

Comparaison d'un texte original et de ses rétrotraductions : que disent les mesures textométriques ?

Olivier Kraif et Pascale Roux

Volume 67, numéro 1, avril-mai 2022

Pour de nouvelles méthodes en traductologie quantitative
Exploring New Methods in Quantitative Translation Studies

URI : <https://id.erudit.org/iderudit/1092196ar>
DOI : <https://doi.org/10.7202/1092196ar>

[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

Kraif, O. & Roux, P. (2022). Comparaison d'un texte original et de ses rétrotraductions : que disent les mesures textométriques ? *Meta*, 67(1), 190–216. <https://doi.org/10.7202/1092196ar>

Résumé de l'article

Cet article présente les premiers résultats de l'analyse d'un corpus littéraire composé de poèmes et d'un essai, ainsi que de leurs 16 rétroversions en français, après traduction dans 8 langues, typologiquement et diachroniquement diversifiées (l'italien, l'allemand, l'arabe, le farsi, le japonais, le coréen, le latin et le grec ancien). Cet ensemble fait l'objet d'une approche interdisciplinaire, interrogeant la rétrotraduction d'un point de vue textométrique et stylistique. À partir d'une évaluation de la distance traductionnelle entre les textes alignés (originaux et rétroversions), fondée sur la longueur, l'indice Dice et la stabilité des lemmes, trois hypothèses principales sont éprouvées : les textes sont-ils davantage transformés lorsqu'ils sont passés par une langue distante du français ? Les poèmes subissent-ils des distorsions plus grandes que l'essai en prose ? Le corpus rétrotraduit peut-il être un outil herméneutique pour l'analyse du texte original ? Pour étudier ces différents aspects dans une perspective de linguistique de corpus outillée et de stylistique française, nous proposons une méthodologie d'alignement multitextuel et de mesures textométriques, s'appuyant sur l'étiquetage et la lemmatisation des unités. L'étude stylistique prend appui sur ces mesures pour aborder ce multitexte riche et complexe. Menée sur un ensemble de taille réduite, l'analyse vise à la mise en place de procédures reproductibles sur d'autres corpus de versions traduites d'un texte dans une même langue.

Comparaison d'un texte original et de ses rétrotraductions : que disent les mesures textométriques ?

OLIVIER KRAIF

Université Grenoble Alpes, Grenoble, France
olivier.kraif@univ-grenoble-alpes.fr

PASCALE ROUX

Université Grenoble Alpes, Grenoble, France
pascale.roux@univ-grenoble-alpes.fr

RÉSUMÉ

Cet article présente les premiers résultats de l'analyse d'un corpus littéraire composé de poèmes et d'un essai, ainsi que de leurs 16 rétroversions en français, après traduction dans 8 langues, typologiquement et diachroniquement diversifiées (l'italien, l'allemand, l'arabe, le farsi, le japonais, le coréen, le latin et le grec ancien). Cet ensemble fait l'objet d'une approche interdisciplinaire, interrogeant la rétrotraduction d'un point de vue textométrique et stylistique. À partir d'une évaluation de la distance traductionnelle entre les textes alignés (originaux et rétroversions), fondée sur la longueur, l'indice Dice et la stabilité des lemmes, trois hypothèses principales sont éprouvées : les textes sont-ils davantage transformés lorsqu'ils sont passés par une langue distante du français ? Les poèmes subissent-ils des distorsions plus grandes que l'essai en prose ? Le corpus rétrotraduit peut-il être un outil herméneutique pour l'analyse du texte original ? Pour étudier ces différents aspects dans une perspective de linguistique de corpus outillée et de stylistique française, nous proposons une méthodologie d'alignement multitextuel et de mesures textométriques, s'appuyant sur l'étiquetage et la lemmatisation des unités. L'étude stylistique prend appui sur ces mesures pour aborder ce multitexte riche et complexe. Menée sur un ensemble de taille réduite, l'analyse vise à la mise en place de procédures reproductibles sur d'autres corpus de versions traduites d'un texte dans une même langue.

ABSTRACT

This article presents the first results of the analysis of a literary corpus composed of poems and an essay, as well as their 16 retroversions in French, after translation into 8 languages, typologically and diachronically diversified (Italian, German, Arabic, Farsi, Japanese, Korean, Latin and ancient Greek). This corpus has been studied from an interdisciplinary point of view, mixing textometric data and stylistic observations. Starting from an evaluation of the translational distance between the aligned texts (originals and retroversions), based on length, Dice index and lemma stability, three main hypotheses are tested: are the texts more deeply transformed when they have been retro-translated through a language distant from French? Are the poems more distorted than the prose essay? Can the retro-translated corpus yield hermeneutic clues for the analysis of the original text? In order to study these different aspects from a perspective of corpus linguistics and French stylistics, we propose a methodology of multi-textual alignment and textometric measurements, based on tagging and lemmatisation. The stylistic study is based on these measures to address this rich and complex multi-text. Conducted on a small set, the analysis aims at setting up procedures that can be reproduced on other corpora of translated versions of a text in the same language.

RESUMEN

Este artículo presenta los primeros resultados del análisis de un corpus literario compuesto por poemas y un ensayo, así como sus 16 retroversiones en francés, tras su traducción a 8 idiomas, tipológica y diacrónicamente diversificadas (italiano, alemán, árabe, farsi, japonés, coreano, latín y griego antiguo). Este corpus se ha estudiado desde un punto de vista interdisciplinar, mezclando datos textométricos y observaciones estilísticas. A partir de una evaluación de la distancia traslativa entre los textos alineados (originales y retroversiones), basada en la longitud, el índice de Dice y la estabilidad de los lemas, se ponen a prueba tres hipótesis principales: ¿se transforman más profundamente los textos cuando han pasado por una lengua distante del francés? ¿Están los poemas más distorsionados que el ensayo en prosa? ¿Puede el corpus retraducido ser una herramienta hermenéutica para el análisis del texto original? Para estudiar estos diferentes aspectos desde una perspectiva de la lingüística de corpus y de la estilística francesa, proponemos una metodología de alineamiento multitextual y de mediciones textométricas, basada en el etiquetado y la lematización. El estudio estilístico se basa en estas medidas para abordar este rico y complejo multitexto. Realizado sobre un conjunto reducido, el análisis pretende establecer procedimientos que puedan reproducirse en otros corpus de versiones traducidas de un texto en el mismo idioma.

MOTS CLÉS/KEYWORDS/PALABRAS CLAVE

rétrotraduction, alignement, stabilité traductionnelle, textométrie, traduction littéraire, stylistique

retrotranslation, aligning, translational stability, textometrics, literary translation, stylistics
retrotraducción, alineamiento, estabilidad de la traducción, textometría, traducción literaria, estilística

Le langage est chimie pour le sens et physique pour les formes. Il est chimie, car il se crée, à partir d'un nombre restreint d'éléments linguistiques, un nombre infini de combinaisons à signification nouvelles; cependant les éléments qui entrent en combinaison pour donner une signification nouvelle ne perdent pas leur identité formelle comme c'est le cas des éléments d'un composé chimique, et la forme du langage est donc pour l'essentiel physique.

(Seleskovitch 1975: 49-50)

1. Introduction

La pratique de la rétrotraduction, qui consiste à traduire un texte puis à le retraduire dans sa langue de départ, est cantonnée pour l'essentiel dans des domaines pédagogique et technique ou scientifique. Dans le champ littéraire, elle est parfois pratiquée à des fins ludiques (par exemple par Montale 1999¹ ou Twain 1903²), mais elle n'est pas mobilisée dans des protocoles de recherche sur des corpus d'étude.

Dans le cadre pédagogique, elle constitue une manière d'entraîner au thème et à la version. Le repérage d'une distance significative entre une séquence du texte de départ et sa rétroversion permet de remonter le cours de la traduction, d'identifier l'étape à laquelle la distorsion s'est produite, d'en analyser les causes et d'y remédier. Dans le cadre professionnel, en particulier de la médecine et des sciences pharmaceutiques, la rétrotraduction sert à valider ou améliorer une traduction. Elle révèle efficacement des erreurs, en raison de la réduplication du processus traductionnel: elle permet «de dépister par amplification les erreurs commises au cours de la traduction, de mettre en évidence les ambiguïtés de formulation de la version source et

certaines interprétations inattendues contenues dans la version cible qui contaminent la version rétrotraduite» (Massoubre, Lang *et al.* 2002). Elle permet aussi, par un effet de retour, de réviser un texte, afin d'en éliminer les formulations équivoques susceptibles d'entraîner des erreurs d'interprétation (Camoin 2019: 42).

La rétrotraduction se présente donc, dans ses domaines d'application privilégiés, comme un outil pour la validation et l'amélioration d'une traduction. Ce protocole de vérification impose une certaine manière de traduire: il faut que les traducteurs successifs se tiennent au plus près de leurs textes de départ respectifs, d'un point de vue structurel et non uniquement sémantique, sans quoi l'épreuve n'est plus probante.

Il paraît ainsi logique que le texte et la traduction littéraires ne soient pas concernés par de telles procédures de validation. S'il est difficile de caractériser globalement la «littérarité», du moins peut-on dire que les textes littéraires ont en général un fort degré de polysémie, que les «formules équivoques» que l'écrit technique cherche à éviter y sont nombreuses et que, bien souvent, le texte est construit pour résister à une interprétation unique ou univoque. Quant à la traduction littéraire, qu'il est tout aussi difficile de caractériser de manière globale tant ses modèles et ses pratiques connaissent de variations en diachronie et en synchronie, elle s'accommode mal d'une injonction de se tenir «au plus près» du texte de départ, et l'on sait combien la notion de «fidélité» est une valeur métatraductive bien plus qu'un critère objectif, ne serait-ce que parce que les réalités que ce terme peut recouvrir sont très diverses (fidélité à l'intention de l'auteur, à la lettre du texte ou à son esprit, à l'effet produit). En outre, la traduction littéraire implique toujours une forme de créativité du traducteur, à divers degrés, et ce, même lorsque celui-ci revendique une hypothétique «transparence» de son style.

On peut néanmoins postuler que la mise en place d'une procédure de rétrotraduction littéraire peut produire des résultats intéressants, selon les principes mêmes qui font qu'on y recourt dans certains domaines, mais pour d'autres fins: puisqu'elle permet de trouver des zones du texte de départ qui posent des problèmes d'interprétation, voire de compréhension, elle doit pouvoir être utilisée à des fins d'analyse stylistique (comprendre le fonctionnement du texte) plutôt que de remédiation (l'améliorer); puisqu'elle permet de voir des distorsions générées au cours de la traduction, elle doit permettre de comprendre, plutôt que d'évaluer ou de valider, ce qui se produit lors du passage d'une langue à l'autre.

