



Repérage des décalages informationnels de traduction au moyen du criblage automatique des segments hétéromorphes d'un corpus parallèle

Éric André Poirier

Traduction et politique(s)

Translation, Politics and Policies

Volume 32, numéro 1, 1er semestre 2019

URI : <https://id.erudit.org/iderudit/1068022ar>

DOI : <https://doi.org/10.7202/1068022ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association canadienne de traductologie

ISSN

0835-8443 (imprimé)

1708-2188 (numérique)

[Découvrir la revue](#)

Citer cet article

Poirier, É. A. (2019). Repérage des décalages informationnels de traduction au moyen du criblage automatique des segments hétéromorphes d'un corpus parallèle. *TTR*, 32(1), 279–308. <https://doi.org/10.7202/1068022ar>

Résumé de l'article

Nous présentons une méthode de criblage quantitatif et automatique des segments hétéromorphes dans les corpus parallèles de traductions qui rationalise la recherche qualitative et manuelle des décalages de traduction informationnels. Ces derniers fournissent des données sur la traduction comme résultat qui servent à caractériser les idiosyncrasies textuelles et linguistiques des langues en contact et qui renseignent aussi sur les biais individuels du sujet traduisant ainsi que sur les biais culturels inhérents à l'opération de traduction pour les langues-cultures en contact. Pour montrer comment fonctionne la méthode de repérage, nous l'appliquons à un corpus de traductions alignées avec leur segment source. Le corpus contient un discours en anglais du président américain John F. Kennedy (1961) et un discours de la première ministre britannique Theresa May (2017) qui ont été traduits en français. Pour le criblage automatique des segments hétéromorphes, l'algorithme PML (précision des mots lexicaux) analyse automatiquement les paires de segments source et cible et compare le nombre de mots lexicaux dans les segments source et cible pour calculer le ratio de la précision de l'information traduite (PIT). Par la suite, les segments hétéromorphes criblés au moyen du ratio PIT sont analysés manuellement pour déterminer l'origine ou la cause de la variation des contenus informatifs de la traduction et définir les repérages qu'elle entraîne, le cas échéant. La méthode offre l'avantage de rationaliser la recherche de décalages dans les segments hétéromorphes négatifs et positifs des textes traduits dans un corpus parallèle, puisqu'elle permet de cibler et de cribler les paires de segments hétéromorphes qui sont le plus susceptibles de contenir des décalages informationnels de traduction.

Repérage des décalages informationnels de traduction au moyen du criblage automatique des segments hétéromorphes d'un corpus parallèle

Éric André Poirier

Université du Québec à Trois-Rivières

Résumé

Nous présentons une méthode de criblage quantitatif et automatique des segments hétéromorphes dans les corpus parallèles de traductions qui rationalise la recherche qualitative et manuelle des décalages de traduction informationnels. Ces derniers fournissent des données sur la traduction comme résultat qui servent à caractériser les idiosyncrasies textuelles et linguistiques des langues en contact et qui renseignent aussi sur les biais individuels du sujet traduisant ainsi que sur les biais culturels inhérents à l'opération de traduction pour les langues-cultures en contact. Pour montrer comment fonctionne la méthode de repérage, nous l'appliquons à un corpus de traductions alignées avec leur segment source. Le corpus contient un discours en anglais du président américain John F. Kennedy (1961) et un discours de la première ministre britannique Theresa May (2017) qui ont été traduits en français. Pour le criblage automatique des segments hétéromorphes, l'algorithme PML (précision des mots lexicaux) analyse automatiquement les paires de segments source et cible et compare le nombre de mots lexicaux dans les segments source et cible pour calculer le ratio de la précision de l'information traduite (PIT). Par la suite, les segments hétéromorphes criblés au moyen du ratio PIT sont analysés manuellement pour déterminer l'origine ou la cause de la variation des contenus informatifs de la traduction et définir les repérages qu'elle entraîne, le cas échéant. La méthode offre l'avantage de rationaliser la recherche de décalages dans les segments hétéromorphes négatifs et positifs des textes traduits dans un corpus parallèle, puisqu'elle permet de cibler et de cribler les paires de segments hétéromorphes qui sont le plus susceptibles de contenir des décalages informationnels de traduction.

Mots-clés : segment hétéromorphe, mot lexical, ratio PIT, décalage informationnel de traduction, corpus parallèle de traductions alignées

Abstract

We present a method of quantitative and automatic screening of heteromorphic segments in corpora of aligned translations that streamlines the qualitative and manual search for informational translation shifts. The latter provide data on translation as a result that serve to characterize the textual and linguistic idiosyncrasies of the languages in contact and also provide information on the individual biases of the subject translating as well as the cultural biases inherent in the translation operation for the languages-cultures in contact. To show how the tracking method works, we apply it to a bilingual corpus of translations aligned with their source segment. The corpus contains a speech by American President John F. Kennedy (1961) and a speech by British Prime Minister Theresa May (2017) that have been translated into French. For automatic screening of heteromorphic segments, the LWA (lexical word accuracy) algorithm automatically analyzes all source and target segment pairs and compares the number of lexical words in the source and target segments to calculate the translated information accuracy ratio (TIA). Subsequently, the heteromorphic segments screened using the TIA ratio are analyzed manually to determine the origin or cause of the variation in the translation's informational content and to define the shifts it causes, if any. The method has the advantage of rationalizing the search for translation shifts in negative and positive heteromorphic segments of translated texts in a parallel corpus since it allows targeting and screening of heteromorphic segment pairs that are most likely to contain informational translation shifts.

Keywords: heteromorphic segment, lexical word, TIA ratio, informational translation shift, parallel bilingual corpus

1. Introduction

Le présent article vise à montrer l'utilité des corpus parallèles¹ pour la mise au jour des opérations cognitives qui sont à l'œuvre dans l'opération de traduction. Nous proposons une méthode originale et inédite de criblage automatique des paires de segments d'un corpus parallèle à l'aide du ratio de la précision de l'information traduite (PIT), qui se calcule en divisant le nombre de mots lexicaux du segment cible par le nombre de mots lexicaux du segment source lui correspondant dans un corpus parallèle. Les corpus parallèles qui nous intéressent réunissent un ensemble de bitextes, c'est-à-dire des textes autonomes traduits indépendamment les uns des autres et

1. Nous utilisons le terme « corpus parallèles » pour désigner un corpus de traductions constitué des segments du texte source alignés avec les segments du texte cible, les deux textes ainsi réunis formant ce qu'on appelle aussi un « bitexte ». Les corpus parallèles de traductions ne doivent pas être confondus avec les corpus (unilingues) de textes parallèles, aussi appelés « corpus comparables », qui sont utilisés en traduction et qui réunissent des textes du même domaine ou genre que le texte source, mais qui sont écrits dans la langue cible (Gallego-Hernández, 2018).

dont chaque segment du texte source est aligné avec un, plusieurs ou aucun des segments du texte traduit. Les mots lexicaux correspondent à des noms, des verbes, des adjectifs et des adverbes, par opposition aux mots grammaticaux (déterminants, pronoms, conjonctions et prépositions). L'algorithme PML tire parti du caractère fermé des mots grammaticaux par opposition au caractère ouvert des mots lexicaux, comme l'explique Alain Polguère (2003, p. 80-81). Le choix du dénombrement des mots lexicaux comme élément comparatif offre la possibilité de quantifier le contenu informatif des segments analysés et de repérer automatiquement les paires de segments hétéromorphes dont les segments source et cible contiennent un nombre différent de mots lexicaux et qui sont le plus susceptibles de contenir des décalages informationnels de traduction.

Nous avons choisi d'appliquer cette méthode de repérage à un corpus de discours politiques, parce qu'à la différence des textes pragmatiques (au sens de Delisle et Fiola, 2013, p. 686), les discours politiques semblent contenir davantage de marques idéologiques et culturelles et reposer sur des postulats et des principes (implicites) qui traversent rarement indemnes les frontières des langues-cultures. Le choix d'un tel corpus repose ainsi sur l'hypothèse selon laquelle les discours politiques sont susceptibles d'être plus fréquemment hétéromorphes que les textes pragmatiques et de contenir un plus grand nombre de décalages.

Un décalage de traduction ou *translation shift* est généralement défini comme une transformation du texte source qui se manifeste dans le texte traduit et qui ne constitue pas une différence systémique entre les langues source et cible (Bakker, Koster, et Van Leuven-Zwart, 2011, p. 269). Il s'oppose aux résultats obtenus par la traduction littérale (au sens de Delisle et Fiola, 2013, p. 689) et réunit un ensemble de faits de traduction hétéroclites comme les modulations, les transpositions, les adaptations, les implicitations, les étoffements et les phénomènes de compensation, dans la terminologie de Delisle et Fiola (2013). Les décalages de traduction s'opposent ainsi aux éléments invariants qui sont constatés dans les segments source et cible, et que l'on peut considérer comme des « ressemblances » ou, autrement dit, comme des éléments dont l'emploi est « commun » aux deux langues dans la production de textes. À titre d'exemple, la traduction « Le texte est mal traduit » du segment « The text is poorly translated » est considérée comme une traduction littérale et, de ce fait, ne comporte aucun décalage, tandis que la traduction « Le texte reste controversé »

du segment source «The text is still controversial» comporte un léger décalage avec l'emploi du verbe «rester» comme traduction de la construction «is still»². Suivant notre définition des décalages de traduction, les paires de segments qui ne présentent pas de décalage sont généralement constituées de traductions littérales. De plus, les paires de segments sont considérées comme étant isomorphes lorsque leurs segments source et cible comptent le même nombre de mots lexicaux et comme étant hétéromorphes lorsque ce n'est pas le cas.

