figure suivante présente un

exemple de ce type de rupture tiré de notre corpus.

**Figure 9. Rupture classée comme variante codée**

[...] firms are not expecting a marked change in the pace of business activity.	[...] les entreprises ne s'attendent pas à un <b>changement marqué dans le rythme</b> de l'activité économique.	[...] les entreprises ne s'attendent pas à une <b>variation marquée du rythme</b> de l'activité.
---	---	--

Ce type de rupture est généralement marqué par des unités linguistiques qui entretiennent des relations de synonymie, du type *proche – près*. Ici encore, l'emploi de l'un ou de l'autre mot est davantage une question de cotexte textuel et discursif et de microdomaine thématique. De même, les connotations des mots jouent sans aucun doute un certain rôle.

Une neuvième catégorie de ruptures grammaticales réunit les variantes fonctionnelles qui sont elles aussi définies comme des variations acceptables (ou non fautives) de l'énonciation d'un élément de sens en langue d'arrivée (unité lexicale simple ou complexe ou syntagme). Le critère qui a été retenu ici est le caractère imprévisible et la nature discursive de la variation (par opposition à la dimension linguistique des variantes codées). Contrairement aux variantes codées, les ruptures que nous avons classées parmi les variantes fonctionnelles font partie de variantes dont l'utilisation repose sur une prise en compte du contexte, de la situation et du domaine ainsi que de différents usages de termes qui restent plus difficiles à acquérir que les variantes précédentes. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.

**Figure 10. Rupture classée comme une variante fonctionnelle**

<u>Many firms reported that</u> (...)	<b>Beaucoup d'entreprises</b> ont déclaré que (...).	<b>Bon nombre de répondants</b> ont déclaré que (...).
---------------------------------------	--	--

Dans l'exemple ci-dessus, la situation d'énonciation fait en sorte que les répondants d'un sondage sont en fait des dirigeants d'entreprise (qui incarnent en fin de compte l'opinion de leur employeur) à qui on a demandé leur avis sur les perspectives économiques.

Une dernière catégorie a été envisagée en dernière analyse et il n'a pas été possible de l'appliquer à l'ensemble du corpus. Son intérêt n'a été relevé qu'à la fin de l'analyse. Cette catégorie constitue en fait le pendant de la première catégorie où des erreurs manifestes étaient produites par la traduction machine. Cette catégorie, étudiée seulement en partie et pour laquelle nous n'avons pas de statistiques fiables, réunit plutôt les erreurs manifestes que l'on trouve dans les traductions professionnelles. Le classement de ces éléments s'appuie sur le fait que les segments de la traduction machine ne sont manifestement pas fautifs tandis qu'ils le sont pour ce qui est de la traduction professionnelle. Ou si elles ne le sont pas expressément dans ce dernier cas, elles apportent une lourdeur de formulation qui n'est pas obligatoire et probablement inutile. La figure suivante présente un exemple de ce type de rupture tiré de notre corpus.

**Figure 11. Exemples d'hypercorrection de la traduction humaine**

The increase in this indicator is widespread across all regions and sectors.	L'augmentation de cet indicateur est très répandue dans toutes les régions et les secteurs.	La hausse affichée par cet indicateur s'est observée dans toutes les régions et tous les secteurs.
--	---	--

## 6. Discussion des résultats

Le tableau qui suit présente les résultats validés de l'analyse des ruptures lexicogrammaticales et leur répartition dans les différentes catégories définies ci-dessus.

**Figure 12**

Ruptures	Étudiant 1	Étudiant 2	Total
Post-éditions obligatoires de la TA	121 (14,9 %)	171 (22,0 %)	292 (18,3%)
Interprétations justes de la TH	30 (3,7 %)	47 (6,0 %)	77 (4,8 %)
Expansions	130 (16,0 %)	113 (14,5 %)	243 (15,3 %)
Réductions	73 (9,0 %)	44 (5,6 %)	117 (7,3 %)
Modalités	64 (7,9 %)	47 (6,0 %)	111 (7,0 %)
Transpositions	64 (7,9 %)	14 (1,8 %)	78 (4,9 %)
Modulations	5 (0,6 %)	17 (2,2 %)	22 (1,4 %)
Variantes codées (TA et TH)	178 (21,9 %)	166 (21,3 %)	344 (21,6 %)
Variantes fonctionnelles (TH)	149 (18,3 %)	152 (19,5 %)	301 (18,9 %)
Hypercorrections de la TA*	0*	8* (1,0 %)	8 (0,5 %)
Total	814	779	1593