C'est pour mettre à l'épreuve la productivité scientifique de ce protocole qu'a été généré un corpus littéraire rétrotraduit, à partir de huit langues différentes, dans le cadre d'un projet réunissant des chercheurs et chercheuses d'horizons disciplinaires variés (stylistique française, linguistique et humanités numériques, langues étrangères)³. Nous présentons ici les premiers résultats de l'analyse de ce corpus, effectuée grâce à la textométrie et dans une perspective stylistique, sur la base uniquement de la partie en français (textes de départ et rétroversions).

L'une des ambitions de l'étude étant la mise en place de procédures reproductibles pour d'autres corpus, nous présenterons non seulement les résultats obtenus mais également les différentes étapes de l'expérimentation. Après avoir donné quelques indications sur la constitution du corpus, les objectifs du projet, les hypothèses formulées en amont et les premières observations empiriques, nous exposerons la manière dont les données ont été préparées en vue du traitement automatique du corpus. Les deux dernières parties seront consacrées à l'exposé synthétique des

résultats textométriques, des interprétations qu'on peut en proposer ainsi que des pistes de recherche qu'elles ouvrent.

2. Corpus, objectifs, hypothèses et premières observations empiriques

Nous avons sollicité un poète et essayiste contemporain, Gérard Macé, qui nous a transmis deux inédits: un court essai en prose, «Proche Afrique», et une série de quatre poèmes⁴. Chacun de ces textes a ensuite été traduit vers huit langues, typologiquement et diachroniquement proches ou éloignées du français: l'italien, l'allemand, l'arabe, le farsi, le japonais, le coréen, le latin et le grec ancien (que nous noterons respectivement, par commodité: IT, DE, AR, FA, JA, KO, LA ET GRC). Pour chacune d'entre elles, deux traducteurs ou traductrices⁵ ont été sollicités, afin que puissent apparaître les différences résultant de la créativité propre de chacun d'entre eux, à l'épreuve de la langue traduisante. Les 16 versions obtenues ont ensuite été rétrotraduites en français, par d'autres, qui n'avaient pas connaissance des textes de départ. Pour ces deux étapes, les seules instructions données étaient peu contraignantes mais, dans leur concision, exigeantes: «produire une traduction visant une autonomie et une valeur textuelle qui corresponde dans tous ses aspects à celle du texte de départ, en ne recourant à aucune note explicative». Le corpus généré est ainsi constitué, pour chaque texte (la série de poèmes et la prose) de 16 versions traduites et 16 rétroversions.⁶

Le corpus est de taille modeste: les deux ensembles textuels de départ sont brefs et le nombre de versions (rétro)traduites et de langues représentées est limité. Il n'est donc nullement représentatif et ne permet pas la généralisation, d'autant que le choix des langues comporte en outre une part d'arbitraire et de parti-pris, par exemple l'absence de l'anglais, la présence de langues de traduction minoritaires et de langues dites «mortes».

Mais nos objectifs étaient avant tout méthodologiques. Il s'agissait d'une part de mettre à l'épreuve la productivité de la rétrotraduction dans le champ de la recherche littéraire, d'autre part d'expérimenter des procédures pour l'étude de corpus multitextes traduits, unilingues ou multilingues, alignés ou pouvant l'être. Le postulat était que ces procédures pourraient ensuite être reproduites pour des corpus non pas suscités dans le cadre d'une recherche mais générés «spontanément» dans le champ littéraire, tels un ensemble de traductions dans diverses langues d'un même texte, ou encore une série de versions françaises d'un texte retraduit à de multiples reprises dans l'histoire. Sur le plan disciplinaire, il s'agissait de créer les conditions pour que puissent se croiser trois approches: l'étude stylistique du texte littéraire français (originellement écrit en français ou traduit), la traductologie, les humanités numériques (édition et TAL). Pour que ces objectifs puissent être atteints, il fallait que le corpus soit réduit, du point de vue volumétrique: nous voulions non seulement croiser différentes approches disciplinaires, mais aussi faire dialoguer les chercheurs avec la quarantaine de traducteurs impliqués, mettre en regard les observations empiriques et les données textométriques, enfin comparer les interprétations que les uns et les autres étaient susceptibles de proposer de ces données.

À l'intérieur de ce cadre général, nous projetions au départ de tester deux hypothèses correspondant, dans une certaine mesure, à des énoncés stéréotypés sur le processus de traduction:

- Le passage d'un texte par une langue éloignée (comme le japonais ou le grec ancien) est susceptible de générer davantage de distorsions, observables dans les rétroversions, que le passage du même texte par une langue proche (comme l'italien ou le latin).
- La poésie étant réputée plus sujette à distorsion, dans la traduction, que d'autres types de textes, on s'attendait à ce que les poèmes, en vers libres et blancs, marqués par une grande densité métaphorique, par une certaine opacité sémantique ainsi que par des phénomènes de concision (sémantique, syntaxique, figurale), soient, au retour, plus éloignés des textes de départ que l'essai en prose.

Il est immédiatement apparu que ces hypothèses n'étaient pas massivement validées par le corpus produit. Mais pour pouvoir déboucher sur la production de résultats, il était nécessaire d'évaluer statistiquement, de manière quantitative et qualitative, la distance entre les versions, à plusieurs échelles (depuis celle du texte dans son entier jusqu'à celle du mot, en passant par des séquences de longueur variable) et selon différentes configurations (comparaison du texte de départ et d'une ou plusieurs rétroversions ou encore de plusieurs rétroversions, associées par paires ou par groupes plus étendus).

Le premier examen empirique a par ailleurs fait émerger de nouvelles hypothèses (voir Roux 2018)⁷, dont nous ne retiendrons ici qu'une seule: en se plaçant à une échelle inférieure au texte dans son ensemble, il semble possible, grâce au corpus rétrotraduit, de caractériser des séquences textuelles, en fonction du critère de la stabilité ou de la distorsion traductionnelle, c'est-à-dire de la convergence ou de la divergence des rétroversions par rapport au texte de départ. L'intérêt d'identifier des séquences selon ce critère est en lien direct avec les caractéristiques du protocole de rétrotraduction, décrites en introduction: on peut postuler que les virtualités interprétatives réalisées dans les rétroversions sont en germe dans le texte de départ et qu'elles se trouvent amplifiées par la reduplication de la traduction. Ainsi, la plus ou moins grande stabilité des séquences peut être utilisée comme un indicateur que l'on pourrait dire «herméneutique». Identifier les séquences instables peut permettre de mettre au jour leur caractère équivoque (par exemple lorsque les rétroversions manifestent des choix d'interprétation divergents), des phénomènes de polysémie (par exemple lorsqu'on constate le recours à une série de synonymes partiels différents dans les rétroversions), ou même des phénomènes proprement stylistiques qui viennent parasiter et/ou enrichir la construction du sens (par exemple dans le cas de figures de style perturbant la compréhension du sens et la rendant plus complexe). Le corpus rétrotraduit peut alors être abordé comme un catalogue de lectures et de réceptions des textes de départ, et la (rétro)traduction comme une forme d'interprétation, par définition discutable mais pouvant être consensuelle (dans les cas de convergence des rétroversions) ou non (dans les cas de divergence).

L'analyse textométrique du corpus, centrée sur l'identification et la description de la distance entre le texte de départ et les rétroversions, a été orientée en fonction de ces premières observations, afin de répondre aux questions suivantes:

- La distance entre les textes de départ et les rétroversions est-elle corrélée à la distance entre le français et la langue par laquelle le texte est passé, ainsi qu'à la répartition typologique poésie/essai?
- Peut-on relever dans les textes de départ des séquences stables, moyennement stables ou instables (pour ce qui est des lemmes et des phrases ou encore d'ensembles intermédiaires entre le lemme et la phrase, entre la phrase et le texte) et, si c'est le cas, ces résultats peuvent-ils être utilisés comme un indicateur herméneutique?

3. Préparation des données

Les corpus ont tout d'abord été tokenisés avec un script Perl (tokenisation standard sans dictionnaire, sur le plan de l'espace typographique, de l'apostrophe et des signes de ponctuation, le tiret n'étant considéré comme séparateur que dans le cas des pronoms clitiques et du *-t-* épenthétique). Le nombre de tokens (mots et ponctuations) de chaque partie du corpus est représenté dans le tableau 1 :

TABLEAU 1

Taille des corpus en nombre de tokens

Type	Texte original	16 rétroversions	Totaux
Essai	1062	18 472	19 534
Poèmes	274	4412	4686

Comme on le voit, il s'agit d'un corpus de petite taille, qui pourra donc être traité manuellement mais permettra néanmoins de tirer des observations sur le plan quantitatif.

Une fois tokenisés, les textes ont été lemmatisés et étiquetés avec Treetagger (Schmid 1999). Nous avons tenté ensuite d'effectuer un alignement automatique des versions, mais comme nous allons le détailler plus loin, une segmentation et un alignement manuel se sont avérés plus utiles dans le contexte précis de cette recherche.

3.1. Alignement des textes

La comparaison des rétroversions avec l'original, et entre elles, passe par l'alignement de leurs unités.

L'alignement de corpus parallèle, depuis qu'il a été popularisé par le développement des premières méthodes au début des années 1990 (Kay et Röscheisen 1993; Gale et Church 1991), est une opération désormais bien connue en TAL, pour laquelle de très nombreux algorithmes ont été élaborés, à travers des modèles statistiques ou neuronaux. Aligner implique deux opérations (Isabelle 1992) : d'une part, segmenter les deux textes parallèles en unités plus petites (paragraphe, phrases, syntagmes ou mots), et d'autre part, mettre en correspondance les unités en relation d'équivalence traductionnelle. L'alignement repose en effet sur l'hypothèse de compositionnalité traductionnelle (*ibid.*) : l'équivalence du tout peut se décomposer en l'équivalence de ses parties. En général, on distingue deux familles de systèmes : ceux qui alignent les phrases ou groupes de phrases, et ceux qui alignent les mots ou groupes de mots, ces derniers s'appuyant souvent sur les résultats des premiers.

L'abondante littérature sur le sujet et les progrès des méthodes automatiques pourraient laisser à penser qu'il s'agit là d'une question réglée ; pourtant, toute personne qui s'essaye à pratiquer manuellement l'alignement entre un texte et sa traduction, qu'il s'agisse d'aligner des phrases ou des mots, comprendra rapidement qu'elle avance sur un terrain instable et mouvant.

