Meta
Journal des traducteurs
Translators' Journal

META

L'autoévaluation appuyée sur l'outillage textométrique dans l'enseignement de la traduction

Jun Miao et André Salem

Volume 61, numéro 2, août 2016

URI : https://id.erudit.org/iderudit/1037759ar
DOI : https://doi.org/10.7202/1037759ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Miao, J. & Salem, A. (2016). L'autoévaluation appuyée sur l'outillage textométrique dans l'enseignement de la traduction. *Meta*, *61*(2), 255–275. https://doi.org/10.7202/1037759ar

Résumé de l'article

Dans un contexte d'enseignement de la traduction, nous comparons différentes traductions françaises d'un même texte original en anglais (le discours d'investiture prononcé par le président Barack Obama en 2009). Certaines des traductions ont été réalisées par des traducteurs professionnels, d'autres par des outils automatiques, d'autres enfin par des apprenants traducteurs. Les outils textométriques permettent de mettre en évidence les similitudes et différences entre les traductions proposées. L'utilisation de l'alignement du corpus au niveau du paragraphe permet de construire un réseau de correspondances à partir desquelles les calculs textométriques produisent des résultats particulièrement intéressants. L'*analyse verticale* des traductions permet de localiser des portions du texte original que les traducteurs ont traitées de manière similaire. En utilisant cette même approche, nous pouvons aussi localiser les fragments qui donnent lieu, à l'inverse, à des traductions différentes et explorer l'éventail des traductions proposées. Les ressources numérisées ainsi élaborées constituent une aide précieuse pour juger de la qualité du travail fourni par les apprenants. Elles leur fournissent, par ailleurs, un outil d'autoévaluation efficace.

L'autoévaluation appuyée sur l'outillage textométrique dans l'enseignement de la traduction

JUN MIAO
INALCO, Paris, France
miaojun@miaojun.net

ANDRÉ SALEM
Université Paris 3 – Sorbonne Nouvelle, Paris, France
salem@msh-paris.fr

RÉSUMÉ

Dans un contexte d'enseignement de la traduction, nous comparons différentes traductions françaises d'un même texte original en anglais (le discours d'investiture prononcé par le président Barack Obama en 2009). Certaines des traductions ont été réalisées par des traducteurs professionnels, d'autres par des outils automatiques, d'autres enfin par des apprenants traducteurs. Les outils textométriques permettent de mettre en évidence les similitudes et différences entre les traductions proposées. L'utilisation de l'alignement du corpus au niveau du paragraphe permet de construire un réseau de correspondances à partir desquelles les calculs textométriques produisent des résultats particulièrement intéressants. L'*analyse verticale* des traductions permet de localiser des portions du texte original que les traducteurs ont traitées de manière similaire. En utilisant cette même approche, nous pouvons aussi localiser les fragments qui donnent lieu, à l'inverse, à des traductions différentes et explorer l'éventail des traductions proposées. Les ressources numérisées ainsi élaborées constituent une aide précieuse pour juger de la qualité du travail fourni par les apprenants. Elles leur fournissent, par ailleurs, un outil d'autoévaluation efficace.

ABSTRACT

In the context of teaching translation courses, different French translations of the same English original text (President Barack Obama's 2009 inaugural address) are compared. The translations can be divided into those produced by professional translators, those produced by automated tools, and translation learners. Textometric tools allow us to explore the similarities and differences between translations. Using paragraph alignment makes it possible to build a comparison network from which the textometric calculations produce explicit results. The *vertical analysis* of translations highlights the portions of the original text which translators proceeded in the same way. Using this same approach, we can also, conversely, locate the fragments that leaded to different translations and explore the range of translations. The digital resources compiled during the process of comparison are invaluable for assessing the quality of translation work produced by learners. They also provide the latter with an effective self-assessment tool.

MOTS-CLÉS/KEYWORDS

enseignement de la traduction, autoévaluation, textométrie, corpus parallèles
translation teaching, self-assessment, textometrics, parallel corpora

1. Introduction

Les méthodes de traitement automatique des corpus électroniques font désormais partie intégrante des cursus d'enseignement dans la plupart des départements universitaires qui se préoccupent de l'étude des textes. À côté des recherches qui concernent des corpus particuliers (voir Frérot 2010), on note l'apparition de nombreux logiciels[1] d'aide à la traduction (ex. *Trados*, *Déjà Vu*, *Wordfast*, etc.). Ces logiciels évoluent continuellement pour tenter de s'adapter aux différentes tâches de traduction. Les outils de traduction automatique, encore très imparfaits, ont fait récemment des progrès importants. Ces progrès tiennent avant tout à un changement de stratégie qui s'est opéré dans la traduction du texte par les automates. Prenant le contrepied des méthodes qui tentaient de « comprendre » le texte pour le traduire ensuite, les logiciels modernes s'appuient sur l'existence de mémoires de traduction et sur d'immenses bases de données renfermant des traductions alignées, dont la qualité a été vérifiée par des humains[2].

Ces nouvelles possibilités constituent une aide précieuse pour l'activité de traduction. Elles facilitent à la fois la recherche d'équivalents lexicaux, la mise en parallèle de tournures idiomatiques et, par là même, la compréhension du texte source par les lecteurs non natifs. Les enseignants en traduction se doivent, de plus en plus, d'inclure dans leurs cours des formations à l'utilisation de ces nouveaux outils. Il faut comprendre, et faire comprendre aux apprenants, que ces nouveaux outils ne sont pas à même de fournir des résultats satisfaisants dans toutes les situations de traduction, et qu'ils proposent parfois des solutions complètement fautives. L'enseignement doit conduire les apprenants à distinguer par eux-mêmes, parmi toutes les possibilités de mise en correspondance offertes par les couples de langues, les solutions les plus appropriées au contexte. Il est donc primordial que les étudiants intègrent le fait que la copie pure et simple des résultats proposés par les machines ne peut constituer une solution acceptable dans toutes les situations.

De nombreux chercheurs (Israël 1999 ; Lee-Jahnke 2001 ; Tercedor-Sánchez, López-Rodriguez *et al.* 2005 ; Kiraly 2005 ; Valetopoulos 2012, etc.) soulignent l'importance de l'autoévaluation dans le cursus pédagogique. En contradiction avec les méthodes employées précédemment, pour lesquelles la critique a posteriori, faite par les enseignants, constituait la base de la pédagogie, les méthodes nouvelles incitent les étudiants à développer leurs capacités d'autoévaluation.

L'apprentissage vise avant tout à la réalisation d'une traduction de qualité. Le cas où l'étudiant recopie systématiquement une traduction préexistante dont la qualité est acceptable traduit sans doute un manque d'investissement personnel dommageable pour la formation de l'étudiant. Par contre, le « couper/coller » portant sur de larges portions de traductions erronées, manifeste une insuffisance chez l'apprenant dans la compréhension et l'analyse du texte. Il est impératif de signaler les lacunes de l'apprenant à celui-ci.

Notre travail rend compte d'une expérience de confrontation entre des traductions effectuées par des apprenants, des professionnels et des automates, dans le but de mettre au point des méthodes d'autoévaluation utilisables par les étudiants. La confrontation a été réalisée à l'aide d'outils textométriques utilisés en liaison avec des méthodes d'analyse traductologique plus traditionnelles. Les ressources numérisées, constituées à cette occasion, permettent aux étudiants de mieux cerner la variété de

procédés de traduction employés dans le corpus et d'évaluer par eux-mêmes le travail qu'ils ont fourni.

2. Confronter des traductions

À la suite du projet pionnier d'ICLE (International Corpus of Learner English) mené par Sylviane Granger à l'Université Louvain-La-Neuve depuis 1990, plusieurs groupes d'enseignants-chercheurs ont collecté des traductions d'apprenants afin d'étudier les problèmes rencontrés par ces derniers au cours de leur apprentissage de l'anglais (Granger 1998). Notons, par exemple, l'examen des traductions de Waddington (2001), le travail pédagogique de Lee-Jahnke (2001), les projets *Student Translation Archive* (Bowker et Bennison 2003), *ENTRAD* (Florén 2006) et *Russian Translation Learner Corpus* (Sosnia 2006). Ces travaux tentent de déterminer les difficultés fréquemment rencontrées par les apprenants, dans le but d'améliorer le contenu et le matériel d'enseignement du domaine.