Le présent article décrit une méthode rationnelle de recherche dans les corpus parallèles des décalages de traduction induits par la variation du nombre de mots lexicaux entre les segments source et cible. L'étude des contextes phraséologiques et discursifs de ces décalages permet de caractériser les idiosyncrasies de chaque langue, une information essentielle à l'exercice de la traduction. Elle permet aussi de mieux connaître les activités cognitives qui sont à l'œuvre dans l'opération de traduction et qui témoignent des efforts du sujet traduisant à surmonter les dissonances linguistiques, psychologiques (individuelles) et culturelles qui se présentent dans l'opération de traduction. La recherche puis l'analyse des décalages dans les corpus parallèles permettent ainsi d'étudier un ensemble de phénomènes linguistiques et traductologiques qui se manifestent par l'hétéromorphisme des segments source et cible.

L'intérêt de la méthode que nous proposons est qu'elle peut s'appliquer aux bitextes en anglais, en français et en espagnol, et éventuellement dans toutes les autres langues qui possèdent des mots grammaticaux et des mots lexicaux, lesquels déterminent généralement le dénombrement des contenus informatifs des énoncés. La comparaison du nombre de mots lexicaux dans les paires de segments sert à repérer un certain type de décalage de traduction que nous appelons «décalages informationnels», lesquels sont définis par une variation du nombre de mots lexicaux (et du contenu informatif) entre les segments source et cible. La mesure du ratio PIT peut être faite de façon manuelle par le dénombrement des mots lexicaux, ou de façon automatique à l'aide de l'algorithme PML (précision des mots lexicaux). Les deux méthodes de calcul du ratio PIT sont décrites dans Poirier (2017b).

2. Comme nous l'expliquons à la section 4, les décalages peuvent être figés ou libres selon qu'ils sont obligatoires ou facultatifs. Ici, l'inacceptabilité de la construction «est toujours» dans le microcontexte de la phrase citée en exemple justifie l'analyse d'un décalage obligatoire, donc figé.

Les segments des bitextes sont alignés en amont par un logiciel propriétaire; chaque phrase graphique du texte source est alignée à la phrase graphique à laquelle elle correspond normalement dans le texte traduit. Cette étape n'est pas triviale, puisque l'alignement ainsi produit peut être incorrect du simple fait que le découpage des paragraphes en phrases varie de façon imprévisible d'une langue à l'autre et d'un rédacteur à l'autre, en fonction des besoins et préférences stylistiques des rédacteurs et des traducteurs. Pour les besoins de notre étude, les bitextes mal alignés ont été corrigés manuellement³. Les bitextes se présentent en au moins deux colonnes : la colonne de gauche contient les phrases graphiques ou les segments du texte source, et la colonne de droite contient les phrases graphiques ou les segments du texte cible qui leur correspondent. Le logiciel qui a été utilisé dans le cas présent est LogiTerm^{MC}, avec la fonctionnalité de création de bitextes en format HTML.

Dans les sections qui suivent, nous décrivons le fonctionnement de l'algorithme PML, tel que conçu et programmé en langage Python, et la méthode de calcul du ratio PIT qui a servi à cribler les segments des bitextes choisis. Nous décrivons la méthode de criblage utilisée pour sélectionner les paires de segments qui sont le plus susceptibles de contenir des décalages informationnels. Nous présentons une vue d'ensemble des types de décalage repérés grâce à la méthode de criblage. Pour chacun des bitextes du corpus, nous analysons les décalages trouvés dans les segments sélectionnés parmi l'ensemble des segments hétéromorphes. En conclusion, nous décrivons les avantages offerts par la méthode employée et les perspectives d'amélioration de celles-ci.

2. L'algorithme PML et le calcul automatique du ratio PIT

L'algorithme calcule le ratio PIT de chaque paire de segments en divisant le nombre de mots lexicaux du segment cible par le nombre de mots lexicaux du segment source. Le classement des mots-formes de chaque segment dans l'un ou l'autre des groupes de mots lexicaux ou grammaticaux s'effectue au moyen d'une liste préétablie de mots-formes traités comme des mots grammaticaux qui constituent une classe fermée (Polguère, 2003, p. 81). Une fois établi le nombre de mots-formes appartenant à chacune des deux classes de mots, le ratio

3. La méthode de criblage que nous proposons permet de repérer rapidement certains de ces alignements incorrects comme ceux qui ne sont associés à aucun contenu informatif.

PIT se calcule en divisant le nombre de mots lexicaux du segment cible par le nombre de mots lexicaux du segment source, comme indiqué ci-dessous :

$$\text{PIT} = \frac{\text{Nombre de mots lexicaux dans le segment cible}}{\text{Nombre de mots lexicaux dans le segment source}}$$

Le ratio obtenu fournit une mesure de la précision informationnelle de la traduction par rapport au contenu informatif du segment source. Le ratio PIT ne vise pas à mesurer la qualité de la traduction, mais bien le rapport numérique global entre le volume d'information que contient le segment cible et celui que contient le segment source. L'intérêt de la prise en compte uniquement des mots lexicaux dans le calcul du volume d'information tient au fait que leur traduction est généralement plus stable que celle des mots grammaticaux, auxquels sont souvent associées de nombreuses traductions différentes (Villada Moirón, Begona et Tiedmann, 2006). En effet, la contribution sémantique de ces deux classes de mots au sens des énoncés est bien différente : le sens des mots grammaticaux est de nature relationnelle tandis que le sens des mots lexicaux est lexicalisé, donc suffisamment stable pour qu'un contenu informatif leur soit associé (indépendamment du fait que leur signification varie selon le contexte).

La distinction entre les deux classes de mots n'est pas toujours aisée. Un même mot-forme peut en effet appartenir à l'une ou à l'autre, selon l'emploi qui en est fait. Par ailleurs, des mots-formes d'une classe de mots lexicaux peuvent à l'occasion jouer le rôle de formants ou d'éléments constitutifs dans une expression figée, comme c'est le cas du mot-forme « bien », qui est généralement un adverbe, mais dont le statut adverbial dans la locution conjonctive « bien que » est indiscernable. Dans l'application de l'algorithme, c'est l'appartenance à la classe la plus fréquente qui a été privilégiée, même si ce choix peut entraîner un faux décalage (voir plus loin) lorsque le mot-forme est employé dans une acception qui appartient à l'autre classe de mots. C'est ainsi que les verbes *être*, *avoir*, *pouvoir*, *to be*, *to have*, *to do*, *shall*, *can*, *will*, *would* et leurs formes conjuguées ont été systématiquement traités comme des mots grammaticaux, même s'ils peuvent à l'occasion être employés comme verbe principal.

3. Criblage des segments hétéromorphes les plus marqués

La première étape de la méthode consiste à calculer le ratio PIT de toutes les paires de segments d'un bitexte, puis à classer et à coter

les paires hétéromorphes au moyen du ratio PIT. Les paires sont considérées comme isomorphes lorsque leur ratio PIT est de 1,0; elles sont considérées comme hétéromorphes lorsque leur ratio PIT est inférieur ou supérieur à 1,0. Lorsque le ratio PIT des paires de segments est inférieur à 1,0, ce sont des segments hétéromorphes négatifs; lorsque leur ratio PIT est supérieur à 1,0, ce sont des segments hétéromorphes positifs.

Le calcul du ratio PIT des paires de segments permet de les trier numériquement et de chercher les décalages dans les paires qui sont le plus susceptibles de contenir des décalages, c'est-à-dire celles qui sont le plus marquées. Pour les paires de segments hétéromorphes négatifs (dont le ratio est compris entre 0 et 1), les plus marquées numériquement sont celles dont le ratio PIT est le plus éloigné de 1,0 (ou le plus près de 0). Pour les paires hétéromorphes positifs (dont le ratio est supérieur à 1,0), les plus marquées numériquement sont celles dont le ratio PIT est le plus élevé et donc le plus éloigné de 1,0 également. Après avoir criblé les paires hétéromorphes positives et négatives en fonction de leur ratio PIT, il est possible de sélectionner celles qui sont le plus marquées numériquement.

