



HAL
open science

Décoder le texte : le développement d'un comparateur numérique des traductions d'oeuvres littéraires

Amélie Derome

► **To cite this version:**

Amélie Derome. Décoder le texte : le développement d'un comparateur numérique des traductions d'oeuvres littéraires. 2018. hal-01892284

HAL Id: hal-01892284

<https://hal.science/hal-01892284>

Submitted on 10 Oct 2018

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

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

Décoder le texte : le développement d'un comparateur numérique des traductions d'œuvres littéraires

Amélie Derome, Aix Marseille Univ, LERMA, Aix-en-Provence, France

Borges disait des traductions innombrables de l'*Odyssee* qu'elles formaient une bibliothèque de la littérature mondiale¹. Voici le précipice qu'ouvre l'idée vertigineuse de l'écrivain argentin : les traductions d'une seule œuvre, à la condition toutefois qu'elle figure au panthéon des belles lettres, peuvent former un corpus au sein duquel les infinies variations linguistiques et culturelles créent une richesse similaire à celle d'une bibliothèque classique, dont les rayonnages accueillent les ouvrages de différents auteurs. Borges réaffirme ainsi la primauté des traductions : loin de l'image de texte dérivé qui leur est souvent assignée, elles seraient, en fin de compte, des œuvres littéraires comme les autres.

Ce propos paraît trouver un écho chez les chercheurs contemporains qui s'intéressent à la question de la retraduction, que j'entends ici comme la traduction nouvelle d'une œuvre déjà traduite. L'un d'entre eux, Matthew Reynolds², bat en brèche la pensée commune selon laquelle les traductions nouvelles vaudraient mieux que leurs aînées. Cette vision téléologique de la retraduction repose avant tout sur la tentation, pour les acteurs du monde du livre, de proposer une offre économique neuve – et parfois artificielle – des textes classiques. L'appareil critique des livres retraduits constitue ainsi, habituellement, une entreprise de démolition en bonne et due forme des traductions antérieures. L'idée selon laquelle la qualité des traductions croîtrait en fonction du nombre des années repose avant tout sur l'opportunisme commercial : son socle théorique paraît bien faible en comparaison. À la suite de Reynolds, il me semble en effet que les différentes traductions d'une œuvre ne devraient pas se supplanter, mais au contraire coexister, au sein de la bibliothèque qu'imaginait Borges. Cette bibliothèque est au cœur de cet article : je souhaiterais en effet la faire vivre en élaborant un comparateur numérique de traductions, c'est-à-dire un espace où les retraductions pourraient coexister.

Cet article retracera brièvement la genèse du projet avant de détailler ses sources d'inspirations et de proposer une première solution de réalisation possible.

Genèse et objectifs du comparateur de traductions

Ma thèse est consacrée aux traductions françaises de *Gulliver's Travels* de Jonathan Swift. L'enjeu majeur de cette étude se trouve dans la question de savoir si – et comment – les 14 traductions françaises intégrales de l'œuvre destinées à un public adulte ont influencé la réception du texte du doyen de Saint Patrick de ce côté-ci de la Manche. Cette interrogation m'amène naturellement à comparer, entre elles et au texte original, ces 14 traductions – travail aussi laborieux que minutieux. J'ai pensé qu'un comparateur numérique me fournirait une aide précieuse pour mener à bien cette tâche indispensable à ma thèse.

Bien que formée à la création de sites internet, je demeure une jeune chercheuse littéraire, plus familière de l'énonciation discursive que de la formulation algorithmique. Il me fallait donc, afin de poursuivre ce projet, m'associer à un développeur. J'ai donc soumis mon idée à un ingénieur, dont les conseils m'accompagnent depuis. Diplômé de École nationale de la statistique et de l'administration économique, il travaille comme ingénieur en *machine learning*

¹ Jorge Luis Borges, « Les Traductions d'Homère » in *Œuvres I* (Paris : Gallimard, Pléiade, 1993), p. 291.

² <https://www.creativeml.ox.ac.uk/research/prismatic-translation>

appliqué à la cryptosécurité tout en poursuivant un master de mathématiques fondamentales à Paris VII.