Prenons l'exemple suivant d'un vers tiré de notre corpus :

des nappes de lumière dans lesquelles on voit des lacs (Macé 2017b)⁸

Et maintenant, observons la rétrotraduction de la traduction en grec ancien du vers de Macé :

[e]t des semblants de lacs dans la lueur étale (Dell’Oro 2017, traduit du grec ancien par Capponi)⁹

Ici, l’alignement des vers est relativement trivial. Si l’on cherche à décomposer plus finement l’équivalence traductionnelle, on pourra faire correspondre les deux parties du vers, en inversant leur séquence :

- des nappes de lumière ↔ dans la lueur étale
- dans lesquelles on voit des lacs ↔ Et des semblants de lacs

Cette fragmentation affaiblit l’équivalence traductionnelle, notamment sur le plan syntaxique, les syntagmes correspondants assumant des fonctions différentes. Mais on peut décomposer plus avant, et trouver des correspondances lexicales :

- des ↔ la
- nappes ↔ étale
- de ↔ ∅
- lumière ↔ lueur
- dans lesquelles ↔ ∅
- on voit ↔ des semblants
- des ↔ de
- lacs ↔ lacs

Bien que l’on puisse établir assez aisément le lien génétique entre les formes originales et les formes issues de la rétrotraduction, on peut se demander légitimement ce que signifient ces équivalences, qui semblent décrire tout un continuum de degrés, allant de l’identité (*lacs* ↔ *lacs*) à des formes distantes à la fois sur les plans sémantique et syntaxique (*nappes* ↔ *étale*, *on voit* ↔ *des semblants*) en passant par des cas intermédiaires de synonymie (*lumière* ↔ *lueur*). Par ailleurs, y a-t-il un sens à associer des mots outils apparemment équivalents (*des* ↔ *la* ou *des* ↔ *de*) quand les constructions syntaxiques ne sont pas isomorphes ?

Il semblerait que la recherche de finesse dans la granularité de l’alignement implique l’affaiblissement de la notion d’équivalence – l’équivalence traductionnelle étant avant tout une relation globale construite par le traducteur sur le plan du message et de ses fonctions communicatives, et non sur le plan des unités qui composent celui-ci. À chercher un grain trop fin, les équivalences s’étiolent dans les sables mouvants de la reconstruction du sens. Ainsi, on se retrouve condamné à osciller entre une relation d’équivalence forte mais entre des segments trop larges (paragraphe, vers ou phrases) ou bien des équivalences fragmentaires, voire absentes, entre des unités plus fines telles que les mots ou les syntagmes.

C’est que la compositionnalité traductionnelle, comme nous l’avions montré dans des travaux antérieurs (Kraif 2002), n’est pas définie a priori à un niveau homogène de segmentation. Dans un même texte, elle peut jouer tantôt dans une phrase entière, tantôt mot à mot, tantôt à un niveau intermédiaire, en fonction des choix de traduction et de l’effet recherché. Ce qui remet en question la notion même d’alignement lexical, telle qu’elle est pratiquée dans le domaine du TAL.

En ce qui concerne ce travail, une contrainte forte s’est imposée pour le choix de la segmentation : comme nous avons cherché à aligner les 17 versions ensemble, en un même multitexte, il n’était pas question d’ajuster la segmentation en fonction des équivalences entre les textes pris 2 à 2, car il n’était pas possible de réunir ensuite ces

différentes segmentations de manière cohérente – sauf à appliquer une clôture transitive qui au final aurait abouti à une granularité trop grossière.

De ce fait, nous avons opté pour un alignement asymétrique vers les textes originaux, qui jouent donc le rôle de pivot, en partant d'une segmentation a priori de ceux-ci, et ceci à deux niveaux :

- segmentation en vers pour les poèmes, segmentation en phrase pour le texte en prose ;
- puis à l'intérieur des segments alignés (vers ou phrases), segmentation en mots ou groupes de mots constituant des unités sémantiques.

3.2. Alignement sur le plan phrastique (phrases et vers)

Comme l'a montré Boboshko (2019), l'utilisation d'aligneurs automatiques tels que Yasa (Lamraoui et Langlais 2013) s'est révélée problématique, car ceux-ci effectuent des regroupements de phrases de façon symétrique (aussi bien du côté du texte original que du côté de la rétroversion), ce qui aboutit à des alignements mettant en jeu des groupements de segments non concordants du texte original (l'application de la clôture transitive aboutissant par la suite à des regroupements trop larges). En outre, les aligneurs automatiques s'appuient sur une segmentation des phrases a priori. Or, dans le cas spécifique de notre corpus (notamment la partie en vers), aligner correctement les rétroversions sur les segments originaux impliquait d'effectuer une segmentation *ad hoc*, qui imposait parfois de fragmenter les segments rétrotraduits, vers ou phrases, d'où la nécessité d'effectuer un alignement préalable à la main. Le tableau 2 illustre la solution adoptée pour représenter l'alignement phrastique : les textes étant segmentés grâce à un balisage structurel (ici <l> représente un vers), la balise simple <milestone/> permet de marquer l'alignement sur les vers du texte source, même si le découpage en vers de la rétroversion n'est pas congruent (un vers de la source pouvant correspondre à plusieurs fragments de vers de la cible).

TABLEAU 2

Alignements des segments (vers)

Texte balisé du texte original	Texte balisé et aligné de la rétroversion LA
<milestone/><l>Un monde où l'on ferrait les chevaux</l>	<milestone/><l>de ceux qui fixent aux canassons les sabots sonores
<milestone/><l>dans l'odeur de la corne, des chevaux battus</l>	<milestone/>dans l'odeur de corne.</l><l>Mais le cheval,
<milestone/><l>qui regardaient droit devant eux, à cause des œillères</l>	<milestone/>regard au sol, fonçait tout droit sans qu'on califourchonne,</l><l>vue couverte et sans lucarne,</l>
<milestone/><l>que n'osaient pas porter les hommes.</l>	<milestone/><l>portant les œillères que les hommes s'épargnent.</l>

Comme on le voit, l'étalon de découpage du texte source est le vers, mais celui-ci ne permet pas toujours d'établir des correspondances consistantes entre vers ou groupes de vers, d'où la nécessité d'opérer un découpage à la main.

Cet exemple présente un cas extrême où la relation d'équivalence est particulièrement vague et inexacte – même en restant dans les vers ou fragments de vers. Le

problème vient souvent d'une différence de séquence dans la répartition des sens, cas de figure que notre système de représentation (points d'ancrage avec <milestones/>) ne permet pas de traiter. On trouve également dans la prose quelques cas, assez rares, où une phrase se trouve enchâssée dans une autre, ce qui rend impossible la mise en place d'une équivalence stricte avec la ou les phrases de l'original. La plupart du temps, on arrive cependant à une correspondance exacte entre les segments (on nommera désormais segments les phrases ou groupes de phrases, vers ou groupes de vers alignés).

3.3. Alignement sur le plan lexical

À l'intérieur des segments ainsi alignés, nous avons procédé à un alignement plus fin sur le plan lexical. Cet alignement des mots ou groupes de mots a été effectué dans un premier temps pour les poèmes, en s'appuyant essentiellement sur des *chunks* ou des unités polylexicales¹⁰.

Le tableau ci-dessous montre un échantillon des alignements obtenus :

TABLEAU 3

Alignements sur le plan lexical (poème)

Original	AR	KO1	KO2
En sortant de	Une fois sortis de	à la sortie	Si tu sors de
la caverne	la caverne	une grotte	la grotte
nous verrons	nous verrons	qui éblouissent	tu verras
les mirages	le mirage	les mirages	le mirage

Pour l'alignement de l'essai, nous avons recherché plus de finesse, en procédant à un découpage intuitif du texte source en unités lexicales ou polylexicales correspondant à des unités sémantiques, et en autorisant la non-contiguïté des mots composant ces unités.

TABLEAU 4

Alignements sur le plan lexical (essai)

Original	DE	DE	AR	AR
L'	L'	L'	l'	l'
absence	absence	absence	absence	absence
d'	d'	d'	d'	d'
écriture	écriture	écriture	écriture	écriture
si	certes pourtant	même si	Bien que	Même si
elle		elle		
ne pas	ne pas	n' pas	ne pas	ne pas
relève d'	relève d'	est fondée sur	soit	relève d'
tout à fait	tout à fait	à strictement parler	forcément	
un	un	une	un	un
choix	choix	décision	choix	choix

Il faut noter que ces différences de granularité dans le découpage n'ont pas d'influence sur les résultats ultérieurs, notamment sur le calcul de la stabilité des

tokens sources, ceux-ci étant par la suite repérés automatiquement à l'intérieur de chaque groupe de mots ainsi délimité manuellement. Par exemple, la conjonction *si* sera considérée comme stable dans 2 versions sur 4 ci-dessus. Si les choix des annotateurs pouvaient varier dans certains cas limites, entre aligner sur une expression distante ou ne pas aligner du tout, ces variations étaient sans incidence sur le calcul de stabilité, qui s'intéresse seulement à la présence ou à l'absence d'un lemme identique dans le groupe aligné.

4. Observations textométriques

Les textes du corpus une fois alignés sur les niveaux phrastique et lexical, toutes nos observations quantitatives vont s'articuler sur un axe unique, celui de la *stabilité traductionnelle*. Par le simple comptage de ce qui se conserve et de ce qui se transforme, il nous sera ainsi possible de mesurer l'influence éventuelle du type de texte, de la langue intermédiaire de traduction, des classes morphosyntaxiques des unités lexicales mises en jeu, voire de leurs propriétés sémantiques.

4.1. Influence de la langue intermédiaire de traduction

La mesure la plus simple pour essayer de corrélérer les rétroversions d'une même langue est celle de la longueur des textes, en nombre de tokens.

TABLEAU 5

Taille des corpus en nombre de tokens

Version	Nombre de tokens			
	Essai		Poèmes	
FR	1062		274	
	V1	V2	V1	V2
DE	1138	1245	275	280
KO	1045	1118	247	297
GRC	1158	1266	236	308
IT	1126	1140	264	265
JA	1137	1168	235	273
LA	1143	1171	297	293
FA	1110	1187	280	264
Moyenne	1154,5		275,75	

Les valeurs en gras correspondent aux deux extrêmes, pour chaque série. On constate que pour l'essai, les rétroversions sont presque toutes plus longues que l'original. Ce phénomène était plutôt attendu, et correspond à un des « universaux de traduction » mentionné par Baker (1993) et corroboré depuis par de nombreux auteurs, concernant une tendance des traductions à accroître le niveau d'explicitation. On en donnera des exemples dans la dernière partie.

Ce qui est plus marquant, c'est que ce phénomène n'est pas confirmé pour les traductions poétiques, dont la longueur moyenne est à peu près identique, malgré des variations locales. Contrairement à un texte de type argumentatif, le texte poétique ne semble pas requérir ce travail d'explicitation : les images poétiques, en effet,

fonctionnent par la condensation de plusieurs dimensions (sonore, sensorielle, symbolique, etc.) qui ne peuvent être traduites simultanément que par la recherche d'une image équivalente, tout aussi condensée et économe dans ses moyens.