On constate tout d'abord que les ruptures lexicogrammaticales sont très nombreuses. En effet, sur un corpus de 4767 mots, on relève pas moins de 1593 ruptures, tous types confondus, ce qui représente une rupture à tous les trois mots environ. Ce phénomène n'a rien de bien surprenant puisqu'il témoigne de l'importance de la variation énonciative qui est importante en rédaction et qui a été constatée chez les mêmes personnes à propos d'un même événement dans une expérience rédactionnelle dont fait état Gile (2011 : 86) et qui montre que 61,5 % des couples d'énoncés (le même énoncé rédigé une fois puis une seconde fois deux ou trois heures plus tard) présentaient des différences de formulation. Dans notre analyse, les variations énonciatives ne révèlent pas tant une grande variété de contenus informationnels que des procédés de traduction différents s'appliquant à un même texte de départ.

Si on répartit grossièrement les erreurs dans les textes de la traduction machine en erreurs de langue pour les ruptures à post-édition obligatoire et en erreurs de méthode de traduction pour les ruptures provenant d'une interprétation juste, on constate qu'un très grand nombre des erreurs de la traduction machine, 292 sur 369 (292+77), ou 79 % d'entre elles, donnent lieu à des fautes de langue ou de formulation qui sont caractérisées par une méconnaissance des normes d'usage ou des



règles d'utilisation du code linguistique. D'après cette répartition des erreurs de la traduction machine, on peut estimer qu'il serait très « rentable » d'utiliser un module de correction linguistique qui permettrait de relever et de corriger automatiquement des erreurs d'usage comme les accords, les non-traductions, les régimes et la conjugaison des verbes, la coordination adéquate des éléments dans la phrase et autres erreurs faciles à détecter par rapport aux erreurs de traduction et d'interprétation.

À partir du nombre de ruptures lexicogrammicales relevées dans le corpus, on peut dégager trois principaux procédés de traduction les plus fréquents : les variantes codées, les variantes fonctionnelles et l'expansion qui représentent, respectivement, 344 ou 21,6 %, 301 ou 18,9 % et 243 ou 15,3 % de toutes les ruptures que nous avons relevées. Le nombre le plus élevé de ruptures dans la catégorie des variantes codées montre bien le principe selon lequel la traduction procède aussi sur le plan intralinguistique qu'interlinguistique. Si les variantes codées illustrent l'importance du phénomène qui est attribuable aux mécanismes internes de la langue, les variantes fonctionnelles illustrent quant à elles les mêmes possibilités créatives qui relèvent plutôt du savoir-faire et de la créativité des traducteurs qui ont su tirer parti des liens complexes qui unissent les mots, les constructions et leurs usages textuels.

Le troisième procédé le plus fréquent de traduction est représenté par une démarche qui est bien connue en traduction et qui a fait de travaux récents sur les corpus de textes traduits. L'étude de Becher (2011) montre que les procédés d'expansion et de réductions sont en fait conditionnés par des facteurs lexicogrammicales et pragmatiques qui doivent être approfondis et testés sur des données empiriques. Notre travail s'inscrit dans ce programme de recherche et nous entendons mettre en évidence ces facteurs dans une analyse plus approfondie des déterminismes lexicogrammicales qui sont associés aux ruptures relevées dans notre corpus et notamment aux ruptures par expansion et réduction. Ici encore, les déterminismes de ces ruptures illustrent probablement la créativité et le savoir-faire des traducteurs sur le plan rédactionnel. Une meilleure connaissance de ces facteurs et de ce savoir-faire pourra être très utile à la compréhension de l'activité cognitive de la traduction et à l'élaboration d'activités de formation à la traduction professionnelle.

Une dernière information sur la compilation des données relatives aux ruptures lexicogrammicales tient au fait que nous avons dénombré un nombre relativement peu élevé de transpositions et de modulations. Pour les transpositions, cette situation s'explique peut-être par la décomposition en éléments significatifs des variations énonciatives entre les deux modes de traduction. Dans le cas des modulations, on peut mentionner que nous n'avons pas eu le temps de caractériser de façon systématique et exhaustive tous les types de relations analogiques du type partie-tout que l'on pouvait associer aux 1593 ruptures lexicogrammicales. Un certain nombre de variantes codées et de variantes fonctionnelles s'appuient peut-être sur un type ou un autre de modulation.