Masschelein et Verschueren (2005) fournissent une recherche intéressante et solide sur *l'évaluation formative* qui s'oriente vers un apprentissage semi-autonome de la traduction. À l'aide d'un logiciel (*Markin*), ces chercheurs évaluent les exercices des étudiants avec plus de 90 codes (positifs et négatifs) selon des critères d'évaluation préalablement définis. Par exemple, le code MS signifie que l'étudiant a commis une faute morphologique, alors que STA souligne qu'il a y des problèmes stylistiques de tout genre. De manière similaire, à travers des extraits de textes comptant environ 350 mots chacun et des annotations faites à partir d'une typologie prédéfinie des erreurs, le projet européen *MeLLANGE* (Castagnoli, Ciobanu *et al.* 2009) entreprend de cerner les problèmes de traduction dans un environnement de corpus multilingues. Popescu-Belis *et al.* (2002) comparent pour leur part la traduction automatique, les traductions des étudiants et les traductions professionnelles. L'objectif de cette dernière étude est de mettre en évidence dans ces traductions des types d'erreurs grâce à des mesures statistiques permettant de montrer la corrélation entre la répartition des erreurs et les notes attribuées à chaque traduction.

Dans les recherches mentionnées ci-dessus, on peut dégager les points suivants :

- les enseignants occupent souvent un rôle central dans l'évaluation, les étudiants suivant leurs commentaires ;
- le processus d'annotation des usages (erronés ou acceptables) est un travail coûteux en temps et les critères d'évaluation sont souvent complexes ;
- la maîtrise simultanée des outils informatiques, la réalisation des différents exercices de traduction et la mise à profit des évaluations données par les enseignants durant le temps de formation sont difficiles pour les apprenants ; les textes abordés dans les exercices de traduction constituent souvent des extraits difficiles à relier avec l'intégralité du texte ;
- la plupart des efforts se concentrent sur les erreurs commises par les étudiants, au détriment de la recherche de possibilités de traduction optimales ;
- la remise en contexte partielle, réalisée dans la plupart des logiciels (ex.*WordSmith*, *AntConc*, etc.), ne donne pas une vision globale du texte qui est, cependant, indispensable pour connaître l'organisation du texte.

Dans l'enseignement de la traduction, il est important d'apprendre aux étudiants à juger de la qualité d'une traduction et à développer leurs capacités d'autoévaluation.

La comparaison collective des différentes traductions réalisées par un groupe d'apprenants à partir d'un même texte-source nous semble constituer une activité intéressante pour aborder ces questions. La comparaison inclut également des traductions proposées par des traducteurs professionnels, ainsi que des traductions réalisées par des procédures automatisées.

Notre programme d'enseignement ainsi conçu doit, tout à la fois, permettre aux apprenants de prendre conscience de leurs erreurs récurrentes, de leur signaler des solutions de traduction auxquelles ils n'avaient pas songé lors de la réalisation de leur propre traduction, de leur signaler aussi les erreurs grossières couramment commises par les automates de traduction afin d'accroître leurs facultés d'autoévaluation.

Pour traiter ces corpus de traduction, nous utilisons ici les méthodes de la textométrie. La *textométrie* rassemble une série de méthodes statistiques qui permettent de réorganiser formellement les séquences textuelles et d'effectuer des analyses statistiques portant sur leur vocabulaire. L'*analyse des spécificités* (Lafon 1980 ; Lafon 1984) articulée avec *l'analyse factorielle des correspondances* (Bénzécri 1973, 1977, 1981) permet de dresser des typologies à partir de textes et de repérer les formes que chacun d'entre eux emploie, ou sous-emploie, de manière privilégiée (voir plus loin les sections 3.1 et 4.1). La plupart de ces méthodes sont implantées dans le logiciel *Lexico3* (créé par l'équipe universitaire SYLED-CLA2T, sous la direction du professeur André Salem, Université Paris 3). Parmi d'autres logiciels de textométrie, *Lexico3* offre également une méthode de *cartographie textuelle* qui fournit une localisation visuelle des occurrences de chaque unité textuelle étudiée dans l'ensemble du corpus (Lamalle et Salem 2002). On utilise cette approche cartographique pour la détection des accords et des discordances entre différentes traductions.

2.1. Les documents rassemblés

Dans un cursus de formation à la traduction (master première année - G1, deuxième année - G2) donné à l'Institut National des Langues et Civilisations Orientales (INALCO) à Paris, nous avons enseigné devant deux groupes d'étudiants aux origines langagières et culturelles diverses. Les cours portaient sur la traduction pragmatique, principalement liée au couple anglais-français pour les exemples. Par ailleurs, pour stimuler l'intérêt des étudiants, nous les avons encouragés à découvrir par eux-mêmes les phénomènes du domaine à l'aide d'outils informatiques.

Afin de permettre aux étudiants de constituer un corpus textométrique de taille réduite, nous avons opté pour un texte court : le discours d'investiture d'Obama en 2009[3] (2417 occurrences/tokens). Les étudiants disposaient, par ailleurs, de cinq traductions françaises de ce même texte, publiées dans la presse : *Le Monde*, *Libération*, *La Croix*, *RFI* et *Maison-Blanche*[4], ainsi que de traductions en ligne, réalisées par des automates *Google Translation*, *Systran*, *Reverso*[5]. Après une initiation des deux groupes d'apprenants aux traitements textométriques des corpus (une dizaine d'heures, environ) comparant les cinq traductions professionnelles et le discours original, nous leur avons demandé de traduire par eux-mêmes le discours étudié en dehors des séances de cours et de noter, simultanément, les difficultés qu'ils rencontraient. Chaque étudiant a donc constitué un dossier comprenant : sa propre traduction et un rapport sur les principales difficultés rencontrées. Les différentes traductions produites par les étudiants ont intégré, par la suite, notre corpus d'étude.

À l'aide de l'outil d'alignement *Alignator*[6], nous avons constitué un corpus aligné dans lequel chaque paragraphe de chacune des traductions est mis en parallèle avec la partie originale correspondante. Le corpus rassemble donc cinq grands types de documents : le document original ; 5 traductions de ce document réalisées par des traducteurs professionnels ; 3 traductions réalisées par des automates ; 16 traductions effectuées par des apprenants de première année et 9 traductions effectuées par des apprenants de deuxième année. Malgré le nombre d'étudiants non homogène dans les deux groupes (16 c. 9), les méthodes textométriques telles que l'analyse des spécificités et l'analyse des facteurs de correspondances, basées sur le calcul probabiliste (voir ci-dessous), peuvent nous fournir des résultats pertinents. Ce corpus aligné, constitué de 34 textes (1 texte original + 33 traductions), a été appelé *Obama1* (voir tableau 1).

Tableau 1
Les 34 textes du corpus *Obama1*

Groupes	Nombre		
TS	1	texte source (en anglais)	Barack Obama's 2009 presidential inaugural address
G0	3	traductions automatiques	logiciels : *Google*, *Systran* et *Reverso*
G1	16	étudiants (groupe G1)	réalisées par les étudiants de première année
G2	9	étudiants (groupe G2)	réalisées par les étudiants de deuxième année
G3	5	traductions professionnelles	*La Croix*, *Libération*, *Le Monde*, *RFI*, *Maison-Blanche*

Lors de la correction des travaux, nous avons été amenés à nous poser deux séries de questions :

a) Questions sur la comparaison des traductions :
- En quoi les traductions professionnelles diffèrent-elles des traductions effectuées par des apprenants ? En quoi sont-elles meilleures ?
- Peut-on distinguer des niveaux de compétence dans le travail de traduction en fonction de l'année d'apprentissage ?
- Peut-on corréler le niveau de la traduction effectuée au niveau de maîtrise du français par chacun des étudiants ?
- La qualité des traductions automatiques permet-elle de distinguer ces dernières traductions des traductions concurrentes réalisées par des humains ?

b) Questions sur le recours à l'outil informatique dans l'évaluation des traductions :
- Quelles sont les possibilités d'utilisation des outils informatiques dans ce type d'évaluation ?
- Peut-on utiliser ces outils dans des programmes d'évaluation et d'autoévaluation lors du processus d'enseignement ?

C'est pour tenter de répondre à ces questions que nous avons procédé à des analyses textométriques du corpus *Obama1* à l'aide de l'outil *Lexico3*.

2.2. Les deux groupes d'apprenants

Le groupe des étudiants de première année (G1) se partage en deux parties égales entre étudiants français et étudiants étrangers. Celui des étudiants de deuxième année (G2) ne compte qu'une seule étudiante dont le français est la langue maternelle, les autres étudiants étant originaires de cultures très diverses (arabophones, turcophones, etc.).