En ce qui concerne l'effet des langues traduites, on constate une corrélation linéaire assez nette entre les deux séries V1 et V2, avec un coefficient de Pearson de 0,69 pour l'essai¹¹. Autrement dit, pour l'essai, il semblerait qu'il y ait un effet de la langue intermédiaire: on constate que les deux versions du grec ancien et de l'allemand sont plutôt longues, tandis qu'à l'opposé les deux versions de l'italien et du coréen sont plutôt courtes. Pour interpréter cet effet, il faudrait une étude contrastive plus approfondie, en lien avec les caractéristiques morphosyntaxiques de chaque langue. Notons cependant que la poésie, là encore, semble échapper à ce constat, car on n'observe aucune corrélation linéaire significative – et le faible volume de données ne permet pas d'en tirer de conclusion.

Un deuxième indicateur de la stabilité lexicale, pour chaque version, est le calcul du Dice entre segments alignés. La mesure du Dice indique la proportion de vocabulaire partagé entre deux segments (sans tenir compte de l'ordre des mots), et se calcule comme le nombre de lemmes communs divisé par la moyenne des longueurs des segments¹²:

$$Dice = \frac{2|Seg1 \cap Seg2|}{|Seg1| + |Seg2|}$$

|Seg1| et |Seg2| désignent respectivement les longueurs des segments 1 et 2. Pour l'essai, si on calcule le Dice entre toutes les versions (originale et rétroversions), on trouve une valeur moyenne de 0,553. Si on compare entre elles seulement les versions issues des mêmes langues, on trouve un score de 0,575 – les versions semblent donc légèrement plus similaires entre elles quand elles sont issues d'une même langue, mais la différence n'est pas statistiquement significative, comme l'indique un calcul de t-test (pour une p-valeur à 0,05). Pour les poèmes, on trouve des valeurs similaires: 0,54 en moyenne entre toutes les versions, et 0,56 pour les versions issues d'une même langue. Là encore, la différence n'apparaît pas comme statistiquement significative.

En revanche, les indices Dice entre les deux séries de rétroversions apparaissent comme linéairement corrélés, avec un coefficient de 0,523 pour la prose et 0,399 pour la poésie. Il apparaît, pour la poésie comme pour la prose, que les versions se rapprochant le plus de la version originale sont, par ordre décroissant de similarité: DE > IT > AR > FA – et ceci dans le même ordre pour les deux types de texte. Le grec ancien, le latin et le japonais apparaissent tous les trois parmi les versions les moins similaires obtenant, pour l'essai comme pour le poème, des scores inférieurs à 0,60.

L'hypothèse d'une influence de la langue de rétrotraduction s'en trouve donc renforcée – bien qu'il soit difficile d'en déterminer les causes sans un examen minutieux (on voit notamment que des considérations strictement génétiques auraient placé les versions latines comme étant plus proches, par rapport à des langues comme l'arabe ou le farsi, ce qui n'est pas le cas).

Un autre phénomène intéressant se dégage de ces mesures de similarité: en moyenne, c'est le texte original qui apparaît comme la version la plus semblable à toutes les autres, dans ses comparaisons avec les 16 rétroversions, avec un Dice moyen de 0,639 pour la prose et 0,657 pour la poésie. Si on regarde de plus près chacune des 16 versions rétrotraduites, c'est la version originale qui apparaît systématiquement

comme la plus proche. Ce qui veut dire que les textes originaux, dans les deux cas, se distinguent des rétroversions, sur un strict plan textométrique, par leur centralité dans cet ensemble : si on ne savait pas quel texte est la version originale, on pourrait le retrouver aisément grâce à ces mesures.

Par ailleurs, si on classe les langues en fonction du Dice moyen qu'elles obtiennent dans l'ensemble des comparaisons, on retrouve pratiquement le même ordre que dans la comparaison avec l'original : plus une langue est proche de l'original, à travers ses deux versions, plus elle est proche en moyenne de l'ensemble des autres versions, comme on le voit dans le tableau ci-dessous.

TABLEAU 6

Indice Dice moyen entre versions par langue de rétrotraduction (ou langue originale)

Langue intermédiaire	Dice moyen dans la comparaison avec toutes les autres versions	Dice moyen dans la comparaison avec l'original
FR original	0,639	–
DE	0,593	0,736
IT	0,580	0,720
AR	0,569	0,681
KO	0,529	0,595
JA	0,529	0,589
LA	0,522	0,597
GRC	0,493	0,557

D'un point de vue traductologique, cela semble signifier que toutes les rétroversions divergent de l'original dans des directions différentes. Aucune rétroversion ne se rapproche plus d'une autre rétroversion que de l'original – aucune convergence n'est observée – pour toutes les versions sans exception, la transformation opérée par la rétrotraduction est strictement centrifuge.

4.2. Stabilité en fonction des lemmes

On peut supposer que certains types de lemmes ont vocation à être plus stables que d'autres. Laplace rappelle une célèbre métaphore en citant les travaux de Seleskovitch :

[e]n étudiant non seulement l'interprétation proposée par ses collègues mais également les notes de consécutive qu'ils avaient prises, Seleskovitch constate que certains mots du discours original sont notés et traduits par les participants. Ce sont les chiffres, les appellations, les énumérations et les termes techniques. Par contre d'autres mots, qui possèdent ce qu'elle avait appelé dans *L'interprète dans les conférences internationales* des équivalents conventionnels dans l'autre langue, n'avaient été ni notés ni traduits tels quels. Fondus dans l'opération de chimie du sens, ils avaient fait l'objet d'une réexpression. (Laplace 1994: 239)

Comme des « raisins dans la brioche », ces mots résistent à la traduction, comparée à une opération de cuisson, pendant que d'autres se transforment et se recombinent comme les ingrédients de la pâte. À partir des notes prises par les traducteurs lors d'interprétations consécutives, Seleskovitch conclut en effet que ce qui pousse les interprètes à noter certains mots, et non d'autres, ce n'est pas l'existence d'une

équivalence conventionnelle dans la langue cible « mais la conscience que le mot entendu a une personnalité propre, indépendante du message » (Laplace 1994 : 239).

Pour calculer la stabilité des lemmes, nous avons pris chacun des tokens de la source (respectivement 253 et 939 pour la poésie et pour la prose, hors ponctuation), puis nous avons déterminé si le lemme apparaissait à l'identique dans le groupe de tokens alignés (sur la base des appariements manuels), dans chaque version rétrotraduite. La stabilité exprime en pourcentage la proportion de versions où le lemme est conservé. Par exemple, un lemme qui est aligné à lui-même dans les 16 versions rétrotraduites obtient un score de 100 %. S'il apparaît dans seulement 4 versions, le score tombe à 25 %.

Dans une première série d'observations, nous nous sommes demandé si la tendance à la stabilité était liée à la dichotomie entre mots pleins et mots outils. Nous avons donc calculé les stabilités moyennes en fonction des parties du discours.

TABLEAU 7

Stabilité moyenne des lemmes pour les deux textes

Poème	Stabilité	Essai	Stabilité
Numéral	100,00 %	Nom propre	89,29 %
Nom	73,20 %	Numéral	77,50 %
Déterminant	72,74 %	Déterminant	69,96 %
Conjonction	68,75 %	Nom	62,70 %
Nom propre	68,75 %	Conjonction	62,08 %
Adjectif	56,88 %	Adjectif	56,60 %
Préposition	55,34 %	Préposition	55,41 %
Adverbe	54,86 %	Pronom	51,97 %
Verbe	52,08 %	Adverbe	51,79 %
Pronom	47,92 %	Verbe	47,59 %

Pour les poèmes, certaines statistiques s'appuient sur des données trop petites pour être concluantes (p. ex. on a un seul nom propre et deux adjectifs numéraux). Néanmoins, on observe une certaine convergence entre les deux types de texte : globalement – et comme suggéré dans la citation de Laplace (1994) –, les noms communs, les noms propres et les chiffres apparaissent comme les plus stables. À l'opposé, les pronoms, les verbes et les adverbes apparaissent comme plus instables – ceux-ci sembleraient donc plus exposés à se fondre dans la « chimie du sens ». Étonnamment, certaines catégories de mots outils semblent plutôt stables, comme les déterminants et les conjonctions. Concernant les déterminants, dans la mesure où ceux-ci dépendent des noms, il n'est pas étonnant qu'ils en héritent la stabilité. Pour les conjonctions, on peut supposer que leur rôle de charnière syntaxique dans l'architecture globale des phrases pourrait leur conférer une certaine robustesse.

Quant aux verbes, leur stabilité semble dépendre étroitement de leur emploi, et notamment des temps et modes utilisés.

TABLEAU 8

Stabilité moyenne des verbes en fonction des temps et modes, pour l'essai

Mode - temps	Stabilité
Indicatif futur	62,50 %
Indicatif imparfait	58,20 %
Indicatif présent	48,36 %
Indicatif passé simple	47,92 %
Infinitif	46,99 %
Participe présent	46,88 %
Participe passé	42,61 %
Conditionnel	41,67 %
Subjonctif présent	31,25 %

Le mode indicatif apparaît comme plus stable, peut-être parce qu'il correspond aux emplois narratifs des verbes : on constate que pour le futur et l'imparfait, la stabilité est voisine des noms. À l'opposé, le conditionnel et le subjonctif apparaissent comme très instables.

Si on examine les lemmes individuellement, voici ceux qu'on trouve comme les plus stables :

- Essai : Conrad, Noir, mer, Rouge, Alexandrie, contre, avant, Océan, Indien, Asie, Park, Champollion, japonais, poète, fond
- Poème : prophète, pharaon, mer, lèvres, femme, deux, ou, trois, quatre, sang, œillère

Outre les numéraux et les entités nommées, on y trouve des noms à la fois fréquents et généraux, comme *mer*, *femme*, *poète*, mais aussi des noms plus rares et spécialisés comme *œillère* ou *hiéroglyphe*. À l'opposé, on trouve aussi des noms instables, comme *festin*, *campagne* et *nappe*, pour les poèmes, *breuvage*, *initiation*, *marque* ou *vase* pour l'essai. Des critères purement lexicaux, tels que la fréquence, le caractère concret ou abstrait ou la polysémie semblent insuffisants pour expliquer ce qui fait que certains noms sont moins stables.

Les données textométriques permettent de formuler des hypothèses d'ordre général, en fonction du type de texte, des langues intermédiaires ou des types de lemmes. Mais, lorsqu'on cherche à les interpréter, ces données conduisent souvent à revenir au texte de départ, afin d'examiner précisément les facteurs multiples, pas toujours convergents, qui expliquent la stabilité ou l'instabilité traductionnelle.

5. Interprétation des données textométriques comme indicateurs herméneutiques

C'est à cet effet de retour que nous nous intéresserons maintenant, considérant le corpus rétrotraduit comme susceptible d'éclairer, de manière rétroactive, les textes originaux. Pour cela, on utilisera les statistiques dans leur globalité, sans prendre en compte les différentes langues intermédiaires, mais à une échelle inférieure au texte : la phrase ou le contexte étroit d'occurrence d'un lemme.