L'analyse des paires de segments les plus marquées positivement et négativement est motivée par l'hypothèse selon laquelle ces paires sont celles qui sont le plus susceptibles de contenir des décalages informationnels de traduction. Il est toujours possible d'élargir la recherche de décalages à d'autres paires moins marquées parmi les paires hétéromorphes négatives et positives, puisque toutes les paires de segments hétéromorphes des bitextes ont été triées en fonction de leur ratio PIT compilé par l'algorithme PML. Pour les besoins de la démonstration de la méthode, les dix paires de segments sélectionnées pour chaque texte ont fourni une quantité et une variété de décalages qui nous ont paru tout à fait acceptables, de sorte qu'il n'a pas été nécessaire d'analyser d'autres paires de segments dont le ratio PIT est moins marqué que celles qui ont été sélectionnées.

4. Repérage manuel des décalages de traduction

Les ratios PIT calculés par l'algorithme PML sont des données globales compilées à l'échelle des phrases ou des segments. Pour trouver les décalages, il est nécessaire d'examiner les expressions ou les unités significatives à l'intérieur des segments. Dans le cas des expressions, le calcul du ratio PIT permet de caractériser leur traduction : la traduction est littérale lorsque celui-ci est égal à 1 et elle est non

littérale lorsque celui-ci est inférieur ou supérieur à 1. L'heuristique de l'alignement des mots lexicaux fondé sur le sens (Poirier, 2016) sert à délimiter les expressions et unités significatives prises en compte dans le calcul du ratio PIT. La conséquence de la nature globale du ratio PIT calculé par paire de segments est que chacune peut contenir plusieurs décalages cumulatifs ou antagoniques.

Les décalages informationnels de traduction se répartissent en quatre groupes : les traductions mal alignées, les faux décalages, les décalages figés et les décalages libres, qui découlent des biais individuels ou non (erreurs de traduction) du sujet traduisant ou des biais culturels inhérents à l'opération de traduction pour les langues-cultures en contact⁴.

Les traductions mal alignées sont des segments du texte source qui n'ont pas de traduction ou des segments du texte cible qui ne correspondent pas, en tout ou en partie, aux segments du texte source avec lesquels ils sont physiquement alignés. Ces segments contiennent des décalages techniques qui ne représentent généralement pas des erreurs de traduction ni des biais de traduction individuels ou culturels. L'alignement fautif à l'origine de ces décalages doit être corrigé manuellement.

Les faux décalages sont des paires de segments dans lesquels une expression du texte source ou du texte cible contient des mots-formes équivoques sur le plan du statut de mot lexical ou de mot grammatical. Puisque l'algorithme n'est pas conçu pour discriminer différents emplois de mots-formes qui peuvent être équivoques, une expression qui contient un mot-forme considéré comme un mot grammatical, mais employé comme un mot lexical (ou l'inverse), peut donner lieu erronément à un décalage informationnel de traduction.

Les deux autres types de décalages que permet de repérer le ratio PIT illustrent de réelles variations du volume d'information dans une partie ou dans la totalité du segment du texte cible par comparaison avec une partie ou la totalité du segment du texte source. Ces variations du volume d'information, bien connues en traductologie, sont souvent décrites comme des différences de concentration (v. Ballard, 2006) qui se manifestent par des développements dans le cas des expansions syntaxiques ou par des réductions dans le cas des concentrations lexicales.

4. On trouve aussi dans les paires de segments isomorphes (dont le ratio PIT est égal à 1,0) des faux négatifs pouvant contenir des décalages antagoniques (une traduction non littérale sous forme d'ajout et une traduction non littérale sous forme d'omission, par exemple). L'analyse de ces décalages furtifs n'est pas abordée dans le présent article.

On distingue les décalages de traduction pour lesquels la variation du volume d'information entre les expressions source et cible provient d'une différence structurelle ou compositionnelle entre les langues, et ceux pour lesquels la variation du volume d'information provient d'un biais individuel du sujet traduisant, ou à tout le moins de son intervention dans le cas d'erreurs de traduction, ou qui illustrent des contraintes ou des biais culturels dans l'opération de traduction. Wecksteen-Quinio, Mariaule et Lefebvre-Scodeller (2015) utilisent les adjectifs « libres » et « figés » pour répartir les procédés de traduction obligatoires et facultatifs. Nous les utilisons à notre tour pour différencier les décalages informationnels figés, donc obligatoires et attribuables à des différences structurelles ou compositionnelles entre les langues, des décalages libres, donc librement choisis par le sujet traduisant ou qui manifestent une intervention de sa part, qu'elle soit délibérée ou non.

L'analyse manuelle des segments hétéromorphes criblés à l'aide du ratio PIT permet d'isoler à l'intérieur des segments les décalages libres et les décalages figés. Pour distinguer leur caractère libre ou figé, nous avons eu recours au critère de la traduction littérale concurrente. En vertu de ce critère, si une traduction non littérale est utilisée là où il est possible d'utiliser une traduction littérale acceptable, le décalage créé par la traduction non littérale est alors considéré comme un décalage libre ou facultatif. À l'inverse, si aucune traduction littérale ne peut être substituée à une traduction non littérale, le décalage de la traduction non littérale est alors considéré comme un décalage figé ou obligatoire, c'est-à-dire qu'il provient vraisemblablement de l'exercice de contraintes lexicogrammaticales ou de servitudes linguistiques (Vinay et Darbelnet, 1977 [1958]). Pour aider dans l'analyse et la vérification d'une traduction littérale ou non littérale, nous avons eu recours au service DeepL⁵, en partant du principe que le résultat du transfert linguistique effectué par la traduction automatique est fortement conventionnel et qu'il privilégie les traductions littérales. Par ailleurs, comme nous avons déjà eu l'occasion de montrer que les résultats des services de traduction automatique (statistique) sont principalement défaillants sur le plan de l'acceptabilité grammaticale et stylistique en langue d'arrivée (Poirier, 2017a), nous avons étroitement surveillé la qualité de ces traductions, ce qui a entraîné le rejet de certaines traductions littérales qui ont été jugées incorrectes.

5. La société DeepL GmbH offre gratuitement des services d'intelligence artificielle pour la traduction automatique de textes (voir DeepL GmbH, n. d.).

Les décalages libres comprennent aussi un certain nombre de sous-types de décalages. On compte notamment les erreurs de traduction, les choix délibérés du sujet traduisant (les biais individuels) ainsi que les biais culturels. Il est souvent difficile de départager les deux derniers types de décalage comme il est difficile de départager la part de l'individuel et celle du collectif dans toute activité créative. En traduction, les biais culturels sont souvent des points de vue qui découlent légitimement d'une situation culturelle donnée, soit celle de la langue-culture d'un groupe, d'un pays, d'une société participant à l'opération de traduction. Par définition, la traduction implique deux langues-cultures, donc nécessairement la gestion par le sujet traduisant de points de vue ou de biais culturels différents. En ce sens, la gestion des contradictions et des biais culturels par le traducteur consiste à gérer les légitimités de chaque langue-culture. S'il est difficile d'éliminer les biais individuels et culturels de tout acte de traduction, c'est probablement parce que ceux-ci font partie intégrante de l'opération de traduction. Dans tous les cas d'analyse de ce type de décalage, il n'est pas nécessaire de valider leur interprétation auprès du sujet traduisant puisqu'ils s'établissent à partir de la traduction comme résultat et de la réception du texte traduit.

5. Le discours de John F. Kennedy et de sa traduction en français

Le premier texte que nous avons analysé est le discours d'investiture du président des États-Unis John Fitzgerald Kennedy, prononcé le 21 janvier 1961. Le bitexte qui a servi de départ à l'analyse a été constitué avec les versions anglaise et française publiées sur le site *John F. Kennedy Presidential Library and Museum*.

Dans le format textuel du langage Python (qui est similaire au format HTML du bitexte créé avec LogiTerm^{MC}), le discours compte 1367 et 1474 mots-formes en anglais et en français répartis dans 53 paires de segments alignés. L'analyse de l'algorithme PML donne comme résultat 15 paires de segments isomorphes, 12 paires hétéromorphes négatives (volume d'information du segment cible inférieur à celui du segment source) et 26 paires hétéromorphes positives (volume d'information du segment cible supérieur à celui du segment source). En moyenne, le degré de précision de la traduction est de 1,04, ce qui veut dire que le texte traduit contient légèrement plus de mots lexicaux que le texte source.

L'annexe I contient les résultats complets de l'analyse fournie par l'algorithme et présente les données relatives aux 12 segments

hétéromorphes négatifs et aux 26 segments hétéromorphes positifs qui sont cotés et triés par ordre croissant du ratio PIT. Le tableau A de l'annexe I réunit les segments hétéromorphes négatifs tandis que le tableau B réunit les segments hétéromorphes positifs. Les tableaux 1 et 2 ci-dessous contiennent pour leur part les premiers segments hétéromorphes des tableaux fournis en annexe qui ont été criblés suivant la méthode décrite à la section 3. Les données de chaque segment comprennent son numéro (N°), qui correspond à son ordre d'occurrence dans le bitexte, son ratio PIT (PIT) compilé à partir du dénombrement des mots lexicaux dans le segment source (A) et du dénombrement des mots lexicaux dans le segment cible (F). Rappelons qu'en vertu de la méthode décrite à la section 3, les segments sont triés par ordre croissant du ratio PIT, lequel figure à la deuxième ligne des tableaux.