Le cours est donné en français. À l'exception d'un étudiant qui éprouve des difficultés, la plupart des étudiants possèdent bien le français, du moins à l'oral. Nous notons que la majorité des étudiants du groupe G1 possèdent un diplôme supérieur à la licence et que trois étudiants possèdent déjà un master. Les apprenants du groupe G2 sont presque tous titulaires d'un master.

Afin d'établir une atmosphère de confiance dans le groupe, nous procédons à l'anonymisation des copies en remplaçant les noms des étudiants par des identificateurs de type *xyL*, où :

- *x* indique l'année d'étude (1 ou 2)
- *y* la nationalité de l'étudiant (1 - français ; 2- étranger)
- *L* constitue un identificateur pour chacun des étudiants (A, B, C, …)

Groupe 1 : 11A, 11C, 11D, …
Groupe 2 : 21A, 22B, 22C, …

3. Analyse quantitative du corpus Obama1

Durant le cours, nous calculons avec les étudiants les principales caractéristiques textométriques pour le document original, en anglais (voir tableau 2), puis pour chacun des autres textes.

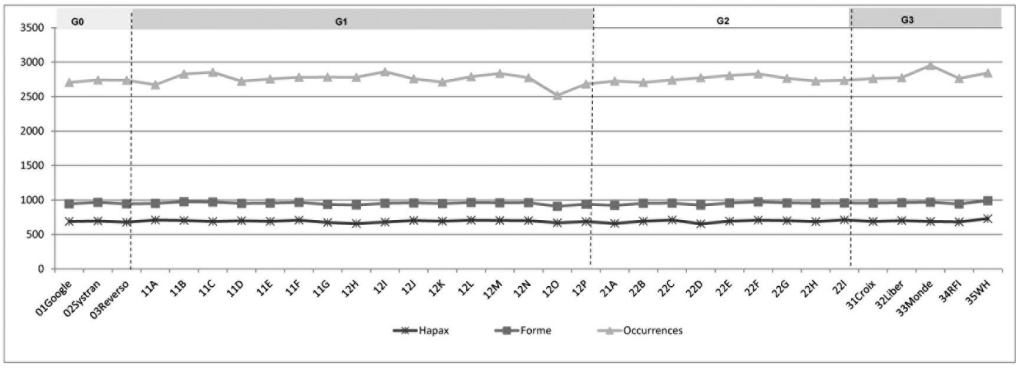
Tableau 2
Principales caractéristiques du discours d'investiture de B. Obama (2009)

Partie	Occurrences	Formes	Hapax	N. Fmax	Fmax
Original	2417	927	645	122	*the*

On note que le texte original est relativement court avec seulement 2417 occurrences (token) au total, dont 927 formes (types) sont des formes différentes. Les formes apparues une seule fois (hapax) atteignent 645. Dans ce texte, l'article défini *the* est la forme la plus fréquente (Fmax), avec 122 fois apparitions (voir la rubrique du N. Fmax).

Les caractéristiques textométriques calculées à partir de chacun des textes ne sont pas toujours directement comparables lorsqu'il s'agit de textes rédigés dans des langues différentes (original anglais c. traductions françaises, par exemple). Par contre, ces caractéristiques deviennent comparables pour des textes rédigés dans une même langue. La confrontation directe de caractéristiques textométriques calculées à partir de différentes traductions d'un même texte source va nous permettre de les comparer utilement.

FIGURE 1
Principales caractéristiques textométriques pour chacune des 33 traductions



Avant d'examiner les données quantitatives obtenues à partir des traductions, nous avons interrogé les étudiants sur les résultats qu'ils attendaient de cette comparaison. Leurs réponses s'accordaient en général sur l'idée qu'une bonne traduction doit posséder un vocabulaire plus riche (plus de formes différentes et plus d'hapax) qu'une traduction de qualité inférieure.

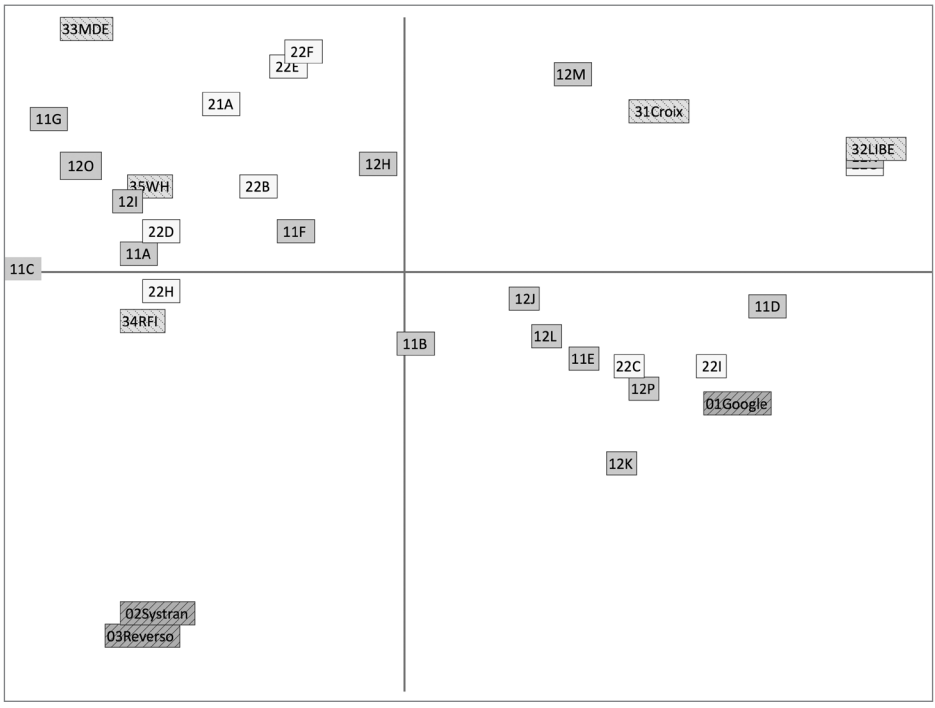
Dans l'analyse comparée des principales caractéristiques textométriques (figure 1), nous constatons cependant que le nombre d'occurrences (marqué par les triangles), le nombre de formes différentes (marqué par les carrés) et celui des hapax (marqué par les étoiles) sont, à quelques exceptions près, assez proches. L'étudiant 12O montre un vocabulaire plus pauvre ; la traduction du *Monde*, un vocabulaire plus varié. À ce moment de l'analyse, nous relevons peu de divergences entre les traductions automatiques et les traductions humaines. Cela est aussi le cas pour les traductions professionnelles et les traductions des étudiants. De plus, les caractéristiques relatives aux deux groupes d'apprenants ne présentent que peu de différences entre elles. À l'intérieur du groupe d'étudiants de première année, la variété semble un peu plus grande, en ce qui concerne le nombre des occurrences.

Il est donc nécessaire d'approfondir les recherches avec des outils de comparaison plus élaborés.

3.1. Typologie sur les traductions

L'analyse factorielle de correspondances (AFC) nous permet de mettre en évidence les principales oppositions pouvant exister dans le corpus des traductions. C'est une méthode statistique d'analyse des données mise au point par Benzécri (1973, 1977, 1981) destinée au traitement des tableaux de données où les valeurs sont positives et homogènes comme les tableaux de contingence. Cette méthode réduit la complexité des données en synthétisant au maximum les informations sur des correspondances entre les variables (non pas les valeurs absolues). Dans un corpus tel que le nôtre, nous obtenons des informations sur l'organisation du vocabulaire. La figure 2 nous fournit une première typologie (sur le plan des facteurs 1 et 2) portant sur les différentes traductions. Pour établir cette typologie, nous avons construit un tableau constitué par les décomptes des 1418 formes de fréquence supérieure à 5 dans les 33 traductions[7]. C'est ce tableau lexical que nous avons ensuite soumis à l'analyse.

Figure 2
Typologie à partir des traductions françaises du corpus *Obama1*



On trouve, sur la figure 2, les principaux résultats issus de cette analyse. Les couleurs permettent de distinguer les différents types de traduction (*traductions automatiques* : gris foncé avec des lignes continues obliques ; *apprenants de première année* : gris clair ; *apprenants de deuxième année* : blanc ; *traductions professionnelles* : gris clair avec des lignes discontinues obliques).

- Groupe A : les traductions automatiques *Systran* et *Reverso*, isolées dans le cadran inférieur gauche de la figure ;
- Groupe B : dans le cadran inférieur droit, la traduction *Google*, entourée par plusieurs traductions d'étudiants ;
- Groupe C : trois traductions professionnelles (*le Monde*, *la Maison-Blanche* et *le RFI*) ainsi que la majorité des traductions G2 et certaines traductions G1[8], en haut à gauche de la figure.
- Groupe D : deux traductions professionnelles sur le haut de la figure (*la Croix* et *Libération*), autour desquelles viennent s'agréger plusieurs traductions produites par des étudiants.