5.1. La longueur des phrases

On a vu qu'à l'échelle des textes, les rétroversions de l'essai ont davantage tendance à l'allongement que celles des poèmes. Cela se vérifie à un niveau inférieur : dans le corpus poétique, le rapport moyen des phrases rétrotraduites aux phrases originales est situé entre 0,90 et 1,09 ; dans la prose, il est situé entre 0,89 et 1,46. Le tableau ci-dessous fait apparaître la proportion des phrases allongées ou raccourcies dans chaque type de texte :

TABLEAU 9

Proportion des phrases allongées ou raccourcies selon 4 intervalles du rapport des longueurs

Rapport moyen de longueur	≤ 0,90]0,9;1]]1;1,10[≥ 1,10
Poèmes (10 phrases)	10 %	50 %	40 %	0 %
Essai (36 phrases)	2,8 %	13,9 %	44,4 %	38,9 %

L'allongement correspond souvent à des phénomènes d'explicitation qui, indirectement, pointent des difficultés interprétatives dans le texte de départ. Ces phénomènes sont similaires à ceux que Berman, dans une autre optique, décrit dans son « analytique de la traduction » : la version traduite procède, par l'allongement, à « un dépliement de ce qui, dans l'original, est “plié” » (Berman 1999 : 56). On le constate par exemple dans la phrase suivante, plus particulièrement la partie en italiques (le rapport moyen des rétroversions au texte de départ est noté entre parenthèses) :

[s]ans parler du latin d'église ou du grec des apothicaires, il suffit de se rappeler les interprétations farfelues à propos des hiéroglyphes, caractères sacrés traduisant les décrets du ciel, ou les révélations d'une religion perdue, jusqu'à ce que Champollion vienne dissiper ces faux mystères. (Macé 2017c)¹³

La relation entre le groupe nominal *les interprétations farfelues à propos des hiéroglyphes* et l'apposition *caractères sacrés [...]* pose des problèmes de compréhension : sur le plan syntaxique, celle-ci dépend du nom *hiéroglyphes* ; mais sur le plan du sens, elle développe en réalité *les interprétations farfelues* (précisément, elle explique ce qu'étaient ces *interprétations* et pourquoi elles étaient *farfelues*). Dans le texte de départ, la discordance entre syntaxe et sens complexifie les relations entre les mots, en partie implicites. Presque toutes les rétroversions les explicitent, allongeant la phrase, par exemple :

il suffit de se représenter les interprétations inouïes des hiéroglyphes : *on a voulu y voir des lettres saintes qui [...]* (Geiser & Béguin 2017, traduit de l'allemand par Rimasson et Schlie)¹⁴

[i]l n'y a qu'à se souvenir des interprétations farfelues *que l'on faisait* des hiéroglyphes, caractères sacrés *censés* traduire [...] (Dell'Oro 2017, traduit du grec ancien par Capponi)¹⁵

Parfois, la complexité du texte original est perdue dans la rétroversion, qui gomme la relation sémantique entre l'apposition et l'adjectif *farfelues* :

il suffit de rappeler les interprétations farfelues des hiéroglyphes, *qui sont* des lettres sacrées [...] (Gassouma, 2017, traduit de l'arabe par Abdi)¹⁶

On peut juger cette explicitation erronée, mais elle fait apparaître ce qui reste un possible interprétatif de la phrase, si l'on se limite à sa syntaxe sans être sensible aux

autres liens. Les rétroversions font ainsi apparaître par contraste le caractère elliptique du texte de départ, et mettent au jour les difficultés qu'il peut générer.

La phrase du corpus qui subit l'allongement le plus marqué l'illustre également :

[d]es manuscrits en guèze, langue morte devenue langue rituelle comme le latin en Occident. (Macé, 2017c)

Les rétroversions recourent très souvent à des tournures explicitant les relations entre les constituants : elles sont nombreuses à rétablir une phrase verbale (*Ces manuscrits sont écrits en guèze...*), à développer, par une subordonnée, la relation de prédication seconde de l'apposition à son support (*en guèze, qui est une langue morte...*) ou de l'épithète (*en guèze, une langue morte qui, comme le latin en Occident, est devenue...*). En outre, les rétroversions insèrent fréquemment des mots expliquant le sens de *guèze*, par exemple un participe passé (*Des manuscrits rédigés en guèze...*), un nom («*en langue guèze*»), ou encore un développement périphrastique (*en guèze, la langue d'Éthiopie*). Ce procédé d'explicitation est certes fréquent dans les textes traduits, mais rien ne le justifie ici : le *guèze* n'est pas plus connu pour un lecteur français que pour un lecteur de l'italien, du coréen ou du farsi ; de surcroît, le contexte est suffisant, puisque les phrases précédentes de l'original mentionnent explicitement *l'Éthiopie, où l'écriture existe depuis l'Antiquité* et que le mot *manuscrit* et la préposition *en* (*Des manuscrits en guèze*) suffisent pour comprendre qu'il s'agit d'une langue écrite.

Qu'il s'agisse d'expliciter la relation entre constituants de la phrase ou le sens d'un mot, on constate que le rapport au lecteur est bien différent dans le texte de départ et dans la quasi-totalité des rétroversions : alors que l'écrivain sélectionne un destinataire capable d'interpréter et de saisir des liens parfois implicites, excluant le lecteur incompetent ou inattentif, les traducteurs manifestent, par l'explicitation, un rapport au destinataire tout autre, dans lequel semble dominer le désir de produire un texte transparent. Cette divergence pragmatique entre l'original et les rétroversions, qui peut aussi s'interpréter comme une différence de posture, est peut-être l'une des raisons pour lesquelles le corpus rétrotraduit fonctionne ici comme indicateur herméneutique et comme outil d'analyse : on peut supposer que les traducteurs ont été d'une part d'attentifs lecteurs, sensibles aux difficultés d'interprétation, et de scrupuleux rédacteurs, soucieux d'élucider, pour le lecteur, ces difficultés et/ou forcés de le faire en raison du passage d'une langue à l'autre.

Les cas où les phrases sont raccourcies peuvent aussi faire ressortir, par contraste, des caractéristiques du texte de départ, par exemple dans ces strophes, extraites de deux poèmes :

[a]jutant de morts à la guerre
 qui ont vu dans *leurs yeux, leurs yeux* hagards,
 la campagne en fleurs derrière les barbelés.
 [...]
 Les étincelles, *c'est* le marteau du philosophe
 qui les fait jaillir en tapant sur le vieux monde. (Macé 2017b)

Dans le corpus rétrotraduit, la première phrase est souvent réduite par suppression de la répétition et la seconde par transformation de la tournure clivée (qui compte, par rapport à la tournure directe, 4 mots supplémentaires, en italiques). Les rétroversions tendent alors à gommer les effets d'insistance et d'oralité que génèrent

ces deux formes, d'autant plus perceptibles qu'ils sont fondés sur un écart par rapport à la norme constituée par le texte de départ lui-même : cette répétition est unique en son genre dans l'ensemble des quatre poèmes, de même que la tournure emphatique, caractéristique de la langue orale et familière alors que le style est très écrit et soutenu. Même si ces transformations peuvent s'expliquer en partie par des caractéristiques morphosyntaxiques et stylistiques des langues intermédiaires, il reste que le « lissage » opéré par les traducteurs indique la difficulté à interpréter ce qu'on pourrait appeler une « aspérité stylistique » : comment comprendre cette répétition marquée, ce changement brutal de niveau de langue ?

La longueur relative des rétroversions par rapport au texte de départ permet, quand on se place au niveau des phrases, de déceler des zones du texte générant une instabilité traductionnelle, que manifeste l'allongement ou la réduction. Cet indicateur serait encore plus efficace si l'on pouvait le mobiliser à une échelle inférieure à la phrase, car il arrive fréquemment que seule une partie soit significativement allongée ou raccourcie – mais il faudrait alors réaliser un alignement intermédiaire entre le lemme et la phrase. Les données peuvent en revanche aisément être utilisées à une échelle plus large, par combinaison de phrases successives, et permettre d'identifier de plus vastes séquences textuelles générant de la stabilité ou de l'instabilité traductionnelle. On se placerait alors à une échelle textuelle de même ordre que celle à laquelle Berman se situe lorsqu'il propose d'identifier intuitivement, dans un texte traduit, indépendamment de la comparaison avec l'original, des « zones textuelles » qui, à la lecture, semblent « problématiques » ou, à l'inverse, « miraculeuses » (Berman 1995 : 66). Dans le texte en prose du corpus, par exemple, on voit que les phrases les plus allongées dans les rétroversions se situent au début du texte, notamment dans les deux premiers paragraphes, la moyenne se rapprochant graduellement de 1 au fil du texte, de manière non pas continue mais tout de même nettement tendancielle.

5.2. La stabilité des lemmes

La stabilité moyenne d'un lemme à une place donnée constitue également un indicateur intéressant, particulièrement pour ce qui concerne les mots pleins. Le fait qu'un lemme soit stable peut être dû aux propriétés du mot (par exemple la spécialisation de *hiéroglyphe*) et/ou à celles de la séquence dans laquelle il est inclus. Dans le premier poème, par exemple, ces deux vers sont stables sur le plan lemmatique :

[p]as de mer Rouge qui s'ouvrirait d'elle-même
Comme les grandes lèvres des femmes (Macé 2017b)

Les mots pleins ont une stabilité moyenne élevée, de presque 77% (*mer*: 100%, *rouge*: 93,75%, *ouvrir*: 4%, *grande*: 62,5%, *lèvre*: 100%, *femmes*: 100%). La stabilité du mot *lèvres* retient l'attention, car son interprétation en contexte n'est pas évidente, deux sens concrets du mot pouvant être actualisés : « partie formant le contour de la bouche » (1), « partie du sexe féminin » (2). Dans l'expression *grandes lèvres*, l'adjectif peut, en fonction du sens du nom, être interprété soit comme un qualificatif (sens 1), soit comme formant avec le nom une locution (sens 2). Il est plus instable dans les rétroversions. Les variantes que l'on trouve dans celles-ci témoignent d'interprétations divergentes, qui sélectionnent un seul sens de *lèvres*: *grosses*, *immenses*, *rondes* (sens 1); *secrètes* (sens 2). Mais, malgré les deux interprétations qu'il peut recevoir, le nom

reste très stable en raison de ses propriétés en langue (la polysémie) et en contexte (celui-ci permet le maintien de la polysémie aussi bien que la restriction à l'un des sens).