## **Conclusion**

L'étude des ruptures lexicogrammicales constatées entre deux traductions réalisées de manière foncièrement différente a permis de mettre en évidence les différences, dans un seul corpus de textes source du domaine de l'économie, entre la traduction humaine professionnelle publiée par un organisme officiellement bilingue du Canada et la traduction machine du même texte source réalisée par un système de traduction automatique probabiliste comme Google Traduction. Le nombre élevé de ruptures et leur variété témoignent nécessairement de la richesse et de la créativité dont font preuve les traducteurs dans leur travail.

Pour l'avancement des connaissances en traduction, la description de ces ruptures au moyen de critères objectifs constitue en soi une réalisation intéressante puisqu'elle a mis en évidence les

particularités de la traduction humaine (les variations lexicales, codées ou fonctionnelles, ainsi que la prépondérance de l'expansion par rapport à la réduction) de même que celles de la traduction machine (qui sont marquées par un grand nombre d'erreurs d'expression en langue d'arrivée qui ont peu à voir avec la méthode ou les procédés de traduction eux-mêmes).

Pour mieux décrire ces données, il reste encore à étudier les facteurs qui sont pris en compte par les traducteurs humains dans les choix qu'ils font grâce à ces procédés de traduction caractérisés par les ruptures lexicogrammaticales. Une meilleure connaissance de ces facteurs permettra de mesurer et de décrire les efforts déployés par les traducteurs dans leur travail, leur savoir-faire, et de mieux cibler les objectifs d'apprentissage de la traduction dans les programmes de formation à la traduction.

## Remerciements

Le projet de recherche a bénéficié du soutien financier du Fonds institutionnel de recherche de l'Université du Québec à Trois-Rivières, concours 2013-2014, Émergence de nouveaux professeurs en recherche. Je tiens à remercier de leurs efforts Marie-Hélène Gauthier et Louise Thériault, deux étudiantes du programme de baccalauréat en traduction qui ont travaillé à la collecte des données à l'été 2013.

## Note

<sup>1</sup> Même si aucune mention dans le texte ne précise que la traduction est orientée de l'anglais au français, nous postulons que les textes publiés par l'organisme en français sont en réalité des traductions des textes rédigés en anglais. Le parallélisme des deux versions des textes ne laisse planer aucun doute sur la relation de traduction qui les unit. Le statut asymétrique de l'utilisation de l'anglais et du français dans les institutions bilingues du Canada, notamment celles qui traitent d'économie et du système financier, de même que le statut de l'anglais comme langue d'usage internationale en matière de commerce et d'économie, permettent aussi de donner plus de poids à notre postulat.

## Références

Becher, Viktor (2011). *Explicitation and implicitation in translation. A corpus-based study of English-German and German-English translations of business texts*. Thèse soutenue à l'Université de Hambourg.

Delisle, Jean, Hannelore Lee-Jahnke and Monique C. Cormier (ed.) (1999). *Terminologie de la Traduction. Translation Terminology. Terminología de la Traducción. Terminologie der Übersetzung*, John Benjamins.

*Financial System Review = Revue du système financier* [périodique] : Banque du Canada, Ottawa, <http://www.bankofcanada.ca/publications-research/periodicals/fsr> [en ligne]. Page consultée le 2 septembre 2013. Documents sources du corpus parallèle de textes traduits par des traducteurs professionnels.

Gile, Daniel (2011). *La traduction. La comprendre, l'apprendre*. Paris : Presses universitaires de France.

Google (2013) : *À propos de Google Traduction. Comment fonctionne Google Traduction?*, [http://translate.google.ca/about/intl/fr\\_ALL/](http://translate.google.ca/about/intl/fr_ALL/) [en ligne]. Page consultée le 6 juin 2013.

Ke Ping (2011) : « Machine Translation » dans Baker et Saldanha (dir.) *Routledge Encyclopedia of Translation Studies*, deuxième édition, Londres et New York : Routledge, p. 162-169.

Kenny Dorothy (2011) : “Corpora” dans Baker et Salsanha (dir.) *Routledge Encyclopedia of Translation Studies*, deuxième édition, Londres et New York : Routledge, p. 59-62.

Poirier, Éric (2003) : « Conséquences didactiques et théoriques du caractère conventionnel et arbitraire de la traduction des unités phraséologiques », *Meta*, vol. 48, no 3, p. 402-410.

Williams, Jenny et Andrew Chesterman (2011) : *The Map. A Beginner's Guide to Doing Research in Translation Studies*, Kinderhook (NY) : St. Jerome Publishing.