Au vu de ce qui précède, on peut avancer l'hypothèse que l'AFC aide à distinguer différents niveaux de traduction : les traductions automatiques, excepté celles de *Google*, apparaissent comme différentes des traductions humaines et les traductions professionnelles se retrouvent proches les unes des autres, dans le haut de la figure. Les étudiants, surtout ceux de première année (12J-12L-11E-12P-12K-11D-22C-22I), semblent s'être inspirés de la traduction fournie par *Google*, ceci est encore plus vraisemblable pour quelques étudiants dont le français n'est pas la langue maternelle.

Les traductions des apprenants ayant obtenu les meilleures notes se regroupent dans le cadran supérieur gauche de la figure. Il s'agit pour la plupart de travaux rendus par des étudiants de deuxième année, disposant souvent d'une certaine expérience de traduction.

Deux apprenants, 12H et 11B, de première année, dont l'un admet avoir utilisé systématiquement la traduction fournie par *Google translation* pour réaliser la seconde partie de sa propre traduction, occupent une position centrale.

En nous référant à la position relative de chaque traduction étudiante par rapport aux traductions professionnelles, nous pouvons supposer que les étudiants n'ont pas utilisé les mêmes ressources. L'étudiant 12M est plus proche de *la Croix*, 12N et 22G se rapprochent plutôt de *Libération* dont ils paraissent s'être largement inspirés pour la traduction de certaines parties ; 12I et 22H ont plutôt utilisé les traductions de la *Maison-Blanche* et de *RFI*[9].

De ce qui précède, nous voyons que l'AFC permet de dresser une première typologie des traductions. L'analyse des emprunts massifs à d'autres traductions, présentes dans le corpus, nous permettra de vérifier plus avant la qualité de chaque traduction.

3.2. Localisation des séquences répétées

Devant une traduction réalisée dans un cadre pédagogique, l'enseignant doit pouvoir reconnaître deux situations distinctes :

1) l'apprenant fournit des solutions de traduction qu'il a lui-même élaborées, en s'appuyant éventuellement sur des outils existants. Le résultat final traduit à la fois son niveau de compétence global et les difficultés qu'il a rencontrées dans cette expérience particulière ;
2) l'apprenant utilise de manière systématique des solutions fournies par les logiciels de traduction automatique, ou par d'autres traductions préexistantes, ce qui ne permet de juger ni de son niveau propre ni de ses progrès.

II est donc important, dans le cadre de l'évaluation d'un travail de traduction, d'être à même de repérer, si possible par des moyens automatiques, le taux d'utilisation directe de traductions proposées par les traducteurs automatiques. Le calcul des segments répétés (voir Salem 1986) fournit des solutions particulièrement adaptées à ce genre d'interrogation. Pour un texte donné, l'algorithme fournit la liste de séquences de plusieurs formes répétées à l'identique dans plusieurs endroits du corpus. Dans le cas d'un corpus comme le nôtre qui rassemble des traductions effectuées par plusieurs types de traducteurs (professionnels, apprenants et automates), il est très peu probable que des traducteurs distincts produisent de longues séquences parfaitement identiques.

Ainsi, dans notre corpus, la duplication massive de la séquence :

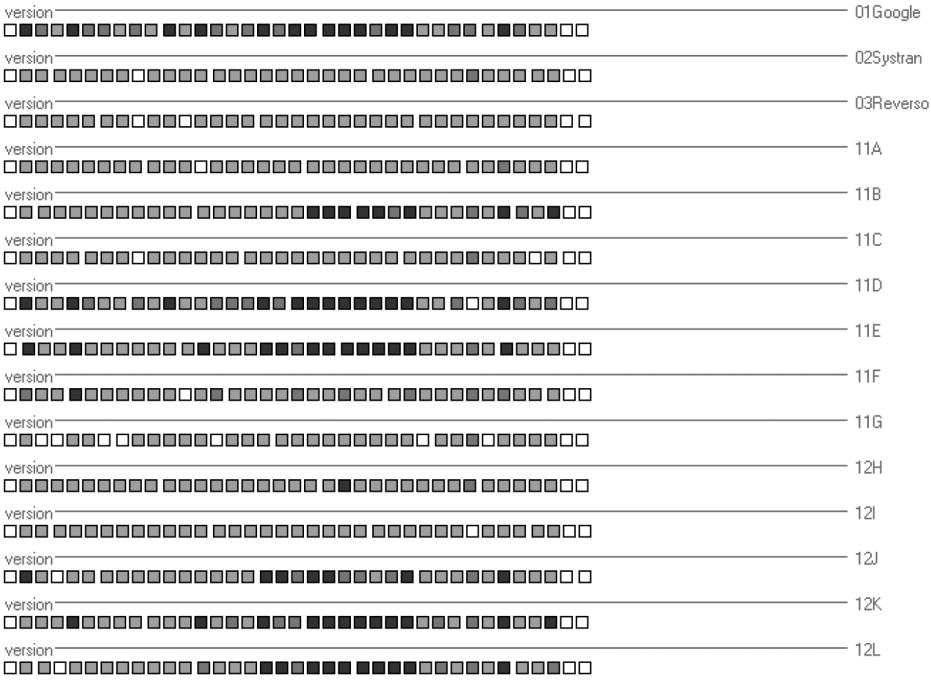
> Je remercie le président Bush pour ses services rendus à la nation, ainsi que pour la générosité et la coopération dont il a fait preuve tout au long de cette (transition /passation de pouvoirs).
>
> L'original en anglais est : *I thank President Bush for his service to our nation, as well as the generosity and cooperation he has shown throughout this transition.*

ne peut être considérée comme une simple coïncidence pouvant résulter de travaux indépendants et simultanés. En effet, près de 30 occurrences apparaissent sans aucune

altération dans onze des traductions du corpus, dont une fois dans la traduction fournie par l'automate *Google*. Si le sens présent dans le texte source était le même, les manières disponibles pour le rendre dans la langue cible étaient, à priori, relativement variées[10]. L'hypothèse de la recopie pure et simple par des moyens de duplication (*copier/coller* ou retranscription à partir d'un document déjà traduit) est beaucoup plus vraisemblable. Car, notons que cette même séquence a été retenue par deux des cinq traducteurs professionnels (*Le Monde* et *Libération*).

Le repérage de segments répétés (séquences de formes répétées plusieurs fois dans le corpus de manière identique) permet de mettre en évidence des coïncidences entre différentes traductions. Le calcul de la proportion des séquences répétées communes à deux traductions censées avoir été réalisées de manière indépendante peut nous aider à localiser des portions de texte dans lesquelles l'utilisation du *couper/coller* doit être considérée comme l'élément moteur de l'élaboration de la traduction : plus les séquences identiques sont longues, plus elles sont nombreuses, plus l'hypothèse d'une coïncidence accidentelle doit être écartée.

Figure 3
Extrait de la localisation des segments répétés du corpus *Obama1*



La figure 3 nous permet d'apprécier globalement les taux de duplication calculés à partir des différentes traductions[11]. Nous avons constitué une unité (*Tgen*)[12] qui rassemble toutes les occurrences situées au début d'une séquence de cinq formes, répétée cinq fois au moins dans le corpus. Une carte des sections a été établie pour le corpus. Les lignes horizontales foncées permettent d'isoler chacune des traductions dont l'identificateur est repérable sur la droite. Chaque carré représente un paragraphe

aligné du corpus. Le calcul de spécificités (voir plus loin section 4.1) permet d'apprécier l'abondance relative des séquences répétées sélectionnées dans chacune des sections du corpus. Une couleur claire traduit la simple présence du *Tgen*. Plus la couleur est foncée, plus l'abondance des segments répétés est jugée spécifique dans la section considérée et plus nous pouvons considérer que la traduction présente de fortes similitudes avec d'autres traductions présentes dans le corpus.

La forte présence dans la traduction *Google* de séquences communes à un grand nombre de traductions remises par les apprenants constitue une présomption supplémentaire du recours systématique à ce premier texte par certains apprenants pour produire les traductions demandées.