Il arrive également qu'on relève une tension entre la stabilité lemmatique et l'instabilité sémantique globale d'une séquence : autrement dit, les mots peuvent être maintenus à l'identique et la séquence avoir un sens différent. C'est ce qu'on observe pour la phrase suivante, extraite de la prose. Avec une stabilité moyenne de 0,63 %, elle figure parmi celles qui sont le plus stables :

[c]ent ans plus tard, à l'ouest du Cameroun, le sultan Njoya fera boire à ses sujets les caractères à peine effacés, dispersés dans l'eau, de l'écriture qu'il vient d'inventer. (Macé 2017c)

Les lemmes qui sont moins stables que la moyenne (en particulier *caractères* et *dispersés*) le sont pour des raisons de substitution synonymique : il s'agit donc d'une phrase très stable dans son ensemble. Pourtant, il existe dans d'assez nombreuses rétroversions d'importantes distorsions de sens global, notamment en ce qui concerne l'organisation temporelle, ou du moins une certaine confusion, comme par exemple :

[b]ien plus tard, dans une région située vers le couchant, le sultan Njoya donnera à boire à ses serviteurs des lettres qu'il viendra de découvrir, légèrement effacées et mélangées à de l'eau. (Cusset 2017, traduit du grec ancien par Kolde)¹⁷

La phrase d'origine est d'une certaine complexité à cet égard, avec une périphrase factitive exprimant le futur dans un contexte narratif passé (*fera boire*), l'expression d'un passé proche (*à peine effacés*) et une périphrase temporelle au présent de narration (*vient d'inventer*). Elle oblige le lecteur à rétablir pour les procès un ordre temporel (*inventer – disperser – faire boire*) qui est l'inverse de celui dans lequel ils sont présentés dans la phrase. Les rétroversions témoignent de cette difficulté, tout en maintenant le lexique de la phrase originale, ou un lexique très proche.

Le phénomène est analogue pour la strophe qui clôt le quatrième poème :

[u]n monde où l'on ferrait les chevaux
dans l'odeur de la corne, des chevaux battus
qui regardaient droit devant eux, à cause des œillères
que n'osaient pas porter les hommes. (Macé 2017b)

Les lemmes correspondant à des mots pleins dans les deux derniers vers sont plutôt stables (*regarder* : 75 %, *droit* : 68,75 %, *œillère* : 100 %, *oser* : 62,5 %, *porter* : 68,75 %, *homme* : 75 %). La stabilité est maximale dans cette rétroversion :

[...] des chevaux battus
Regardant tout droit devant eux à cause des œillères
Et n'osant pas porter les hommes. (Gassouma 2017, traduit de l'arabe par Abdi)¹⁸

Les lemmes correspondants à des mots pleins sont tous identiques à l'original. Pourtant, le sens du dernier vers est très différent, en raison de la redistribution des actants autour du noyau verbal, qui, si l'on simplifie les structures syntaxiques, est la suivante : les hommes n'osaient pas porter les œillères ↔ des chevaux battus n'osaient pas porter les hommes.

Plusieurs autres rétroversions opérant des modifications similaires, on peut supposer que c'est le texte de départ qui favorise ces distorsions et essayer de comprendre pourquoi. Le réseau lexical de la strophe est très cohérent (le champ lexical

équestre) et, dans les trois premiers vers, la mise en scène des relations entre les hommes et les chevaux est polarisée selon une opposition *dominant/dominé*. Le dernier vers opère une volte : les œillères, accessoire des chevaux, sont attribuées aux hommes (sur le mode virtuel) ; la négation de l'auxiliaire *oser* entre en tension avec le pouvoir précédemment associé à la position dominante des humains. Par ailleurs, on s'attendrait à ce que le verbe *porter* ait pour sujet *les chevaux* et pour complément *les hommes*. Ce n'est pas le cas, mais, en plaçant le sujet, *les hommes*, à droite du verbe, le texte renforce le brouillage généré par la volte et indique, tout en ne la réalisant pas, la distribution actantielle attendue. La rétroversion citée, ainsi que d'autres, manifeste des distorsions que l'on peut évaluer comme des erreurs de traduction mais qui sont intéressantes car elles indiquent, sous une forme amplifiée, le brouillage à l'œuvre dans le texte. Dans ce passage, la stabilité des lemmes, qui s'explique entre autres par la cohérence du réseau lexical et par le fait que les mots sont utilisés dans leur sens propre, entre en tension avec l'instabilité des rôles actantiels, due à des effets de structure (la volte, la postposition du sujet), autrement dit à la manière dont les mots sont mis en relation les uns avec les autres.

Les lemmes peu ou pas stables peuvent permettre de relever des zones textuelles sujettes à une distorsion traductionnelle, en particulier lorsqu'un mot est le support d'une métaphore, comme *nappe* qui apparaît à deux reprises (les strophes appartiennent à des poèmes distincts) :

[e]n sortant de la caverne, nous verrons les mirages
qu'on rêve depuis toujours de traverser à pied sec :
des flaques d'eau qui s'évaporent quand on avance,
des *nappes* de lumière dans lesquelles on voit des lacs.

[...]

Avec la neige la *nappe* est mise
sur les prés. Le ciel tout entier
pourrait tenir dans une cuiller en argent,
le ciel et deux ou trois soleils
sur les quatre que comptaient les Aztèques. (Macé 2017b)

Nappe est très instable dans son premier contexte d'apparition (12,5 %), moyennement dans le second (56,25 %). La diversité des variantes dans les rétroversions s'explique par la polysémie de *nappe*, qui peut être entendue dans son sens propre (« linge que l'on étend sur la table pour prendre les repas ») ou dans divers sens métaphoriques.

Dans la première occurrence, le contexte ne permet pas d'actualiser le sens propre, mais mobilise un sens métaphorique lexicalisé (« vaste étendue d'eau, de fluide, à la surface du sol, sous terre »). Cette occurrence est marquée par la plus grande instabilité du lemme, en raison du jeu avec la collocation : l'expression *nappes d'eau*, attendue, est remplacée par *nappes de lumière*. Cette substitution établit une analogie implicite eau/lumière et complexifie la compréhension, ce dont témoignent les noms que l'on trouve dans les rétroversions :

- (1) *étendue* (4 occ.), *couche* (2), *strate*, *étale*
- (2) *puits*, *source*
- (3) *plaine*, *feuille*, *nuée*
- (4) *s'ourdir*

Dans la majorité des cas (1), les mots choisis comprennent le sème /surface horizontale/. D'autres proposent des images qui éclairent, de manière rétroactive, le texte de départ: la série (2) insiste sur l'analogie eau/lumière; la série (3) inclut le mot dans l'isotopie du paysage; la relative *qui s'ourdissent* (4) attire l'attention sur la polysémie de *nappe* en référant à son sens propre de textile (*ourdir* signifie, au sens premier, «disposer sur l'ourdissoir les fils de la chaîne réunis en nappe»). Quelle que soit par ailleurs leur justesse sur le plan de la traduction, les rétroversions font apparaître une série d'interprétations en germe ou en puissance dans le texte de départ.

Dans la seconde occurrence du mot, le sens propre est mobilisable (en raison de l'expression *mettre la nappe* et des mots *cuiller, festin, repas*), ce qui explique sa plus grande stabilité, même si son incompatibilité avec le référent décrit (un paysage naturel: *neige, prés, ciel, soleil*) peut provoquer une interprétation métaphorique. On note que seul un quart des rétroversions répète à l'identique le mot utilisé à la première occurrence, favorisant la structure formelle (la répétition) plutôt que sémantique (la divergence des emplois).

L'instabilité d'un lemme peut être due à d'autres facteurs qu'un emploi métaphorique, comme dans la phrase suivante, parmi celles de la prose qui sont les plus instables:

[*Les résistants, les clandestins, les êtres pourchassés* prenaient la précaution d'avaler les documents qui auraient pu les compromettre. (Macé 2017c)

La faible stabilité des lemmes s'explique ici par des phénomènes de synonymie, notamment au sein de la série initiale des trois groupes nominaux. Voici les variantes présentes dans les rétroversions pour les mots pleins:

- (1) *résistants* (50%) ↔ *séditieux, dissidents, rebelles, opposants, partisans, ceux qui combattent l'oppression, armée de la Résistance, forces de résistance*
- (2) *clandestins* (62,5%) ↔ *hors-la-loi, criminels, qui vivent dans l'ombre, saboteurs*
- (3) *êtres* (0%) ↔ Ø, *personnes, sujets, ceux*
- (4) *pourchassés* (6,25%) ↔ *persécuté(e)s, fugitifs, proscrits, qu'on chassait de leur patrie, qui étaient recherchés*

Ici, les phénomènes de synonymie et, plus largement, de proximité sémantique, sont rendus plus sensibles par le fait qu'ils n'engagent pas seulement l'axe paradigmatique (la série des mots alternatifs) mais aussi l'axe syntagmatique puisque les expressions, quoique distinctes, partagent certains sèmes.

Sur le plan du sens, l'accumulation ne désigne pas, dans le texte d'origine, trois classes d'humains différentes, mais deux, suivies d'une expression hyperonyme englobant les précédentes (*résistants* et *clandestins* ont comme sèmes communs ceux qui sont repris par *êtres pourchassés*; on pourrait ainsi paraphraser la série: «les résistants, les clandestins, bref tous les êtres pourchassés»). Une difficulté interprétative est générée par cette structure, que manifeste la divergence des propositions dans les rétroversions. La difficulté tient, une fois de plus, à la propension du texte à ne pas expliciter les liens logiques. Elle tient également, dans ce cas précis, à des facteurs formels, qui entrent en jeu dans la perception des groupes: le rythme ternaire génère une forme d'assertivité qui inhibe partiellement, chez le lecteur, la recherche de la répartition sémantique (prévalence de la «forme» sur le «fond»), d'autant plus qu'il est redoublé (*résistants, clandestins* et *pourchassés* comportent trois syllabes) et associé à un jeu phonique de répétition et de variation (avec en particulier les nasales

[ã] et [ê], les voyelles [e] et [ɛ], les consonnes [s], [z] et [t]). On peut supposer que la forte instabilité traductionnelle du dernier terme est liée à l'ensemble de ces caractéristiques du texte de départ, concernant la structuration sémantique, logique et formelle de la séquence.

La conjonction de facteurs sémantiques, formels et stylistiques explique également la très faible stabilité du lemme *vase* (6,25%), dans le contexte suivant :

[j]’avais mis dans une bouteille d’eau les dernières paroles de ma femme. Les lettres noires se diluèrent bientôt, et l’eau devint couleur de *vase*. (Macé 2017c)

Les nombreuses variantes proposent un panel interprétatif intéressant :

(1) *trouble* (3 occ.), *se troubler* (2 occ.), *rosâtre* (1 occ.)

(2) *boue* (3 occ.), *boueuse* (3 occ.)

(3) *bouteille* (1 occ.), *pot* (1 occ.), *flacon* (1 occ.)