Tableau 1. Ratios et variables PIT des 5 premiers segments hétéromorphes négatifs

N°	25	46	7	2	42
PIT	0,75	0,85	0,86	0,87	0,88
A	4	13	7	15	8
F	3	11	6	13	7

Tableau 2. Ratios et variables PIT des 5 derniers segments hétéromorphes positifs

N°	33	21	18	50	10
PIT	1,33	1,38	1,40	1,43	1,50
A	19	23	12	7	2
F	22	27	14	10	3

Les tableaux 1 et 2 montrent que les cinq premières paires de segments hétéromorphes négatifs ont un ratio PIT qui s'établit entre 0,75 et 0,88 tandis que les cinq dernières paires de segments hétéromorphes positifs ont un ratio PIT qui s'établit entre 1,33 et 1,50. Ces segments correspondent aux segments traduits hétéromorphes dont le ratio PIT est le plus éloigné de 1; les autres segments hétéromorphes étant tous plus près de 1,0, soit dans l'intervalle de 0,89 à 0,99 pour les segments négatifs et dans l'intervalle de 1,01 à 1,32 pour les segments positifs. Le bitexte contient un plus grand nombre de segments

hétéromorphes positifs que négatifs, ce qui semble confirmer l'hypothèse selon laquelle les textes traduits de l'anglais au français sont en général plus longs, et probablement plus explicites.

Après suppression ou correction des paires de segments mal alignés (décalages techniques), la deuxième étape a consisté à analyser qualitativement les décalages qui se trouvent dans les segments hétéromorphes qui ont été criblés par l'algorithme. Cette analyse est décrite ci-après dans les tableaux 3 et 4. La première colonne contient la paire de segments, le nombre de mots lexicaux et le ratio PIT calculé par l'algorithme PML, et la deuxième contient la traduction de DeepL ainsi qu'une analyse qualitative des décalages observés.

L'utilisation de DeepL a donné lieu à quelques non-sens qui n'ont pas été retenus dans la délimitation des possibilités de traduction littérale. C'est le cas de l'expression « à son heure de danger maximum » proposée pour rendre « in its hour of maximum danger » (segment 46). De sorte que « lorsqu'elle était grandement menacée » a été traité comme un décalage figé (obligatoire) et non pas libre.

Tableau 3. Analyse des décalages dans les paires de segments hétéromorphes négatifs

Paires de segments et ratio PIT

(25) We dare not tempt them with weakness. = Nous ne les tenterons pas par notre faiblesse. (3/4=0,75)

(46) in the long history of the world only a few generations have been granted the role of defending freedom in its hour of maximum danger = tout au long de l'histoire du monde seules quelques générations ont été appelées à défendre la liberté lorsqu'elle était grandement menacée (11/13=0,85)

(7) We dare not forget today that we are the heirs of that first revolution. = Nous ne saurions oublier aujourd'hui que nous sommes les héritiers de cette première révolution. (7/6=0,86)

Analyse qualitative

DeepL : Nous n'osons pas les tenter par la faiblesse.

Faux décalage : l'auxiliaire *dare* est traité comme un verbe lexical par l'algorithme.

DeepL : dans la longue histoire du monde, seules quelques générations se sont vu accorder le rôle de défendre la liberté à son heure de danger maximum.

Décalages figés (3) : le nom *few* est rendu par le déterminant *quelques* ; traduction non littérale des expressions *have been granted the role of defending* et *its hour of maximum danger*

DeepL : Nous n'osons pas oublier aujourd'hui que nous sommes les héritiers de cette première révolution.

Faux décalage : négation sans *pas* en français

(2) We observe today not a victory of party but a celebration of freedom--symbolizing an end as well as a beginning--signifying renewal as well as change.
= Nous célébrons aujourd'hui non pas la victoire d'un parti, mais celle de la liberté, qui symbolise aussi bien une fin qu'un commencement, et qui marque un renouveau et un changement.
(13/15=0,87)

(42) The graves of young Americans who answered the call to service surround the globe. = Les jeunes Américains qui ont répondu à cet appel reposent dans le monde entier. (8/7=0,88)

DeepL : Nous observons aujourd'hui non pas une victoire du parti, mais une célébration de la liberté - symboliser une fin aussi bien qu'un début - signifiant un renouveau aussi bien qu'un changement.

Décalages figés (2) : traduction non littérale de l'expression *observe a celebration of sth* rendue par *célébrer qqch* ; traduction non littérale de l'expression *as well as [...] as well as [...]* par *aussi bien [...] que [...]* et qui ne contient pas de répétition de la locution conjonctive.

DeepL : Les tombes des jeunes Américains qui ont répondu à l'appel au service entourent le globe.

Décalages figés (2) : traduction non littérale de *call to service* par *appel* ; traduction non littérale de *world* par *monde entier*

Décalage libre : non-traduction de *graves* rendu par compensation avec le verbe *reposer* qui fait référence au fait que le sujet du verbe est enterré. Une façon élégante de rendre le sens en anglais de façon plus économe.

Comme pour le tableau 3, le tableau 4 (page suivante) décrit dans la première colonne l'application de la méthode quantitative sur les segments hétéromorphes positifs criblés dans le premier bitexte de notre corpus. Nous proposons dans la deuxième colonne un relevé qualitatif des décalages de traduction qui expliquent la variation du contenu informatif constaté entre le segment source et sa traduction.

Dans les 10 segments criblés qui ont été analysés dans les tableaux 3 et 4, on trouve 8 faux décalages et 21 décalages informationnels de traduction, dont 13 décalages figés et 8 décalages libres. Les faux décalages repérés par l'algorithme PML représentent du bruit dans le repérage des décalages de traduction, puisqu'ils ne découlent pas de l'opération de traduction d'un contenu informatif, mais d'un décompte inexact des mots lexicaux dans l'un ou l'autre des segments source et cible. Pour cette raison, ce sont des décalages techniques, attribuables à une difficulté précise de dénombrement des mots lexicaux.

Tableau 4. Analyse des décalages dans les paires de segments hétéromorphes positifs

Paires de segments et ratio PIT

(33) let both sides seek to invoke the wonders of science instead of its terrors = que les deux camps cherchent à évoquer les merveilles de la science plutôt que ses abominations (8/7=1,33)

(21) Let all our neighbors know that we shall join with them to oppose aggression or subversion anywhere in the Americas. = Que tous nos voisins sachent que nous nous unirons à eux pour faire face, ensemble, à toute agression ou subversion sur l'ensemble du continent américain. (11/8=1,38)

(18) If a free society cannot help the many who are poor, it cannot save the few who are rich = Si une société libre ne peut pas aider la multitude de personnes-vivant dans la pauvreté, elle ne peut pas sauver la minorité de personnes plus aisées (14/10=1,40)

(10) This much we pledge--and more. = Nous en faisons solennellement la promesse. (3/2=1,50)

Analyse qualitative

DeepL : que les deux parties cherchent à invoquer les merveilles de la science au lieu de ses terreurs.

Faux décalages (3) : le verbe *let* est traité comme un auxiliaire ; l'adjectif *both* est traité (à tort) comme un pronom ; la conjonction *plutôt* est traitée (à tort) comme un adverbe

DeepL : Que tous nos voisins sachent que nous nous joindrons à eux pour nous opposer à l'agression ou à la subversion n'importe où dans les Amériques.

Faux décalage : le verbe *let* est traité comme un auxiliaire

Décalages libres (2) : traduction non littérale du verbe *oppose* par *faire face* et du nom *Americas* par *continent américain*

Note : L'adverbe *anywhere* correspond au nom *l'ensemble* dans l'alignement des contenus informatifs.

DeepL : Si une société libre ne peut pas aider les nombreux pauvres, elle ne peut pas sauver les quelques riches.

Décalages figés (2) : traduction non littérale du nom *many* par *multitude de personnes* ; traduction non littérale du nom *few* par *minorité de personnes*

Décalages libres (2) : traduction non littérale de l'adjectif *poor* par l'expression *vivant dans la pauvreté* plutôt que par le nom *les pauvres* ; ajout de l'adverbe *plus* employé comme comparatif elliptique (intensificateur)

DeepL : C'est ce que nous promettons - et plus encore.

Décalages libres (3) : traduction non littérale de *This much* rendu par le pronom *en* ; ajout fautif de *solennellement* comme compensation à *more* ; traduction non littérale du verbe *pledge* par une locution verbale

(50) and so my fellow americans ask not
what your country can do for you ask
what you can do for your country = ainsi
mes chers compatriotes américains ne
demandez pas ce que votre pays peut faire
pour vous mais bien ce que vous pouvez
faire pour votre pays (10/7=1,43)

DeepL : et donc mes compatriotes américains ne demandent pas ce que votre pays peut faire pour vous demander ce que vous pouvez faire pour votre pays.