L'analyse de la localisation des répétitions segmentales permet de tirer plusieurs conclusions supplémentaires. Nous notons que les segments sélectionnés pour constituer notre *Tgen* (longueur et fréquence supérieures ou égale à cinq) apparaissent plus fréquemment chez les apprenants étrangers de première année (12J, 12K, 12L, par exemple). Cependant, deux étudiants dont le français est la langue maternelle (11D et 11E) emploient également ces segments de manière massive. Ces dernières traductions sont celles que nous avions repérées autour de la traduction *Google* dans la typologie obtenue à l'aide de l'AFC (voir section précédente). L'hypothèse selon laquelle les traductions ont été produites à l'aide d'un recours systématique à la fonctionnalité *couper/coller* se confirme. Cette même méthode permet également de vérifier que d'autres apprenants (11A, 11F et 12I, par exemple) n'ont eu recours à cette facilité qu'à de rares occasions.

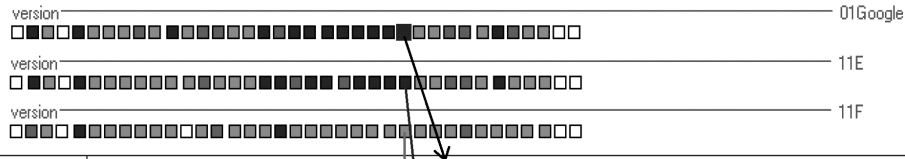
Comme nous l'avons signalé dans la section précédente, la traduction 11B (voir Figure 2) présente des caractéristiques particulières. La localisation des séquences répétées délimite ici assez nettement deux zones différentes dans le travail de traduction (voir Figure 3) : la première moitié présente peu de carrés foncés, ce qui souligne le caractère plutôt original de la traduction des paragraphes concernés. Dans la deuxième moitié, au contraire, le recours aux outils de traduction automatique, en l'occurrence *Google translate,* apparaît avoir été massivement utilisée.

L'affichage de la carte des sections permet un accès direct à chacun des paragraphes de chacune des traductions. Nous voyons sur la figure 4 que la traduction fournie par l'apprenant 11E ne s'écarte que très faiblement de la traduction réalisée par le traducteur automatique *Google translate* tandis que la traduction 11F contient des corrections plus nombreuses. L'étudiant 11F a corrigé certaines erreurs de syntaxe apparues dans la traduction automatique et ces corrections démontrent une prise en charge plus importante du travail de traduction.

Après une discussion de groupe avec les apprenants, l'enseignant peut insister sur les passages qui ont posé le plus grand nombre de problèmes aux traducteurs, repérer les hésitations et les maladresses dans chaque traduction. Les étudiants apprennent à identifier leurs faiblesses en comparant leur propre travail avec celui des autres.

Figure 4

Traductions du paragraphe 26, fournies par *Google translation* et par les apprenants 11E et 11F[11]



Google translation	<para=01026>Pour les habitants des pays pauvres, nous promettons de travailler à vos côtés pour que vos fermes prospèrent et que l'eau potable coule, de nourrir les corps affamés et les esprits voraces. Et à ces pays qui comme le nôtre bénéficient d'une relative abondance, nous disons que nous ne pouvons plus nous permettre l'indifférence à la souffrance en dehors de nos frontières, nous ne pouvons consommer les ressources planétaires sans nous soucier des conséquences. Car le monde a changé et nous devons changer avec lui. #
11E	<para=11E26>Pour les habitants des pays pauvres, nous promettons de travailler à vos côtés pour que vos fermes prospèrent et que **les eaux saines coulent**, de nourrir les corps affamés et les esprits voraces. Et à ces pays qui comme le nôtre bénéficient d'une relative abondance, nous disons que nous ne pouvons plus nous permettre l'indifférence à la souffrance en dehors de nos frontières, nous ne pouvons consommer les ressources planétaires sans nous soucier des conséquences. Car le monde a changé et nous devons changer avec lui. #
11F	<para=11F26>**Aux peuples** des pays pauvres, nous promettons de travailler à vos côtés pour **faire en sorte** que vos fermes prospèrent et que l'eau potable coule, pour nourrir les corps affamés et les esprits voraces. Et à ces **nations** qui comme **la** nôtre bénéficient d'une relative abondance, nous disons que **nous ne pouvons plus nous permettre d'être indifférents aux souffrances à l'extérieur de nos frontières, ni consommer les ressources mondiales** sans nous soucier des conséquences. Car le monde a changé, et nous devons changer avec lui. #

4. Analyse traductologique

4.1. Spécificités pour chaque groupe

Nous allons tenter, dans cette dernière partie, de mieux caractériser les écarts entre les groupes de traductions obtenus à l'aide de l'AFC (section 3.1). L'analyse des spécificités nous permet de dégager les formes et les segments les plus utilisés par chacun des types de traduction et de les regrouper en quatre groupes à partir des résultats de l'AFC.

Nous commençons par calculer les spécificités de chacun des groupes par rapport à l'ensemble du corpus[13]. Tous les écarts mis en évidence par la méthode des spécificités ne présentent pas le même intérêt pour caractériser les procédés spécifiques de traduction. Certains de ces écarts trouvent leur origine dans des problèmes technologiques, notamment dans des problèmes spécifiques à la traduction automatique, d'autres reflètent plus directement des stratégies de traduction individuelle ou encore la méconnaissance de certaines règles grammaticales des apprenants.

4.2. Différences entre traductions-machine et traductions humaines

Certaines formes apparaissent particulièrement dans les traductions automatiques. C'est le cas, par exemple, pour des formes comme : *avant*, *pour*, *laissez*, etc. Dans ces traductions, par exemple, la fréquence élevée de la forme *avant* trouve son origine dans la traduction systématique par les automates de la forme anglaise *before* par la forme française *avant*. Cette traduction est inadaptée, voire fautive dans certains cas.

Ainsi :

I stand here today humbled by the task **before** us

est traduit automatiquement par :

(a) Je me tiens ici aujourd'hui humilié par la tâche **avant** nous (*Systran*)
(b) Je suis debout ici aujourd'hui humilié par la tâche **avant** nous (*Reverso*)

Alors que dans une des traductions humaines, réalisée par le *Monde* par exemple, elle est traduite par :

Je me tiens aujourd'hui devant vous avec un sentiment d'humilité, devant la tâche qui nous attend

4.3. Différences entre traductions professionnelles et traductions d'apprenants

Dans certains cas, des apprenants qui ne maîtrisent pas totalement la langue d'arrivée ont du mal à opter pour une solution appropriée. Ainsi, pour rendre l'expression *turn back*, dans la phrase :

we refused to let this journey end, that we did not turn back nor did we falter,

nous trouvons des variations entre les traductions :

01Google et: 11B, 11E, 12J, 12K, 12L, 12P, 21A et 22C	*revenir/retourner en arrière*
31Croix, 32Libération	*tourner le dos*
33Le Monde	*détourner*
34RFI	*faire un demi-tour*
35WH	*faire tourner*

Comme nous le voyons, les traductions professionnelles font preuve d'une plus grande recherche et manifestent un plus grand souci d'expression littéraire en utilisant des expressions telles que *tourner le dos*, *faire un demi-tour.*

Une lecture comparative des listes des spécificités de chacun des groupes de traductions nous permet d'entrevoir des habitudes de traduction propres à chaque traducteur. Le contexte permet de cerner des écarts majeurs entre les traductions humaines réunies dans notre corpus. Ces écarts, qui concernent principalement des utilisations différentes des procédés grammaticaux et des procédés de mise en forme du texte, peuvent être regroupés en six grandes catégories :

1) les **déictiques** (ex. : *on*, *les*, *ceci*…) ;
2) les **adverbes** (ex. : *toute*, *même*, *simplement*…) ;
3) les **temps** (ex. : *va*, *vont*, *laissez*…) ;
4) les **noms** (ex. : *disponibilité*, *gouvernants*, *états*, *unis*…) ;
5) les **prépositions** (ex. : *arrière*, *avant*, *pour*…) ;
6) les **notes** (ex. : *ndlr*).

Dans ce qui suit, nous montrons quelques exemples d'utilisation différenciée de ces catégories chez les différents traducteurs.

Déictiques : Le pronom indéfini *on* est rarement utilisé dans les traductions professionnelles du discours de B. Obama qui a servi de texte-source. Il est nettement plus utilisé par les apprenants étrangers (en particulier : 12H et 22B) pour rendre le pronom we du texte original.

That **we** are in the midst of crisis is now well understood. […]

Nous savons maintenant fort bien que nous sommes en crise. (*Le Monde*)
On est désormais bien conscient qu'**on** traverse une crise (*22B*)

L'emploi du pronom we est relativement fréquent dans le texte original, dans la mesure où le discours est construit autour de la première personne du pluriel[14]. Le transfert du pronom défini de la première personne du pluriel en pronom indéfini (*on*) est opéré systématiquement par l'étudiant 22B. Ce dernier tente d'éviter les répétitions du pronom *nous* et semble céder à une habitude de langage courante chez les jeunes générations d'utiliser le pronom indéfini (Fonseca-Gréber et Waugh 2003, Thomas 2015).