La série (1) ne retient du mot d’origine que l’idée d’impureté, que l’on retrouve dans le suffixe de *rosâtre*. La *boue* (2), mélange de terre et d’eau, est proche sémantiquement de la *vase*, qui, elle, contient des particules organiques en décomposition. Ce dernier détail n’est pas anodin, car on peut supposer que si l’écriture dissoute est décrite par analogie avec la vase, c’est en raison de la polysémie de *feuille*, le texte décrivant une feuille (végétale/de papier) en décomposition. On peut même faire l’hypothèse que le texte de départ oblitère sa matrice, en évitant le mot dans la phrase précédente : *J’avais mis dans une bouteille d’eau [une feuille sur laquelle étaient écrites] les dernières paroles de ma femme*. Mais cette interprétation, fondée sur la polysémie d’un terme absent, n’est soutenue par aucun autre indice contextuel et n’apparaît pas dans les rétroversions, ce qui signale son caractère incertain et discutable, au moins sur le plan de la réception. La série (3) met au jour une autre difficulté interprétative, l’homonymie *vase* (n. m.)/*vase* (n. f.), rendue sensible, dans le texte, par l’absence d’article (empêchant l’identification du genre) et par la présence, dans la phrase précédente, de l’expression *bouteille d’eau*, dont *vase* (n. m.) pourrait être une anaphore infidèle.

Ce qui importe ici n’est pas d’évaluer le catalogue des interprétations fournies par les rétroversions, mais de les exploiter comme une série de points de vue sur le texte. Leur divergence indique une zone d’instabilité de la réception, donc de la traduction, et permet d’identifier des interprétations dominantes (l’impureté, la *boue*), possibles mais minoritaires (l’homonymie de *vase*) ou absentes (la polysémie de *feuille*).

6. Conclusion

Le corpus élaboré dans le cadre du projet « Épreuves de l’étranger », bien que de taille réduite, constitue un objet fascinant et complexe. Avec ses 9 langues (en comptant le français), dont certaines typologiquement et génétiquement éloignées, et les 38 traducteurs qu’il a impliqués, il constitue un multitexte complexe qui engage, on l’a vu, de nombreux phénomènes tant sur les plans contrastif qu’herméneutique. Le prisme de la stabilité traductionnelle, qui nous a permis de dégager des données textométriques quantitatives, ouvre la voie à de nombreuses pistes de recherche.

Ainsi, nous avons pu montrer que les langues impliquées dans la rétrotraduction exerçaient un effet déterminant sur la stabilité considérée globalement : ce que

montrent tant les rapports des longueurs que le calcul de proximité lexicale. Cet effet requiert cependant, pour être expliqué, des recherches plus approfondies : contrairement à notre hypothèse initiale, la stabilité globale n'est pas immédiatement liée à la proximité génétique des langues impliquée, le latin ou le grec ancien obtenant des scores beaucoup plus faibles que l'arabe ou le farsi. Par ailleurs, la poésie ne semble pas aboutir à des distorsions plus importantes que la prose, en ce qui concerne aussi bien le Dice que le nombre de mots : là encore, notre hypothèse de départ a été battue en brèche.

Sur le plan de la stabilité des lemmes, des tendances générales se dégagent : les noms, les numéraux et les entités nommées apparaissent beaucoup plus stables que les adverbes ou les verbes. Ces mots avec « une personnalité propre », pour reprendre l'expression de Seleskovitch (1975), manifestent des rapports extralinguistiques qui échappent aux fluctuations des contrastes linguistiques. De manière plus inattendue, les conjonctions apparaissent comme relativement stables – et manifestent peut-être des régularités textuelles qui dépassent également le simple plan contrastif.

Mais un examen minutieux des principaux lieux de stabilité ou d'instabilité, pour ce qui est des phrases (et toutes langues confondues), montre que ce sont surtout des phénomènes interprétatifs locaux, liés à des configurations syntaxiques particulières, à des métaphores aux multiples facettes, à des ambiguïtés lexicales non résolues, qui engendrent des fluctuations dans les choix de traduction. Dans certains cas, c'est la posture même du traducteur, cherchant à restituer une lecture fidèle tout en l'élucidant, qui apparaît comme principal facteur de distanciation avec l'original.

Au final, chacune de ces versions livre une lecture originale du texte source – et de façon étonnante, ces lectures ne semblent pas converger dans une direction définie. Car si sur le plan textométrique toutes les versions s'éloignent de l'original, elles se rapprochent beaucoup plus de celui-ci que des versions produites par les autres traducteurs. Si l'on a parfois décrit une tendance à la normalisation dans la traduction (Baker 1996), pouvant se manifester par un plus grand degré d'homogénéité des textes traduits par rapport aux textes originaux (Laviosa 2002 : 72), force est de constater que les fluctuations interprétatives observées dans ce type de traduction littéraire aboutissent, bien au contraire, à un véritable déploiement des virtualités interprétatives en germe dans les textes originaux.

Pour pousser l'exploration plus avant d'un point de vue traductologique, il faudra cependant étudier la part manquante de ce corpus, ici restée dans l'ombre, à savoir les 16 traductions produites en amont – et il conviendra d'interroger les 16 traducteurs et traductrices sur leurs choix, notamment quand ils divergent. Il pourra être intéressant, notamment, de croiser stabilité des lemmes et comparaison entre langues, une instabilité marquée dans une langue pour un lemme globalement stable dans les autres pouvant être l'indice d'une difficulté traductionnelle. L'enquête ne fait donc que commencer...

REMERCIEMENTS

Merci à nos stagiaires Kevin Woringer et Ekatarina Boboshko pour leur aide précieuse dans l'alignement et dans l'étude de ces deux textes. Merci également aux traducteurs et traductrices ayant participé à l'expérience : Sana Abdi, Sara Abdollahzade, Florian Barrière, Marc Béghin, Matteo Capponi, Nicolò Cecchella, Cha Hyang-mi, Christophe Cusset, Francesca Dell'Oro, Dominique Dias, Jalel El Gharbi, Sadok Gassouma, Myriam Geiser, Alireza Ghafouri, Laetitia

Gonon, Charles Guittard, Houssam Jalal, Antje Marianne Kolde, Kohei Kuwada, Thierry Laplanche, Lee So-yeong, Lee Tae-yeon, Jacques Lévy, Aline Marchand, Thierry Maré, Katja Meister, Moon So-young, Emanuela Nanni, Hessam Noghrehchi, Midori Ogawa, Erika Padova, Claire Pellissier, Natacha Rimasson, Ferdinand Schlie, Pierre Siegenthaler, Nina Soleymani Majd, Olivier Thévenaz, Marie-Line Zucchiatti.

NOTES

1. MONTALE, Eugenio (1999): *Poesia travestita*. Novara: Intralinea.
2. TWAIN, Mark (1903): *The Jumping Frog: In English, Then in French, and Then The Clawed Back into a Civilized Language Once More by Patient, Unremunerated Toil*. New York/Londres: Harper & Brothers.
3. Intitulé «Épreuves de l'étranger», ce projet, impliquant des chercheurs de trois laboratoires de l'Université Grenoble Alpes (ILCEA4, LIDILEM, Litt&Arts), coporté par Emanuela Nanni, Filippo Fonio et Pascale Roux, s'inscrit dans le cadre du programme «Poétique et stylistique du texte traduit» de l'UMR Litt&Arts.
4. L'essai et les quatre poèmes, inclus dans un ensemble plus vaste, ont ensuite été publiés (Macé 2018). MACÉ, Gérard (2018): Images de la caverne. Proche Afrique. In: Ridha BOULAËBI et Claude COSTE, dir. *Les Mondes de Gérard Macé*. Mazères: Le Temps qu'il fait, 15-25, 109-110.
5. Ce sont pour la plupart des professeurs et professeures ou universitaires spécialistes dans le domaine des langues. Quelques-uns sont traducteurs ou traductrices à plein temps. Quelques rétroversions résultent d'une cotraduction. Aucune version traduite ou rétrotraduite n'a été produite, à notre connaissance, en contexte pédagogique. La liste des traducteurs et traductrices est consultable à l'adresse suivante. FONIO, Filippo, NANNI, Emanuela et ROUX, Pascale (2018): Le projet «Épreuves de l'étranger». Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/418-le-projet-epreuves-de-l-etranger>>.
6. L'ensemble du corpus est disponible sur le site suivant. ROUX, Pascale, dir. (2017): Épreuves de l'étranger. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/224-l-epreuve-de-l-etranger>>. Une édition alignée est en préparation sur Roux, Pascale, dir. (s. d.): Épreuves de l'étranger. Université Grenoble Alpes. Consulté le 15 avril 2022, <<http://epreuves-etranger.elan-numerique.fr/>>. Il existe également une édition papier des quatre poèmes et de leurs traductions (Macé 2017a). MACÉ, Gérard (2017a): *Chenilles et papillons*. Droue-sur-Drouette: La Pionnière.
7. ROUX, Pascale (2018): L'analyse stylistique du corpus en français: cadres de l'étude et hypothèses. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/420-l-analyse-stylistique-du-corpus-en-francais-cadres-de-l-etude-et-hypotheses>>.
8. MACÉ, Gérard (2017b): *Chenilles (& papillons)*: Poèmes. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/333-chenilles-amp-papillons>>.
9. DELL'ORO, Francesca (2017): Traduire Gérard Macé: Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/301-traduire-gerard-mace-poesie>>.
10. Le critère de découpage est essentiellement sémantique, mais peut être variable en fonction des rétrotraductions, l'unité retenue apparaissant intuitivement comme une unité de sens relativement stable au niveau de la traduction. Ainsi, on a des segments assez simples tels que DET + NOM, PRON + VERB, ou VERB + PREP, mais parfois l'annotateur a jugé plus cohérent de ne pas segmenter des expressions comme «la goutte de sang» ou «pour marcher devant nous», dans la mesure où les rétrotraductions s'ajustent mieux avec ces syntagmes élargis.
11. Ce coefficient permet de mesurer à quel point les deux séries sont linéairement corrélées – en prenant la première série en abscisse et la deuxième en ordonnée, les points ainsi formés se rapprochent d'une droite lorsque le coefficient est proche de 1 ou -1.
12. On compare les segments des lemmes et non des formes. Par exemple, dans la comparaison de «s'ouvrirait» avec «pour s'ouvrir», on a les deux ensembles de lemmes («se», «ouvrir») et («pour», «se», «ouvrir»), avec une intersection de cardinal 2. Le Dice est donc égal à: $2 * 2 / (2+3)$.
13. MACÉ, Gérard (2017c): Proche Afrique: Prose. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/387-proche-afrique-prose>>.
14. GEISER, Myriam, BÉGHIN, Marc (2017): Traduire Gérard Macé: Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/388-traduire-gerard-mace-prose>>.

15. DELL'ORO, Francesca (2017) : Traduire Gérard Macé : Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/395-traduire-gerard-mace-prose>>.
16. GASSOUMA, Sadok (2017) : Traduire Gérard Macé : Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/395-traduire-gerard-mace-prose>>.
17. CUSSET, Christophe (2017) : Traduire Gérard Macé : Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/395-traduire-gerard-mace-prose>>.
18. GASSOUMA, Sadok (2017) : Traduire Gérard Macé : Poésie. Acta Litt&Arts. Consulté le 15 avril 2022, <<http://ouvroir-litt-arts.univ-grenoble-alpes.fr/revues/actalittarts/280-traduire-gerard-mace-poesie>>.