Faux décalages (2) : le verbe principal *do* a été traité (à tort) comme un auxiliaire à deux reprises

Décalage figé : l'adjectif *cher* est obligatoire dans le contexte des formules d'appel usuelles

Note : le mot-forme *bien* est traité comme un adverbe alors qu'il forme une conjonction complexe avec *mais* ; ce faux décalage positif est annulé par la répétition de *ask* en anglais qui n'est pas rendu directement en français, ce qui crée un segment isomorphe furtif (voir 4.0).

On a trouvé près de 2 décalages de traduction pour chacun des 10 segments analysés dans le discours de Kennedy. Ce nombre élevé peut s'expliquer par la méthode de criblage que nous proposons, qui vise à rationaliser le repérage des décalages. Les segments hétéromorphes qui font l'objet de l'analyse manuelle sont en fait les segments dont le ratio PIT est le plus éloigné de 1 (donc des segments isomorphes), lesquels sont le plus susceptibles de contenir des décalages de traduction.

Les faux décalages sont attribuables principalement à l'ambiguïté de certains mots-formes qui peuvent s'employer comme mots lexicaux ou comme mots grammaticaux. Un cas est attribuable à l'utilisation de la particule « ne » sans « pas » pour exprimer la négation, qui fait en sorte que l'expression de la négation n'est pas reconnue dans le calcul du volume informatif du segment. Un autre cas est attribuable à la non-reconnaissance par l'algorithme d'une expression constituée de plusieurs mots-formes. À titre d'exemple, le mot-forme « plutôt », qui est le plus souvent un adverbe lorsqu'employé seul, est dans le segment 33 un élément constitutif de la locution conjonctive « plutôt que », à l'intérieur de laquelle le statut de « plutôt » est indiscernable. Même problème dans le segment 50 avec l'adverbe « bien », utilisé dans la locution conjonctive « bien que ».

Rappelons que les décalages figés ont un caractère obligatoire et, par conséquent, s'imposent au sujet traduisant. Le plus souvent, ils proviennent de l'impossibilité en langue cible d'établir une correspondance littérale avec l'emploi d'un mot-forme de la langue source

(ou de l'agrammaticalité d'une telle correspondance, ce qui revient au même), comme c'est le cas des décalages figés observés dans les segments 2, 42 et 18 (tableaux 3 et 4). Dans le cas du segment 50 (l'emploi de l'adjectif « cher »), le décalage est attribuable au caractère obligatoire de l'adjectif en question dans une formule figée de salutation. Le caractère figé en langue cible est souvent attribuable à la normalisation d'usages linguistiques stéréotypés. Pour les segments 46 et 10, les décalages sont associés à l'emploi en langue source d'une expression idiomatique non compositionnelle dont il importe pour bien la traduire d'en rendre uniquement le sens.

Les décalages libres expriment des choix effectués par le ou les traducteurs qui divergent d'une traduction littérale qui serait pourtant correcte. Dans les segments hétéromorphes analysés, on trouve 6 occurrences de décalages libres. Une première occurrence dans le segment 42⁶ témoigne d'une forme de naturalisation de la formulation en langue source par une compensation. Ce type d'aménagement semble découler d'un souci d'économie puisqu'il constitue la réduction d'une traduction littérale plausible (« Les tombes des » est omis). Un deuxième décalage libre dans le segment 21 témoigne de l'importance accordée à la signification en contexte, qui rend obligatoire une formulation non littérale avec l'emploi d'un verbe complexe (« faire face à qqch » pour rendre « to oppose sth ») rendant plus justement le sens exprimé par le mot correspondant dans le segment source. Dans le deuxième décalage du segment 21 (traduction de « anywhere in the Americas » par « sur l'ensemble du continent américain »), il semble que ce soit surtout le souci d'employer un registre plus soutenu qui ait motivé le choix de traduction, la formulation littérale « n'importe où dans les Amériques » relevant d'un registre plus familier ou de la langue courante. Un troisième exemple de décalage libre dans le segment 18 illustre un biais qui témoigne de la dissonance linguistique et culturelle qui est à l'œuvre dans l'opération de traduction. En effet, un premier décalage libre s'applique à la traduction euphémisante non littérale de l'adjectif « poor » (rendu par « personnes vivant dans la pauvreté ») tandis qu'un second témoigne d'un mouvement inverse qui consiste à intensifier l'adjectif « rich » au moyen d'un comparatif elliptique (« plus ») qui vient souligner et renforcer l'opposition

6. Soit la non-traduction de « graves » rendu par compensation avec le verbe « reposer » dans : « The graves of young Americans who answered the call to service surround the globe. » = « Les jeunes Américains qui ont répondu à cet appel reposent dans le monde entier. » (8/7=0,88)

entre les riches et les pauvres (« the few who are rich » rendu par « la minorité de personnes plus aisées »). Un dernier décalage libre dans le segment 10 témoigne d'une erreur de traduction : un contenu informatif (« solennellement ») a été ajouté comme compensation de « more »⁷. Ce dernier cas montre que le criblage des segments hétéromorphes par l'algorithme PML peut aussi servir à relever des erreurs de traduction dans le cadre du contrôle de la qualité et de la révision des textes traduits⁸.

6. Le discours de Theresa May et sa traduction en français

Le deuxième discours analysé est celui de la première ministre britannique Theresa May sur la sortie de la Grande-Bretagne de l'Union européenne, qui a été prononcé le 17 janvier 2017. Le bitexte qui a servi à l'analyse a été constitué avec la transcription du discours publiée par le magazine *Time*⁹ (Anonyme, 2017). Pour ce qui est de la version française, nous avons utilisé la traduction publiée par le magazine *L'Obs* (Riché, 2017)¹⁰. Le fait qu'elle ait été produite par une personne vivant vraisemblablement en France, à l'extérieur du Royaume-Uni, explique qu'elle soit susceptible de contenir, à première vue, un plus grand nombre de décalages de traduction¹¹.

Dans le format textuel du langage Python, le discours compte 6396 et 6881 mots-formes en anglais et en français répartis dans 336 paires de segments alignés. L'analyse des segments traduits par l'algorithme PML donne comme résultat 116 segments isomorphes, 64 segments hétéromorphes négatifs et 156 segments hétéromorphes

7. Dans le segment « This much we pledge—and more » qui a été traduit par « Nous en faisons solennellement la promesse ».

8. Cela dit, compte tenu des bruits (faux décalages) et de la difficulté de les distinguer des décalages figés et des autres types de décalages libres, la méthode ne permet toutefois pas encore de repérer exclusivement les erreurs de traduction.

9. La version anglaise officielle (May, 2017a) lui est pour l'essentiel identique. Sauf pour ce qui est de l'emploi de l'apostrophe et des traits d'union, aucune différence n'a été observée dans la version publiée par le magazine *Time*.

10. Une traduction française est aussi disponible sur le site du gouvernement britannique (May, 2017b), mais elle a été publiée le 3 février 2017, soit un peu plus de deux semaines après le discours de Theresa May. Cette traduction officielle contient des différences par rapport à la traduction du journaliste Pascal Riché (2017), pour la plupart des améliorations. La version de Riché revêt cependant un plus grand intérêt pour montrer l'utilité de la méthode que nous proposons.

11. Une partie des décalages contenus dans la traduction du journaliste peut aussi s'expliquer par la rapidité d'exécution, puisque la traduction a été publiée le jour même où le discours a été prononcé.

positifs. En moyenne, le degré de précision de la traduction est de 1,07, ce qui veut dire que le bitexte contient légèrement plus de mots lexicaux en langue cible qu'en langue source.

L'annexe II présente les résultats complets de l'analyse fournie par l'algorithme et les données relatives aux 64 segments hétéromorphes négatifs et aux 156 segments hétéromorphes positifs. Le tableau A de l'annexe II réunit les segments hétéromorphes négatifs et le tableau B réunit les segments hétéromorphes positifs. À l'instar des tableaux 1 et 2 présentés précédemment, les tableaux 5 et 6 se rapportent aux premiers segments hétéromorphes des tableaux fournis en annexe qui ont été criblés suivant la méthode décrite à la section 3.

Tableau 5. Ratios et variables PIT des 5 premiers segments hétéromorphes négatifs

N°	254	240	314	323	17
PIT	0,50	0,67	0,67	0,67	0,69
A	2	3	3	3	13
F	1	2	2	2	9

Tableau 6. Ratios et variables PIT des 5 derniers segments hétéromorphes positifs

N°	141	238	264	276	329
PIT	2,00	2,00	2,00	2,00	2,00
A	1	2	1	2	2
F	2	4	2	4	4

Les tableaux 5 et 6 montrent que les cinq premières paires de segments hétéromorphes négatifs ont un ratio PIT qui s'établit entre 0,50 et 0,69 tandis que les cinq dernières paires de segments hétéromorphes positifs ont un ratio qui s'établit à 2,00. Ces segments correspondent aux segments traduits hétéromorphes dont le ratio PIT est le plus éloigné de 1; les autres segments hétéromorphes étant tous plus près de 1,0, soit dans l'intervalle de 0,69 à 0,99 pour les segments négatifs et dans l'intervalle de 1,01 à 2,00 pour les segments positifs.