Adverbes : l'adverbe *toute* est peu utilisé par les apprenants étrangers de première année. En revanche, il l'est fréquemment par *Systran* et les apprenants 11G et 22B. Les traductions automatiques de l'anglais au français utilisent fréquemment l'adverbe *tout* (et ses flexions), pour rendre all et throughout. Dans les traductions humaines, cet emploi est plus contrôlé et dépend de l'intensité que le traducteur veut rendre dans sa propre production. Voyons un exemple :

Tableau 3
Diversité d'intensité dans les traductions françaises via l'emploi de *toute* dans corpus *Obama1*

Original	I stand here today humbled by the task before us, grateful for the trust you have bestowed, mindful of the sacrifices borne by our ancestors. I thank President Bush for his service to our nation, as well as the generosity and cooperation he has shown ***throughout*** this transition.
02Systran	Je me tiens ici aujourd'hui humilié par la tâche avant nous, reconnaissant pour la confiance que vous avez accordée, conscient des sacrifices soutenus par nos ancêtres. Je remercie le président Bush de son service à notre nation, aussi bien que la générosité et la coopération qu'il a montrée dans ***toute*** cette transition.
21I	Je suis ici aujourd'hui pour vous avouer mon plein engagement envers la tâche qui nous incombe, ***toute*** ma reconnaissance de la confiance que vous m'avez faite et mon respect envers les sacrifices de nos ancêtres. Je remercie Président Bush pour le travail qu'il a fait pour notre pays, ainsi que pour sa générosité et sa coopération pendant cette transition.
34RFI	Je suis là devant vous humble face aux tâches qui nous attendent, reconnaissant de votre confiance et attentif aux sacrifices de nos ancêtres. Je remercie le président Bush, pour ses services rendus à la nation, ainsi que pour ***toute*** la générosité et la coopération qu'il a montrées lors de ***toute*** cette période de transition.
35WH	Je me présente devant vous aujourd'hui en ***toute*** humilité face à la tâche qui nous attend, reconnaissant de la confiance que vous m'avez accordée et conscient des sacrifices consentis par nos ancêtres. *je remercie le président *bush des services rendus à notre nation, ainsi que de la générosité et de la coopération dont il a fait preuve durant ***toute*** la transition.

Systran utilise *toute* pour traduire throughout dans le texte original, lorsque dans la même partie citée, nous constatons que cet adverbe est utilisé à divers endroits chez les traducteurs humains. L'étudiant *12I* met *toute* devant le nom *reconnaissance*

pour traduire l'adjectif *grateful*, alors que RFI souligne la *générosité* et le temps *période*. Cependant, la *Maison-Blanche* met l'accent sur l'*humilité* de *moi* en tant que président.

Temps: dans notre corpus, les trois formes du verbe *aller*: *va, vont et allons* ne concernent que le futur proche. La traduction du *Monde* les utilise seize fois alors que le *RFI* ne les utilise jamais. La *Maison-Blanche* s'en sert une seule fois; la *Croix* et *Libération* l'utilisent, respectivement, cinq et six fois. Parmi les traductions automatiques, seul *Google* a recours au futur proche, alors que *Systran* et *Reverso* l'évitent complètement. Dans les traductions des étudiants, on ne note pas de distinction nette entre les étudiants en deuxième année de scolarité et les étudiants français et étrangers.

Noms: l'étudiant de première année 11C traduit systématiquement America par *Etats-Unis d'Amérique* alors que les autres utilisent principalement la forme *Amérique*. Mais lorsqu'il s'agit d'un appel ou d'une invocation, la traduction de ce mot peut varier. Par exemple, dans une des phrases de la conclusion de B. Obama:

> **America**, in the face of our common dangers, in this winter of our hardship, let us remember these timeless words.

Le nom propre America a été rendu par *Etats-Unis* dans *Libération*, mais par l'adresse *Chers concitoyens* dans la traduction *Maison-Blanche*. Ceci laisse transparaître une stratégie d'écriture visant à impliquer plus directement le destinataire.

Prépositions: la préposition *pour* apparaît fréquemment dans les traductions automatiques *Systran* et *Reverso*. Elle est systématiquement utilisée pour traduire: *to, for, so that, etc.* Certaines des traductions produites par des apprenants (par exemple, 12L et 22H) ainsi que la traduction du *Monde* semblent également marquer une préférence pour cette préposition. Il en résulte que ces traductions ont tendance à sous-utiliser d'autres prépositions telles que *de* et *à*.

Notes: l'acronyme *ndlr* signifie *note de la rédaction*. Il apparaît exclusivement dans *Libération*, à deux endroits: une fois pour introduire des précisions sur la base militaire de Khe Sanh (Vietnam), l'autre pour donner des précisions sur le pays natal du père du président Obama (le Kenya). Ces deux notes reflètent la préoccupation du traducteur pour son lectorat français. Dans un ordre d'idées comparable, l'emploi par les étudiants de l'article contracté *au*, de la préposition *dans*, des articles définis *le* ou *la,* devant la forme *Khe Sanh*, laisse, avant tout, transparaître un manque d'information sur la nature exacte et l'histoire de ce lieu.

4.4. Analyse verticale des traductions

Au cours de notre démarche qui vise à étudier les différentes façons de traduire un texte, le calcul des spécificités nous permet de repérer les écarts les plus importants dans les variations de traductions. Les concordances et les méthodes de cartographie textuelle nous permettent de localiser facilement les contextes qui manifestent ces variations de manière remarquable. Pour une même séquence (paragraphe, phrase, segment répété), il est alors possible d'analyser de manière synthétique les variations produites par les différents traducteurs. Nous appelons *analyse verticale* ce type d'approche qui peut être centré sur chacun des différents problèmes rencontrés lors de la traduction du texte source.

La matérialisation des écarts mis en évidence par le calcul des spécificités, sous forme de soulignage des séquences textuelles correspondant à une même portion du texte source, permet de visualiser de manière particulièrement suggestive les convergences et les discordances qui existent entre les différentes traductions d'un même texte.

Tableau 4
La première phrase du discours de B. Obama (2009) et ses cinq traductions professionnelles

Original	My fellow citizens: I stand here today ***humbled by*** the task before us, ***grateful for*** the trust you have bestowed, ***mindful of*** the sacrifices borne by our ancestors.
31Croix	Mes chers concitoyens, Je me tiens devant vous, ***mesurant humblement*** la tâche qui nous incombe, ***reconnaissant pour*** la confiance que vous avez témoignée, ***conscient des*** sacrifices consentis par nos ancêtres.
32Liber	Mes chers concitoyens, Je me tiens devant vous, ***mesurant humblement*** la tâche qui nous incombe, ***reconnaissant pour*** la confiance que vous avez témoignée, ***conscient des*** sacrifices consentis par nos ancêtres.
33Monde	Chers compatriotes, je me tiens aujourd'hui devant vous ***avec un sentiment d'humilité***, devant la tâche qui nous attend, de ***reconnaissance pour*** la confiance que vous m'avez manifestée, ***gardant à l'esprit*** les sacrifices consentis par nos ancêtres.
34RFI	Je suis là devant vous ***humble face aux*** tâches qui nous attendent, ***reconnaissant de*** votre confiance et ***attentif aux*** sacrifices de nos ancêtres.
35WH	Mes chers concitoyens: Je me présente devant vous aujourd'hui ***en toute humilité face à*** la tâche qui nous attend, ***reconnaissant de*** la confiance que vous m'avez accordée et ***conscient des*** sacrifices consentis par nos ancêtres.