RÉFÉRENCES

- BAKER, Mona (1993) : Corpus Linguistics and Translation Studies: Implications and Applications. In: Mona BAKER, Gill FRANCIS et Elena TOGNINI-BONELLI, eds. *Text And Technology, In Honour of John Sinclair*. Amsterdam/Philadelphie: John Benjamins, 233-249.
- BAKER, Mona (1996) : Corpus-based Translation Studies: the challenges that lie ahead. In: Harold SOMERS, ed. *Terminology, LSP and Translation. Studies in Language Engineering: In Honour of Juan C. Sager*. Amsterdam/Philadelphie: John Benjamins, 175-186.
- BERMAN, Antoine (1995) : *Pour une critique des traductions: John Donne*. Paris: Gallimard.
- BERMAN, Antoine (1999) : *La traduction et la lettre ou l'auberge du lointain*. Paris: Seuil.
- BOBOSHKO, Ekaterina (2019) : *Analyse textométrique sur les rétrotraductions du projet « Épreuves de l'étranger »*. Mémoire de Master 2 Mention Sciences du langage. Grenoble: Université Grenoble Alpes.
- CAMOIN, Ariane (2019) : Le chirurgien-dentiste face au refus de soins chez l'enfant avec troubles psychiques ou cognitifs. Thèse de doctorat. Aix-Marseille: Université.
- GALE, William A. et CHURCH, Kenneth W. (1993) : A Program for Aligning Sentences in Bilingual Corpora. *Computational Linguistics*. 19(1):75-102.
- ISABELLE, Pierre (1992) : La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie. *Meta*. 37(4):721-731.
- KAY, Martin et RÖSCHEISEN, Martin (1993) : Text-Translation Alignment. *Computational Linguistics*. 19(1):121-142.
- KRAIF, Olivier (2002) : Translation alignment and lexical correspondences: a methodological reflection. In: Bengt ALTENBERG et Sylviane GRANGER, eds. *Lexis in Contrast*. Amsterdam: John Benjamins, 271-290.
- LAMRAOUI, Fethi et LANGLAIS, Philippe (2013) : Yet Another Fast, Robust and Open Source Sentence Aligner. Time to Reconsider Sentence Alignment? In: Andy WAY, Khalil SIMA'AN et Mikel L. FORCADA, dir. *Proceedings of the XIV Machine Translation Summit*. Machine Translation Summit, Nice, 2-6 septembre 2013, 77-84. Disponible sur : <<https://aclanthology.org/2013.mtsummit-papers.10.pdf>>.
- LAPLACE, Colette (1994) : *Théorie du langage et théorie de la traduction*. Paris: Didier Érudition.
- LAVIOSA, Sara (2002) : *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.
- MASSOUBRE, Catherine, LANG, François, JAEGER, Burkard, JULLIEN, Michel et PELLET, Jacques (2002) : La traduction des questionnaires et des tests : techniques et problèmes. *La Revue canadienne de psychiatrie/The Canadian Journal of Psychiatry*. 47(1):63-69.
- SELESKOVITCH, Danica (1968) : *L'interprète dans les conférences internationales: Problèmes de langage et de communication*. Paris: Classiques Garnier.
- SELESKOVITCH, Danica (1975) : *Langage, langues et mémoire, étude de la prise de notes en interprétation consécutive*. Paris: Minard.
- SCHMID, Helmut (1999) : Improvements in Part-of-Speech Tagging with an Application to German. In: Susan ARMSTRONG, Kenneth CHURCH, Pierre ISABELLE, Sandra MANZI,

Evelyne TZOUKERMANN et David YAROWSKY, dir. *Natural Language Processing Using Very Large Corpora*. Dordrecht: Springer.

ANNEXES

Textes originaux de Gérard Macé

Poèmes

En sortant de la caverne, nous verrons les mirages
qu'on rêve depuis toujours de traverser à pied sec:
des flaques d'eau qui s'évaporent quand on avance,
des nappes de lumière dans lesquelles on voit des lacs.

Aucun prophète, aucun pharaon pour marcher devant nous.
Pas de mer rouge qui s'ouvrirait d'elle-même
comme les grandes lèvres des femmes, mais un au-delà
qui recule en même temps que l'horizon.

*

Avec la neige la nappe est mise
sur les prés. Le ciel tout entier
pourrait tenir dans une cuiller en argent,
le ciel et deux ou trois soleils
sur les quatre que comptaient les Aztèques.

Le nôtre est un soleil d'hiver
qui éclaire le lieu du crime:
la goutte de sang, les plumes d'oiseau,
était-ce le festin des dieux
ou le repas du serpent?

*

Tant de noms sous la neige
qui attendent les beaux jours
pour briller comme de l'or.

Autant de morts à la guerre
qui ont vu dans leurs yeux, leurs yeux hagards,
la campagne en fleurs derrière les barbelés.

*

Les montagnes à contrejour ressemblent à des enclumes,
et le soleil éclaire cette forge abandonnée.

Les étincelles, c'est le marteau du philosophe
qui les fait jaillir en tapant sur le vieux monde.

Un monde où l'on ferrait les chevaux
dans l'odeur de la corne, des chevaux battus
qui regardaient droit devant eux, à cause des œillères
que n'osaient pas porter les hommes.

Essai

Proche Afrique

Au cœur des ténèbres, selon Conrad.
Continent noir, pour la plupart d'entre nous.

L'Afrique est réputée sombre et mystérieuse après avoir été impénétrable, et reste à part dans l'histoire des civilisations, quand on ne l'oublie pas tout simplement. À vrai dire elle sert de repoussoir avec ses maladies, ses catastrophes, ses coutumes apparemment étranges, et comme le Moyen Âge elle est réputée obscure alors que la réalité est lumineuse et colorée, autant que le Moyen Âge était vif et polychrome.

Pour aggraver son cas, l'Afrique est réputée sans histoire parce que celle-ci n'est pas écrite. C'est pourquoi je suis allé d'abord en Éthiopie, où l'écriture est connue depuis l'Antiquité, sans doute venue de la péninsule arabique à travers la mer Rouge. L'Éthiopie où le christianisme adopté au IV^e siècle ne doit rien à Rome, et fidèle à l'Ancien Testament n'a rien renié du judaïsme. Son histoire parfaitement autonome s'est prolongée jusqu'à nos jours, en lien avec Alexandrie, grâce à des manuscrits enluminés et des peintures d'église qui réjouissent l'œil autant que l'esprit. Des manuscrits en guèze, langue morte devenue langue rituelle comme le latin en Occident.

Mais alors une question se pose : pourquoi l'écriture, à partir de l'Égypte et de l'Éthiopie, ne s'est-elle pas diffusée dans l'Afrique subsaharienne ?

On pense parfois, un peu trop vite, que l'écriture était inconnue dans ces contrées à l'écart des grandes routes et des échanges, dans ce continent sous-peuplé que les grandes inventions n'atteignaient pas. C'est oublier, en dehors même des migrations, que les rives du Sahel étaient en contact avec le monde arabe, et que les caravanes qui échangeaient les marchandises contre des esclaves, bien avant la traite négrière, apportaient avec elles le Coran. C'est oublier aussi que par l'Océan indien, d'où venait la monnaie sous forme de coquillages (les fameux *cauris*), l'Asie elle-même exerçait une influence, parfois de façon inattendue. En 1730, un voyageur du nom de Snelgrave raconte que dans la bande sahélienne il rencontre des Malais, qui écrivent devant les Africains, et même à leur demande. De ces manuscrits les indigènes font des amulettes, qu'ils portent sur eux comme des porte-bonheurs, ou comme des marques de pouvoir. Et Mungo Park à la fin du même siècle assiste à une scène où l'écriture diluée devient un breuvage magique. Cent ans plus tard, à l'ouest du Cameroun, le sultan Njoya fera boire à ses sujets les caractères à peine effacés, dispersés dans l'eau, de l'écriture qu'il vient d'inventer. Ces deux épisodes si semblables, alors qu'ils sont séparés dans l'espace et dans le temps, témoignent d'un rapport magique à l'écriture, que nous avons connu nous aussi. Sans parler du latin d'église ou du grec des apothicaires, il suffit de se rappeler les interprétations farfelues à propos des hiéroglyphes, caractères sacrés traduisant les décrets du ciel, ou les révélations d'une religion perdue, jusqu'à ce que Champollion vienne dissiper ces faux mystères.

L'absence d'écriture, si elle ne relève pas tout à fait d'un choix, est un phénomène dont la complexité devrait éveiller notre intérêt, car il concerne l'humanité tout entière. Qu'il suffise de rappeler que le sanscrit dans l'Antiquité, ou le japonais jusqu'au VIII^e siècle de notre ère, furent des langues non écrites. Et que les langues sans écriture, aujourd'hui encore, sont les plus nombreuses. L'adoption de l'écriture, dont il est difficile de rendre compte, est donc un fait anthropologique essentiel, qui nous renseigne sur notre propre histoire, à condition de ne pas se complaire dans l'ethnocentrisme, et d'abandonner une vision linéaire d'on ne sait quel progrès. D'autant que nous vivons plusieurs âges à la fois : Barthes disait très justement que nombre d'entre nous ont encore une mentalité « pré-voltairienne ».

Le rapport magique à l'écriture subsiste, et la crainte de sa disparition nous hante. En Afrique où l'enseignement a changé le rapport aux signes, il reste quelque chose de l'initiation et du secret, qui ne nous sont pas tout à fait étrangers. Si l'on privilégie la parole, c'est

pour choisir celui à qui l'on s'adresse, alors qu'un document écrit peut tomber entre toutes les mains. Avec la parole on peut rester entre soi, avec l'écriture on porte la loi au loin et l'on recule les frontières de l'empire, mais on peut aussi précipiter sa chute.

Les résistants, les clandestins, les êtres pourchassés prenaient la précaution d'avaler les documents qui auraient pu les compromettre. Ce qu'on sait par cœur est invisible, inviolable, et combien de prisonniers, de poètes ont confié à leur mémoire (ou à celle de leurs proches) ce qui devait les faire vivre ou survivre.

Le hasard, qu'on croirait dirigé par une main bienveillante, vient de mettre sous mes yeux la fable suivante, de Pierre Bettencourt :

« J'avais mis dans une bouteille d'eau les dernières paroles de ma femme. Les lettres noires se diluèrent bientôt, et l'eau devint couleur de vase. Quelques années plus tard, la retrouvant dans un placard, l'eau s'était évaporée, et dans le fond de la bouteille les lettres avaient repris leur forme. »

Bettencourt n'avait pas entendu parler du sultan Njoya, qui n'avait pas prévu l'existence de Bettencourt, mais c'est aussi grâce à cet imaginaire partagé qu'on peut approcher l'Afrique, autant que par le voyage.