De manière plus prononcée que dans le bitexte précédent, le bitexte du discours de Theresa May contient un plus grand nombre de segments hétéromorphes positifs que négatifs, ce qui semble confirmer l'hypothèse selon laquelle les textes traduits de l'anglais au français sont en général plus longs, et probablement plus explicites.

Comme pour le premier discours, après suppression ou correction des paires de segments mal alignés (décalages techniques), la deuxième étape a consisté à analyser qualitativement les décalages qui se trouvent dans les segments hétéromorphes qui ont été criblés par l'algorithme. Cette analyse est décrite ci-après dans les tableaux 7 et 8, suivant le même format que pour les tableaux 3 et 4 décrits précédemment.

Tableau 7. Analyse des décalages dans les paires de segments hétéromorphes négatifs

Paires de segments et ratio PIT

(254) That is only right = C'est légitime.
(1/2=0,50)

(240) Strengthening the United Kingdom =
Renforcer le Royaume-Uni (3/2=0,67b)

(314) And another thing that's important =
Un autre point est capital (2/3=0,67c)

(323) And the country comes together = Et
le pays se rassemble (3/2=0,67d)

(17) One that gets us the right deal abroad
but also ensures we get a better deal for
ordinary working people at home. = Un
plan propre à améliorer la situation du pays
dans le monde et celle des gens ordinaires
sur le territoire. (9/13=0,69)

Analyse qualitative

DeepL : C'est **tout à fait juste**

Décalage libre : erreur de traduction par omission de l'adverbe « only » qui sert d'intensificateur.

DeepL : **Renforcement du Royaume-Uni**

Décalage figé : traduction non littérale des deux mots lexicaux « United Kingdom » par le mot-forme « Royaume-Uni », qui est traité comme un mot lexical.

DeepL : **Et une autre chose qui est importante.**

Faux décalage : l'adjectif « autre » est traité comme son mot-forme homonyme qui est classé comme un pronom, c.-à-d. un mot grammatical.

DeepL : **Et le pays se rassemble.**

Décalage figé : le verbe à particule « come together », formé d'un verbe et d'un adverbe, est traduit par un verbe simple.

DeepL : **Une qui nous permet d'obtenir le bon accord à l'étranger, mais qui nous permet également d'obtenir un meilleur accord pour les travailleurs ordinaires dans leur pays d'origine.**

Décalages libres (2) : erreur de traduction par contresens de l'expression « the right deal abroad » = « améliorer la situation du pays dans le monde » ; erreur de traduction par omission de l'élément « working dans ordinary working people » = « gens ordinaires ».

Note : « home »/« territoire » ; bonne traduction culturelle ou situationnelle.

Le tableau 8 décrit l'application de la méthode quantitative sur les segments criblés et une analyse qualitative des décalages. Cette fois, l'analyse s'applique aux segments hétéromorphes positifs.

Tableau 8. Analyse des décalages dans les paires de segments hétéromorphes positifs

Paires de segments et ratio PIT

(141) And that is what we will deliver. = C'est ce que nous allons mettre en œuvre. (2/1=2,00)

(238) Certainty wherever possible. = De la certitude, chaque fois que c'est possible. (4/2=2,00)

(264) And that is what I intend to do. = Et c'est ce que j'entends faire. (2/1=2,00)

(276) more of it makes us all more prosperous. = plus il est important, plus nous serons prospères. (4/2=2,00)

(329) And let that be the legacy of our time. = Et que ce soit là la grande œuvre de notre époque. (4/2=2,00)

Analyse qualitative

DeepL : *Et c'est ce que nous allons livrer.*

Décalage libre : traduction non littérale de « deliver » par « mettre en œuvre ».

Note : « deliver » peut se rendre par de nombreux verbes simples : « réaliser », « exécuter », « fournir », « offrir », etc.

DeepL : *Certitude chaque fois que c'est possible.*

Décalage figé : traduction non littérale de « wherever » par « chaque fois », une locution conjonctive formée de deux mots lexicaux.

DeepL : *Et c'est ce que j'ai l'intention de faire.*

Faux décalage : le verbe « do » est traité à tort comme un auxiliaire.

DeepL : *plus de cela nous rend tous plus prospères.*

Décalage figé (2) : les deux déterminants « more » sont correctement traduits par l'adverbe « plus », qui est traité comme un mot lexical.

Décalage libre : explicitation de « more of it » avec « plus il est important » plutôt que « plus il y en a » ou « ?plus de cela » (acceptabilité incertaine de cette expression fournie par DeepL).

Note : le pronom « it » renvoie à la notion de commerce.

DeepL : *Et que ce soit l'héritage de notre temps.*

Décalage libre : erreur de traduction du déictique ; le mot-forme « that », qui est pronom en anglais, est traduit par l'adverbe « là », qui est un mot lexical.

Décalage libre : traduction non littérale de « legacy » par deux mots lexicaux.

Dans les segments criblés inclus dans les tableaux 7 et 8, on trouve 2 faux décalages et 12 décalages de traduction, dont 5 figés et 7 libres. Ces données font état d'une proportion élevée de décalages libres, qui comptent pour 50 % de tous les décalages du deuxième bitexte comparativement à 33 % dans le premier bitexte. Cette différence peut s'expliquer par le choix d'une traduction non officielle pour la création du bitexte, cette dernière étant susceptible de contenir de plus grands écarts par rapport à une traduction officielle, et ce, pour les raisons évoquées précédemment. Un autre élément qui a pu jouer est le nombre beaucoup plus élevé de segments (336 par rapport à 53) que l'on trouve dans le deuxième bitexte. L'efficacité de la méthode de criblage semble en effet augmenter en fonction du nombre de segments. De fait, le criblage des segments hétéromorphes permet d'isoler les segments qui sont le plus susceptibles de présenter des décalages de traduction, et ceux-ci sont plus nombreux quand le bitexte contient un nombre plus élevé de segments.

Dans les décalages libres du deuxième discours, on trouve 5 erreurs de traduction et 2 expansions facultatives. Ces dernières concordent avec l'hypothèse d'explicitation, qui consiste, dans le texte cible, à décrire avec plus de moyens formels que le texte source le sens ou le message transmis. Un exemple d'explicitation est le décalage libre du segment 276 («more of it» rendu par «plus il est important»). La traduction non littérale de «legacy» par «grande œuvre» fournit par ailleurs un exemple d'explicitation de ce qui aurait pu être traduit littéralement par «héritage».

Comme l'indiquent les décalages relevés dans les segments sélectionnés, le nombre d'erreurs de traduction semble élevé dans le deuxième bitexte, qui, rappelons-le, repose sur une traduction réalisée par un journaliste qui n'est pas un professionnel de la traduction et qui a dû, de surcroît, respecter des délais très serrés pour produire son texte. Ces décalages libres renseignent à la fois sur les contraintes linguistiques, mais aussi sur les biais individuels du sujet traduisant et sur les biais culturels des langues-cultures en contact. Dans le cas de la traduction de «deliver» par «mettre en œuvre» dans le segment 141¹², il s'agit d'un décalage libre, parce qu'on trouve de nombreuses solutions de traduction avec un seul mot lexical. Cela dit, le décalage informationnel introduit par la locution «mettre en œuvre» constitue tout de même une erreur de traduction sur le plan discursif. En

12. «And that is what we will deliver» traduit par «C'est ce que nous allons mettre en œuvre».

effet, l'emploi de cette formule abstraite, qui relève dans une certaine mesure de la langue administrative, est mal avisé si l'on considère que la locutrice concernée semblait chercher à éviter les formules creuses et abstraites. L'emploi du verbe « faire » aurait donc été plus juste dans ce cas-ci.

Nous avons aussi relevé d'autres erreurs de traduction qui témoignent d'un biais culturel dans la reformulation du message source. Une première a consisté à omettre l'adverbe « tout à fait » (ou un autre intensificateur) dans le segment 254¹³, ce qui donne à penser que le sujet traduisant (un Français, vraisemblablement) a voulu réduire la force illocutoire de l'énoncé « That is only right » parce que celui-ci renvoie à la sortie par les Britanniques de l'Union européenne. Si on replace le segment en contexte, toutefois, la légitimité dont il est question dans le discours de Theresa May est celle des négociations (en Grande-Bretagne) qui doivent avoir lieu sur le plan de sortie de l'Union européenne. L'interprétation de l'omission de l'intensificateur « only » pouvant être rendu par « tout à fait » par le sujet traduisant doit donc être nuancée. Sans doute a-t-il jugé (avec raison) qu'il n'était pas nécessaire d'insister sur ce point puisque des négociations allaient devoir de toute façon avoir lieu. Par contraste, pour la première ministre britannique, il est normal et même souhaitable d'insister sur le fait que ses adversaires pourront faire entendre leur désaccord, ce qui contribue à légitimer sa propre démarche consensuelle (après un référendum serré, ne l'oublions pas), dans la mesure où ceux qui s'opposent à son projet auront eu l'occasion de se faire entendre. Dans l'opération de traduction, on peut penser que ces subtilités discursives n'ont pas à être rendues puisque les destinataires de la traduction ne participeront pas aux négociations; la démarche de traduction vise pour ainsi dire à les prendre à témoin puisqu'ils n'ont pas voix au chapitre.