L'examen vertical des traductions professionnelles permet d'explorer les possibilités de traduction offertes par les deux langues et de nous concentrer sur les choix effectués par les différents traducteurs. Dans l'ordre, nous posons plusieurs questions aux étudiants : comment traduire l'appel dans un discours politique ? Quelle est la façon usuelle de le faire en français ? Quelle ponctuation utilise-t-on ? Comme rendre en français le verbe et l'indication de location de *I stand here* ? Comment traduire le sens figuré de la localisation contenue dans *the task before us* ? Est-il toujours obligatoire de traduire un adverbe temporel du texte de départ (*today*) ? Comment rendre dans la langue cible la musicalité du texte original qui résulte du la mise en parallèle des sentiments : *humbled by…grateful for…mindful of… ?*

Après avoir analysé l'ensemble des traductions réalisées par les traducteurs professionnels, il est intéressant d'examiner les traductions fournies par les apprenants. L'exemple de la forme française *face*, dont la répartition irrégulière parmi les traductions des apprenants est mise en évidence par le calcul des spécificités, nous fait découvrir que pour rendre le sens figuré de la localisation exprimé dans le segment *the task before us /la tâche qui nous attend*, on peut utiliser plusieurs procédés : *devant, face à*, utiliser le pluriel *face aux tâches*, recourir au participe présent *mesurant*, lequel permet de réaliser un parallélisme sonore avec *reconnaissant… conscient…*

Tableau 5

Plusieurs traductions de la séquence *the task before us* dans la première phrase du discours de B. Obama (2009)

Original	the task ***before us***	**21A**	devant la tâche qui nous attend
31Croix	mesurant humblement la tâche qui nous incombe	**22B**	face à la tâche qui nous attend
32Liber	face à la tâche qui nous attend	**22C**	face à la tâche qui nous attend
33Monde	devant la tâche qui nous attend	**22D**	devant la tâche que nous avons à accomplir
34RFI	face aux tâches qui nous attendent	**22E**	mesurant la tâche qui nous attend
35WH	face à la tâche qui nous attend	**22F**	devant la tâche qui nous attend
		22G	face à la tâche qui nous attend
		22H	devant la tâche qui nous attend
		22I	devant la tâche qui nous attend

Cette approche permet de faire prendre conscience aux apprenants qu'il existe plusieurs façons de rendre le sens d'un segment lorsqu'on le traduit d'une langue à l'autre. Par delà l'indispensable conservation du sens, une traduction qui prend en compte des éléments de sonorités manifeste un travail plus élaboré. À travers de telles comparaisons, les apprenants intègrent naturellement l'idée de l'évaluation et se familiarisent avec les techniques de la traduction. De cette manière, ils développent également une méthode d'apprentissage qui peut leur servir dans les futures études.

5. Conclusion

Dans le cadre de l'enseignement de la traduction, la confrontation de plusieurs traductions d'un même texte original permet à l'enseignant de présenter différentes possibilités de traduction et d'inciter les élèves à distinguer différents niveaux de traduction. À travers ce type d'observation, les apprenants peuvent développer leur sens de l'évaluation du travail de traduction et parvenir à une autoévaluation du travail qu'ils ont fourni. Les analyses que nous avons effectuées à l'aide des méthodes textométriques sur trois séries de traduction d'un même texte (des traductions fournies par les apprenants, des traductions professionnelles et des traductions réalisées par des automates) ont montré une similitude dans l'emploi des verbes et des adjectifs et une variation sur l'emploi des mots-outils (prépositions, adverbes, déictiques, etc.). Nous avons évalué les difficultés spécifiques éprouvées par les apprenants dont la langue de travail n'était pas leur langue maternelle.

Grâce aux outils textométriques, nous avons exploré, avec une grande efficacité, la variété des traductions rassemblées dans le corpus. L'utilisation de l'alignement du corpus en paragraphes permet de construire un réseau de comparaisons sur lesquelles les calculs textométriques peuvent ensuite s'appuyer pour produire des résultats particulièrement explicites. La représentation des différents textes réunis dans le corpus sous forme de cartes des sections alignées permet de visualiser des phénomènes de répartition qui attirent alors l'attention de l'analyste. Les méthodes d'analyse statistique (AFC, localisation des segments répétés, analyse de spécificités) mettent en évidence des traits d'écriture propres à chacun des groupes de traductions.

Les méthodes de la textométrie peuvent aider les enseignants à comprendre les procédés employés lors des traductions effectuées par les apprenants. Ils permettent de repérer les problèmes que ceux-ci ont rencontrés, de percevoir leurs hésitations.

Enfin, l'examen vertical des traductions permet de localiser les portions du texte original qui ont reçu un traitement uniforme de la part des différents traducteurs. Grâce à cette même approche, nous pouvons également localiser les fragments du texte original ayant donné lieu à des traductions particulièrement variées et explorer l'éventail des possibilités attestées dans le corpus.

Lors de l'analyse comparative des traductions proposées par les étudiants, l'enseignant trouve l'occasion de mettre ces derniers en garde contre l'utilisation systématique aux solutions proposées par les automates de traduction. En effet, ces traductions sont parfois fautives. Si l'on peut accepter qu'une traduction s'inspire très fortement de celles des solutions proposées par les automates de traduction qui se révèlent être acceptables sur le plan traductologique, la reproduction systématique, dans un même travail, des erreurs commises par ces automates, témoigne à coup sûr d'une compétence insuffisante des étudiants.

À l'issue de ce travail, nous sommes convaincus que l'enseignement de la traduction trouvera une aide précieuse dans l'utilisation des méthodes textométriques appliquées aux corpus alignés multilingues.

REMERCIEMENTS

Les auteurs remercient sincèrement Kim Gerdes, Serge Fleury, Jean-Michel Daube, Mathieu Valette, Colette Laplace, Marianne Lederer et Sylvie Royer pour leur aide précieuse et leurs conseils dans la réalisation de cet article.

NOTES

1. Les références de tous les logiciels mentionnés dans l'article se trouvent dans la deuxième partie de la bibliographie.
2. À l'opposé des procédures telles que BLEU (Papineni *et al.* 2002), ROUGE (Lin 2004) et NIST (http://www.nist.gov/itl/iad/mig/mt.cfm), qui visent l'évaluation intrinsèque de la performance des programmes de traduction automatique, notre expérience se concentre sur l'étude des écarts entre divers types de traductions et la mise en place de systèmes d'autoévaluation destinés aux apprenants.
3. Le texte original a été prélevé sur le site du *White House* <https://goo.gl/MOVdwn>.
4. *La Croix* <http://goo.gl/qo0HYY>H ; *Libération* (AFP) <http://goo.gl/ZPbYMj>H ; *Le Monde* (traduit par Arianne Cobin-Favier) <http://goo.gl/h7cI3N>H ; *RFI* (traduit par Chérif Ezzel) <http://goo.gl/h7HE9yH> ; White House <http://goo.gl/vBz4L0> (Tous les fichiers ont été téléchargés le 10 septembre 2012, et vérifiés.)
5. Nous avons obtenu le résultat des traductions automatiques le 25 novembre 2012.
6. Ce logiciel a été conçu par Kim Gerdes, Paris 3. (Pour plus d'informations sur celui-ci, voir Gerdes (2008) et Gerdes et Miao (2008).
7. Un tableau lexical soumis à l'analyse factorielle des correspondances est un tableau de contingence : les différents *types* (*i*) de vocabulaire occupent les lignes du tableau, les différentes *parties* (*j*) du corpus les colonnes, et le nombre d'occurrences (*Kij*) correspond à celui de la forme *i* dans la partie *j*. À l'aide d'analyses statistiques et de graphiques, les facteurs variables du tableau seront mesurés et représentés par des proximités géométriques entre points-lignes et points-colonnes. Cette méthode rend particulièrement compte des principales oppositions qui sous-tendent le corpus : on s'appuie dans un premier temps sur la distance du chi-deux pour calculer la distance entre chacune des paires de textes constituant le corpus ; puis on décompose les distances sur une succession hiérarchisée d'axes factoriels à l'aide des pourcentages d'inertie. Les pourcentages d'inertie issus de notre analyse font apparaître la succession suivante : $\tau 1 = 14\,\%$, $\tau 2 = 8\,\%$, $\tau 3 = 7\,\%$, $\tau 4 = 5\,\%$…

8. Notons cependant que la traduction du *RFI* et celle de *22H* ont une position relativement particulière, elles se situent légèrement au-dessous de l'axe X.
9. On se souvient que les *Exercices de style* de Raymond Queneau (1947/2006) présentent quatre-vingt-dix-neuf manières différentes de raconter une même histoire.
10. On consultera sur ces questions les travaux sur la notion de *résonance textuelle* (Salem 2004).
11. Le type généralisé *Tgen* permet de constituer des unités textométriques rassemblant les occurrences de formes graphiques différentes liées par une même propriété (Lamalle *et al.* 2003). Cette propriété est implantée sous le nom de : *Groupes de formes* dans le logiciel ***Lexico3***.
12. Pour obtenir des paragraphes dans l'état que nous les présentons dans la figure 4, nous avons transformé les résultats bruts fournis par le logiciel *Lexico3*. Nous avons utilisé le logiciel pour repérer la première occurrence de chacune des séquences de longueur supérieure ou égale à 5, répétées au moins 5 fois. Ce marquage nous a permis de souligner, dans un second temps, les répétitions et les altérations contenues dans les paragraphes étudiés.
13. Nous nous concentrons ici sur les formes spécifiques majeures, les formes dont le coefficient de spécificité est supérieur à 4.
14. La forme *we* apparaît 62 fois dans le corpus original, et toutes les formes du pronom de la première personne au pluriel (*we*, *our*, *ours*) occupent 61.9 % de l'ensemble des pronoms utilisés.