Une deuxième erreur qui va dans le sens d'un biais culturel se manifeste par l'omission des « travailleurs ordinaires » dans le segment 17¹⁴. Ici, sans doute que l'omission marque le changement de point de vue et le biais culturel sur les personnes qui seront touchées par la sortie de l'Union d'un pays membre. Dans le texte source, Theresa May met l'accent sur le fait que les travailleurs restent sa priorité dans ses négociations avec les autres pays. Du point de vue d'un Français, la sortie de l'Union européenne aura des conséquences

13. « That is only right » traduit par « C'est légitime ».

14. L'expression « ordinary working people at home » est rendue par « gens ordinaires sur le territoire ».

pour tous les citoyens ordinaires d'un pays membre et non pas que pour ses travailleurs. Cet exemple illustre bien la différence de point de vue qui infléchit en quelque sorte la traduction d'un segment particulier du texte source. Deux autres erreurs de traduction illustrent une méconnaissance des difficultés et des principes de base de la traduction, à savoir, dans le segment 17, un contresens qui découle de la mauvaise interprétation du mot « deal »¹⁵, et, dans le segment 329¹⁶, la traduction littérale du déictique « that » qui n'est généralement pas traduit par un déictique comme « là », lequel peut être perçu comme redondant en français.

7. Conclusion et perspectives

La méthode décrite dans le présent article mise sur le criblage des segments hétéromorphes d'un corpus parallèle de traductions grâce au ratio PIT. Les résultats obtenus avec cette méthode et que nous avons présentés pour les 10 segments d'un bitexte qui sont les plus susceptibles de contenir des décalages donnent à penser qu'elle s'acquitte honorablement de la recherche de décalages informationnels de traduction, lesquels sont essentiels pour la connaissance de différents phénomènes liés à la traduction (phraséologie contrastive, difficultés de traduction et leurs solutions idiosyncrasiques) et aux opérations cognitives de la traduction (erreurs, biais individuels et culturels). L'examen et la catégorisation de 33 décalages informationnels ont montré que les décalages figés et libres sont des catégories pertinentes pour l'étude de ces phénomènes. Une meilleure connaissance de ceux-ci peut contribuer à l'avancement de la traductologie et fournir du matériel instructif dans la formation des traducteurs professionnels.

Même s'il est difficile de mesurer avec exactitude l'efficacité de la méthode quantitative, elle est assurément plus efficace que la lecture intégrale des textes traduits comme méthode de criblage des segments hétéromorphes, lesquels sont susceptibles de contenir des décalages libres et des erreurs de traduction. Cela étant établi, il reste aussi à montrer dans d'autres bitextes et avec un nombre plus important de segments l'hypothèse que les 20 paires de segments décrits ici ont permis de formuler, à savoir que les décalages informationnels de traduction permettent de caractériser les contrastes d'usage entre les langues (les servitudes avec lesquelles les traducteurs doivent

15. L'expression « right deal abroad » rendue par « améliorer la situation du pays ».

16. « And let that be the legacy of our time. » = « Et que ce soit là la grande œuvre de notre époque. »

composer) et d'acquérir une meilleure connaissance des opérations de traduction telles qu'elles peuvent être mises au jour par des emplois linguistiques attestés dans des corpus parallèles.

Par ailleurs, la méthode que nous proposons a jeté un éclairage sur certaines particularités de la traduction de discours politiques. L'examen des décalages de traduction informationnels libres dans le discours de Theresa May a révélé que les décalages induits par des biais culturels peuvent être doublement perçus comme des erreurs de traduction ou des décalages légitimes, et que la différence entre les deux réside dans la reconnaissance ou non de la légitimité du biais culturel du sujet traduisant dans l'interprétation du texte source. Par contre, comme la traduction de ce discours a été faite par un journaliste et non par un professionnel de la traduction, il est difficile de déterminer si l'équivoque dans la catégorisation de ces décalages est le propre des discours politiques ou des textes non traduits par des professionnels de la traduction. L'absence d'information sur l'auteur ou les auteurs du discours de John F. Kennedy, et la nature intraculturelle de sa traduction (aux États-Unis), ne nous offrent aucun ancrage pour examiner le statut des décalages induits par des biais culturels.

Un discours comme celui de John F. Kennedy met toutefois en évidence une autre particularité de la traduction des discours politiques, qui est celle de la traduction d'actes de langage axés sur la persuasion et sur l'action (plutôt que sur l'information, le divertissement ou l'explication). L'étude des décalages libres est susceptible de jeter un éclairage intéressant sur la traduction de ce type particulier de discours. Dans les quelques exemples que nous avons pu relever, il semble que la traduction tende à omettre ces actes de langage, ceux-ci n'ayant plus de pertinence dans la restitution du message dans un contexte discursif distinct de la situation du texte source. Il faudrait voir jusqu'à quel point ce procédé est répandu. Le petit nombre de décalages que nous avons analysés ne nous permet que d'émettre une hypothèse à ce sujet.

Malgré l'intérêt du calcul du ratio PIT comme méthode d'évaluation quantitative des segments traduits, il subsiste un nombre important de difficultés d'application de l'algorithme, qui donnent lieu à de faux résultats positifs que nous avons appelés « faux décalages ». Une voie intéressante d'amélioration de la méthode consisterait à réduire le nombre de faux décalages obtenus par le criblage des segments hétéromorphes avec l'algorithme PML. Différentes solutions sont envisageables, mais il faudra essentiellement qu'une analyse lexico-

grammaticale automatique soit faite dans la catégorisation pertinente et contextuelle des mots-formes en mots lexicaux et en mots grammaticaux.

Références

- Anonyme (2017). « Read Theresa May's Speech Laying Out the U.K.'s Plan for Brexit ». *Time*. January 17. [<https://time.com/4636141/theresa-may-brexits-speech-transcript/>].
- Bakker, Matthijs, Koster, Cees et Kitty Van Leuven-Zwart (2011). « Shifts ». In M. Baker et G. Saldanha, dir. *Routledge Encyclopedia of Translation Studies*, 2nd ed. Londres et New York, Routledge, p. 269-274.
- Ballard, Michel (2006). « À propos des procédés de traduction ». *Palimpsestes*. Numéro hors série, p. 113-130. [<http://palimpsestes.revues.org/386>].
- Chesterman, Andrew (2004). « Hypotheses about Translation Universals ». In G. Hansen, K. Malmkjær et D. Gile, dir. *Claims, Changes and Challenges in Translation Studies: Selected contributions from the EST Congress, Copenhagen 2001*. Amsterdam, John Benjamins, p. 1-14.
- DeepL GmbH (n.d.). *DeepL Translator*. [<https://www.deepl.com/translator/>].
- Delisle, Jean et Marco A. Fiola (2013). *La traduction raisonnée. Manuel d'initiation à la traduction professionnelle de l'anglais vers le français*. 3^e éd. Ottawa, Les Presses de l'Université d'Ottawa.
- Fraser, Bruce (1975). « Hedged performatives ». In P. Cole et J. Morgan, dir. *Speech Acts, Syntax and Semantics*. Academic Press, New York, p. 187-210.
- Gallego-Hernández, Daniel (2018). « Testing a Methodological Framework for Retrieving Parallel Texts in the Domain of Business Translation ». *Perspectives*, 26, 1, p. 39-53.
- Gottesman, Catherine (2006). « Quelques réflexions sur la traduction littérale ». *Études de linguistique appliquée*, 141, p. 95-106. [<https://www.cairn.info/revue-ela-2006-1-page-95.htm#>].
- Kennedy, John F. (1961). « Inaugural Address, 20 January 1961 ». *John F. Kennedy Presidential Library and Museum*. [<https://www.jfklibrary.org/learn/about-jfk/historic-speeches/inaugural-address>]. [Retranscription en anglais et traduction officielle en français].
- May, Theresa (2017a). « Plan for Britain: Prime Minister's Speech on Brexit Negotiating Objectives ». *GOV.UK*. 17 January. [<https://www.gov.uk/government/speeches/the-governments-negotiating-objectives-for-exiting-the-eu-pm-speech>].
- May, Theresa (2017b). « Un "Global Britain" : discours prononcé par la Première ministre Rt Hon Theresa May, Membre du Parlement, Lancaster House, 17 Janvier 2017 ». *GOV.UK*. 3 février. [<https://www.gov.uk/government/speeches/the-governments-negotiating-objectives-for-exiting-the-eu-pm-speech.fr>].
- Poirier, Éric (2016). « Meaning-based Content Word Alignment Heuristic ». *Proceedings of the 8th International Conference on Management of Digital EcoSystems*. ACM, p. 208-214.

- Poirier, Éric (2017a). « Entre comparaison et raison : la qualité de la traduction automatique ». *Circuit*, 133. [<http://www.circuitmagazine.org/dossier-133/entre-comparaison-et-raison-la-qualite-de-la-traduction-automatique>].
- Poirier, Éric (2017b). « A Comparison of Three Metrics for Detecting Cross-linguistic Variations in Information Volume and Multiword Expressions between Parallel Bitexts ». *Proceedings of EUROPHRAS 2017*, London, UK, November 13-14, 2017, p. 1-10.
- Polguère, Alain (2003). *Lexicologie et sémantique lexicale : notions fondamentales*. Montréal, Les Presses de l'Université de Montréal.
- Riché, Pascal (2017). « Comment va se dérouler le Brexit : le discours de Theresa May traduit en français » *L'Obs*. 17 janvier. [<https://www.nouvelobs.com/brexit/20170117.OBS3936/comment-va-se-derouler-le-brexit-le-discours-de-theresa-may-traduit-en-francais.html>].
- Villada Moirón, Begoña et Jörg Tiedemann (2006). « Identifying Idiomatic Expressions Using Automatic Word-alignment ». *Proceedings of the Workshop on Multi-word-expressions in a multilingual context*. [<https://www.aclweb.org/anthology/volumes/W06-24/>].
- Vinay, Jean-Paul et Jean Darbelnet (1977 [1958]). *Stylistique comparée du français et de l'anglais*. Paris, Didier.
- Wecksteen-Quinio, Corinne, Mickaël Mariaule et Cindy Lefebvre-Scodeller (2015). *La traduction anglais-français. Manuel de traductologie pratique*. Louvain-la-Neuve, De Boeck.

Éric André Poirier

Département des langues modernes et de traduction
Université du Québec à Trois-Rivières
Trois-Rivières (Québec), Canada
Eric.Poirier@uqtr.ca

Annexe I
Cotation et classement automatiques des segments hétéromorphes
Discours de John F. Kennedy

Tableau A. Ratios et variables PIT des 12 segments hétéromorphes négatifs

N°	25	46	7	2	42	14	41	52	53	3	16	8
PIT	0,75	0,85	0,86	0,87	0,88	0,89	0,89	0,91	0,91	0,92	0,94	0,97
A	4	13	7	15	8	18	9	11	22	12	16	38
F	3	11	6	13	7	16	8	10	20	11	15	37

Tableau B. Ratios et variables PIT des 26 segments hétéromorphes positifs

N°	27	36	49	23	43	38	32	20	31	1	24	28
PIT	1,03	1,04	1,08	1,09	1,09	1,10	1,11	1,12	1,14	1,16	1,17	1,17
A	32	24	13	32	35	10	18	8	7	19	23	12
F	33	25	14	35	38	11	20	9	8	22	27	14

N°	6	9	17	51	40	13	44	29	30	33	21	18
PIT	1,18	1,20	1,21	1,22	1,25	1,30	1,31	1,33	1,33	1,33	1,38	1,40
A	32	24	13	32	35	10	18	8	7	19	23	12
F	33	25	14	35	38	11	20	9	8	22	27	14

N°	50	10
PIT	1,43	1,50
A	7	2
F	10	3

Annexe II
Cotation et classement automatiques des segments hétéromorphes
Discours de Theresa May

Tableau A. Ratios et variables PIT des 64 segments hétéromorphes négatifs

N°	254	240	314	323	17	37	94	333	31	280	118	7
PIT	0,50	0,67	0,67	0,67	0,69	0,75	0,75	0,75	0,78	0,78	0,79	0,80
A	2	3	3	3	13	4	8	4	9	18	24	5
F	1	2	2	2	9	3	6	3	7	14	19	4
N°	15	74	157	70	188	5	129	145	204	208	261	305
PIT	0,80	0,80	0,80	0,82	0,82	0,83	0,83	0,83	0,83	0,83	0,83	0,83
A	10	5	5	11	11	12	12	12	6	12	6	6
F	8	4	4	9	9	10	10	10	5	10	5	5
N°	334	135	51	89	125	207	243	260	62	34	142	42
PIT	0,83	0,85	0,86	0,86	0,86	0,86	0,86	0,86	0,87	0,88	0,88	0,89
A	12	27	7	7	7	7	7	7	15	8	8	19
F	10	23	6	6	6	6	6	6	13	7	7	17
N°	159	216	299	102	128	174	217	26	46	93	168	221
PIT	0,89	0,89	0,90	0,91	0,91	0,91	0,91	0,92	0,92	0,92	0,92	0,92
A	9	9	10	11	11	11	22	13	12	12	12	12
F	8	8	9	10	10	10	20	12	11	11	11	11
N°	12	220	225	252	268	270	56	117	127	259	274	295
PIT	0,93	0,93	0,93	0,93	0,93	0,93	0,94	0,94	0,94	0,94	0,94	0,94
A	14	14	14	14	15	15	17	17	16	17	17	17
F	13	13	13	13	14	14	16	16	15	16	16	16
N°	21	320	47	202								
PIT	0,95	0,95	0,96	0,96								
A	19	22	23	25								
F	18	21	22	24								

Tableau B. Ratios et variables PIT des 156 segments hétéromorphes positifs

N°	27	123	173	29	68	182	190	215	311	48	8	33
PIT	1,04	1,05	1,05	1,06	1,06	1,06	1,06	1,06	1,06	1,07	1,08	1,08
A	27	20	20	18	16	16	17	16	17	15	13	12
F	28	21	21	19	17	17	18	17	18	16	14	13
N°	115	134	151	161	308	218	263	103	108	110	171	178
PIT	1,08	1,08	1,08	1,08	1,08	1,09	1,09	1,10	1,10	1,10	1,10	1,10
A	12	24	12	12	25	11	11	10	10	20	31	10
F	13	26	13	13	27	12	12	11	11	22	34	11
N°	266	193	197	213	230	1	18	24	38	65	124	255
PIT	1,10	1,11	1,11	1,11	1,11	1,12	1,12	1,12	1,12	1,12	1,12	1,12
A	41	9	18	9	9	8	16	8	8	17	8	16
F	45	10	20	10	10	9	18	9	9	19	9	18
N°	265	327	59	78	98	104	154	186	194	229	298	63
PIT	1,12	1,12	1,14	1,14	1,14	1,14	1,14	1,14	1,14	1,14	1,14	1,15
A	8	8	7	7	14	7	7	7	14	7	14	13
F	9	9	8	8	16	8	8	8	16	8	16	15
N°	136	66	80	81	97	100	119	206	231	249	256	294
PIT	1,15	1,17	1,17	1,17	1,17	1,17	1,17	1,17	1,17	1,17	1,17	1,17
A	13	6	6	12	6	6	12	12	6	6	6	6
F	15	7	7	14	7	7	14	14	7	7	7	7
N°	109	114	120	160	170	14	55	116	131	175	176	179
PIT	1,18	1,18	1,18	1,18	1,18	1,20	1,20	1,20	1,20	1,20	1,20	1,20
A	11	17	22	11	11	5	5	5	5	5	5	5
F	13	20	26	13	13	6	6	6	6	6	6	6
N°	275	292	297	306	307	326	19	144	164	165	39	50
PIT	1,20	1,20	1,20	1,20	1,20	1,20	1,21	1,21	1,21	1,21	1,22	1,22
A	5	5	5	15	5	5	14	14	14	14	9	27
F	6	6	6	18	6	6	17	17	17	17	11	33
N°	132	195	301	71	90	52	228	20	41	45	113	196
PIT	1,22	1,22	1,22	1,23	1,23	1,24	1,24	1,25	1,25	1,25	1,25	1,25
A	23	9	9	13	13	17	17	12	8	12	8	4
F	28	11	11	16	16	21	21	15	10	15	10	5
N°	245	246	277	291	300	303	313	331	146	10	138	309
PIT	1,25	1,25	1,25	1,25	1,25	1,25	1,25	1,25	1,26	1,27	1,27	1,27
A	4	4	12	12	8	8	28	4	19	11	15	11
F	5	5	15	15	10	10	35	5	24	14	19	14

ÉRIC ANDRÉ POIRIER

N°	64	72	192	232	236	23	191	3	6	16	43	86
PIT	1,29	1,29	1,29	1,29	1,29	1,30	1,30	1,33	1,33	1,33	1,33	1,33
A	7	7	7	7	14	7	10	10	6	3	3	3
F	9	9	9	9	18	13	13	8	4	4	4	4

N°	88	106	156	198	205	318	328	227	105	121	325	75
PIT	1,33	1,33	1,33	1,33	1,33	1,33	1,33	1,36	1,38	1,38	1,38	1,40
A	3	3	6	3	6	3	12	11	8	8	8	5
F	4	4	8	4	8	4	16	15	11	11	11	7

N°	95	53	4	61	85	177	223	239	332	224	222	107
PIT	1,40	1,44	1,50	1,50	1,50	1,50	1,50	1,50	1,50	1,54	1,60	1,64
A	5	9	2	6	2	2	2	2	2	13	5	11
F	7	13	3	9	3	3	3	3	3	20	8	18

N°	58	273	233	35	87	91	99	141	238	264	276	329
PIT	1,67	1,67	1,80	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00	2,00
A	3	3	5	1	1	2	2	1	2	1	2	2
F	5	5	9	2	2	4	4	2	4	2	4	4