RÉFÉRENCES

Bénzecri, Jean-Paul *et al.* (1973) : *L'analyse des données*. Tome 1 : *La taxinomie*, tome 2 : *L'analyse des correspondances*. Paris : Dunod.

Benzecri, Jean-Paul (1977) : Analyse discriminante et analyse factorielle. *Les Cahiers de l'Analyse des Données*. II(4):369-406.

Benzecri, Jean-Paul *et al.* (1981). *Pratique de l'analyse des données : linguistique et lexicologie*. Paris : Dunod.

Bowker, Lynne et Bennison, Peter (2003) : Student translation archive : design, development and application. *In* : Federico Zanettin, Silvia Bernardini et Dominic Stewart, dir. *Corpora in Translator Education*. Manchester, UK et Northampton, MA : St Jerome, 103-117.

Castagnoli, Sara, Ciobanu, Dragoş, Kunz Kerstin, *et al.* (2009) : Designing a Learner Translator Corpus for Training Purposes. *In* : Natalie Kübler, dir. *Corpora, Language, Teaching, and Resources : From Theory to Practice*. Bern : Peter Lang. 221-248.

Floren, Celia (2006) : ENTRAD, an English Spanish parallel corpus created for the teaching of translation. Paper presented at *the 7th Teaching and Language Corpora conference* (TaLC), Université Paris 7 Denis Diderot, Paris, 1-4 July 2006.

Fonseca-Greber, Bonnie et Waugh, Linda R (2003) : On the Radical Difference between the Subject Personal Pronouns in Written and Spoken European French. *In* : Pepi Leistyna, Charles F. Meyer, dir. *Corpus Analysis : Language Structure and Language Use*, 225-240.

Frérot, Céline (2010) : Outils d'aide à la traduction : pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs. *Les Cahiers du GEPE*, 2010 (2), *Outils de traduction – outils du traducteur ?* Consulté le 7 juin 2012, <http://www.cahiersdugepe.fr/index1164.php>.

Gerdes, Kim (2008) : L'alignement pour les pauvres. *In* : Serge Heiden et Bénédicte Pincemin, éd. *Actes des 9e Journées internationales d'Analyse statistique des Données Textuelles* (JADT), Université de Lyon, Lyon, 12-14 mars 2008. Vol. 1. Lyon : Presses Université de Lyon, 527-538.

Gerdes, Kim et Miao Jun (2008) : Donner accès à l'œuvre de Fu Lei. *In* : Xu Yun, éd. 傅雷的精神世界及其时代意义 [Le monde spirituel de Fu Lei et sa signification dans le temps], *Colloque international Fu Lei et traduction, traductions chinoises*, Université de Nanjing, Nanjing, 15-18 mai 2008. Shanghai : 中西书局, 351-366.

Granger, Sylviane, dir. (1998) : *Learner English on Computer*. London et New York : Addison Wesley Longman.

Israël, Fortunato (1999) : Principes pour une pédagogie raisonnée de la traduction : le modèle interprétatif. *In* : Ivana Čeňková, Jana Králová-Kullová, dir. *Folia Translatologica*, Vol. 6,

International Series of Translation Studies, Issues of Translation Pedagogy, Helsinki – Paris – Prague : Charles University, Faculty of Arts, 21-32.

Kiraly, Don (2005) : *A Social Constructivist Approach to Translator Education*. Manchester : St. Jerome.

Lafon, Pierre (1980) : Sur la variabilité de la fréquence des formes dans un corpus. *Mots*. 1(1):127-165.

Lafon, Pierre (1984) : *Dépouillements et statistiques en lexicométrie*. Genève/Paris : Slatkine/Champion.

Lamalle Cédric, Salem André. (2002) : Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels. *In* : *Actes des 6e Journées internationales d'Analyse statistiques des Données Textuelles* (JADT), Saint-Malo, 13-15 mars 2002. Inria : 403-412.

Lamalle, Cédric, Martinez, William, Salem, André (2003) : *Lexico3 outils de statistique textuelle et Manuel d'utilisation*. SYLED – CLA2T, Paris : Université de Paris 3.

Lee-Jahnke, Hannelore (2001) : Aspects pédagogiques de l'évaluation des traductions. *Meta*. 46(2):258-271. Consulté le 22 février 2015, <http://www.erudit.org/revue/meta/2001/v/n2/003447ar.html>.

Lin, Chin-Yew (2004) : ROUGE : A Package for Automatic Evaluation of Summaries. *In* : *Proceedings of the Workshop on Text Summarization Branches Out* (WAS 2004), Barcelone juillet 25-26, 2004.

Masschelein, Danny, Verschueren, Walter (2005) : Vers un apprentissage semi-autonome du processus de la traduction. *Meta*. 50(2):560-572. Consulté le 22 février 2015, <http://id.erudit.org/iderudit/011000ar>.

Papineni, Kishore, Roukos, Salim, Ward Todd et Zhu Wei-Jing (2002) : BLEU : A Method for Automatic Evaluation of Machine Translation. *In* : *the 40th Annual meeting of the Association for Computational Linguistics* (ACL), Philadelphie, juillet 2002. Philadelphia : 311–318.

Popescu-Belis, Andrei (2002) : Constitution de banques de textes multilingues : un mécanisme fondé sur le standard XML. *Cahiers du Rifal / Terminologies Nouvelles*. 23:56-61.

Queneau, Raymond (1947/2006) : *Les exercices de style*. Paris : Gallimard.

Salem, André (1986) : Segments répétés et analyse statistique des données textuelles. *Histoire et Mesures*. 1(2):5-28.

Salem, André (2004) : Introduction à la résonance textuelle. *In* : Gérald Purnelle *et al*. éd. *Actes des 7e Journées internationales d'Analyse statistique des Données Textuelles* (JADT), Université de Louvain, Louvain-la-Neuve, 10-12 mars 2004. Louvain-la-Neuve : Presses universitaires de Louvain, 986-992.

Sosnia, E. P. (2006) : Development and application of Russian Translation Learner Corpus. *In* : *Corpus Linguistics*, St. Petersburg, 10-14 octobre 2006. St. Petersburg : 365-373.

Tercedor-Sanchez, Maria Isabel, López-Rodriguez, Clara Inés et Robinson, Bryan (2005) : Textual and Visual Aids for E-learning Translation Courses. *Meta*. 50(4). Consulté le 14 avril 2014, <http://id.erudit.org/iderudit/019904ar>.

Thomas, Alain (2015) : Nous/on : De la réalité linguistique à la salle de classe. *Arborescences : revue d'études françaises* (5):136-138.

Valetopoulos, Freiderikos (2012) : Quand les apprenants doivent observer leurs stratégies métacognitives : une analyse de corpus. *Synergie Pologne*. (9):37-47.

Waddington, Christopher. (2001) : Different Methods of Evaluating Student Translations : The Question of Validity. *Meta*. 46(2):311-325. Consulté le 5 février 2015, <https://www.erudit.org/revue/meta/2001/v46/n2/004583ar.pdf>.

Williams, Malcolm. (2001) : The Application of Argumentation Theory of Translation Quality Assessment. *Meta*. 46(2):326-344. Consulté le 5 février 2015, <http://www.erudit.org/revue/meta/2001/v46/n2/004605ar.pdf>.

OUTILS MENTIONNÉS DANS L'ARTICLE (CONSULTÉS OU UTILISÉS DURANT 2012-2013)

Outils de l'enseignement
Markin : http://www.cict.co.uk/markin/index.php

Outils de traitement de corpus
AntConc : http://www.laurenceanthony.net/software/antconc/
Alignator : http://elizia.net/alignator/alignator.cgi
Lexico3 : http://www.tal.univ-paris3.fr/lexico/
WordSmith : http://www.lexically.net/wordsmith/

Outils d'aide à la traduction
Déjà Vu : http://www.atril.com/
Trados : http://www.sdl.com/fr/cxc/language/translation-productivity/trados-studio/
Wordfast : https://www.wordfast.net/

Outils de la traduction automatique
Google Translate : https://translate.google.fr/
Reverso : http://www.reverso.net
Systran : http://www.systran.fr/