Voici l'idée que je lui ai soumise : il s'agirait de développer un algorithme capable d'identifier les différences entre les traductions d'une même œuvre. Trois catégories de transformation m'ont paru décisives. Les coupes, qui désigneraient les passages censurés, les ajouts, qui indiqueraient les interpolations de la main du traducteur – pratique courante au XVIII^e siècle – et les remplacements, qui décèleraient, d'une traduction à l'autre, les effets de synonymie et de fluctuation de registre. Ces opérations correspondent à certaines des tendances déformantes de la traduction identifiées par Antoine Berman³. Les coupes renvoient à l'appauvrissement quantitatif et les ajouts à l'allongement, tandis que les remplacements recouperaient les tendances de l'ennoblissement, de la vulgarisation, de la destruction des réseaux de signifiants sous-jacents, des réseaux vernaculaires, des locutions et idiotismes. Cependant, je ne les ai pas seulement retenus en raison de leur accointance avec les tendances déformantes de Berman : j'ai en effet appris, grâce à l'ingénieur qui a bien voulu me prêter main-forte, que les algorithmes d'alignement de texte, formules capables de comparer deux textes, reposaient précisément sur les opérations que sont l'insertion, la suppression et le remplacement. Ces trois critères, du point de vue de la traductologie comme de l'ingénierie, constitueraient donc une grille d'analyse efficace des 14 traductions de mon corpus.

Il m'a toutefois semblé, pour deux raisons, insuffisant de réserver ce projet à mon usage particulier. D'une part, et du point de vue universitaire, les éventuels résultats obtenus me paraissent pouvoir concerner l'ensemble des chercheurs en traductologie. D'autre part, l'inspiration borgésienne de la bibliothèque implique la diffusion des traductions. Je souhaite ainsi rendre le programme disponible en open source sur une plateforme communautaire, afin que les chercheurs, éditeurs et particuliers puissent s'en saisir et comparer d'autres œuvres. Je diffuserai, quant à moi, la comparaison des traductions françaises libres de droit de *Gulliver's Travels* sur un site internet sous licence *creative commons*, soumis aux conditions d'attribution et de non-utilisation commerciale. Les portraits des traducteurs des versions libres de droit, les entretiens que j'ai menés avec les traducteurs vivants de Gulliver, une chronologie de l'histoire des traductions de l'œuvre en France, ainsi que des éléments clés concernant la réception de ses traductions figureront également sur le site.

Les comparateurs des traductions bibliques

Après avoir formulé les objectifs du projet, la première étape de son développement exigeait que je me mette en quête de programmes similaires. Trois types se sont distingués lors de mes recherches : les éditions de comparaisons de traduction digitales, les comparateurs unilingues de textes, et les programmes de *natural language processing* (NLP) appliqués à la traduction. Chacun m'a paru source d'inspiration féconde, mais aucun ne m'a semblé suffisant, pour des raisons que j'évoquerai dans l'analyse de chacun d'entre eux.

Je me suis d'abord intéressée aux sites internet de comparaison des traductions de la Bible⁴. Ils présentent l'intérêt d'afficher, sur une même page, deux traductions différentes d'un même texte. Grâce au découpage en versets, le lecteur jouit également de la possibilité d'afficher, pour un seul verset, l'ensemble des traductions disponibles. La présentation de ces sites me semble agréable et fluide. J'aimerais ainsi que le site internet de mon projet réponde à une disposition

³ Antoine Berman, *La Traduction et la lettre ou l'auberge du lointain* (Paris : Seuil, 1999), pp. 52-68.

⁴ <http://www.lueur.org/bible/comparer/>

similaire. Cependant, il ne s'agit là que d'affichage, les comparateurs bibliques se contentent de montrer plusieurs traductions : ils ne comprennent pas d'analyse des données, élément central de mon projet. J'ai pourtant pensé, dans un premier temps, imiter leur fonctionnement. Il me semblait pertinent de découper le texte original et les traductions en unités de sens, que l'algorithme pourrait traiter sous la forme de *strings*, c'est-à-dire de segments. Cette éventualité, quoique tout à fait réalisable, m'a par la suite semblé faillible : elle impliquerait un labeur tout aussi laborieux que celui de comparer les traductions manuellement.

Un comparateur unilingue : MEDITE

Je me suis donc penchée sur les comparateurs unilingues de textes. L'un d'entre eux, le logiciel MEDITE⁵, développé par l'Institut des Textes et des Manuscrits Modernes de l'ENS et le Laboratoire d'informatique de Paris VI, m'a particulièrement enthousiasmée. Ce programme, conçu pour répondre aux enjeux de la génétique textuelle, donne l'opportunité aux chercheurs de comparer, avec une aisance accrue, les différentes versions d'un même texte – il a notamment été utilisé pour une nouvelle édition de l'œuvre de Charles Ferdinand Ramuz. L'algorithme au cœur de ce logiciel s'inspire de ceux d'alignement de texte que j'évoquais tout à l'heure, et s'inscrit dans le *natural language processing*, c'est-à-dire des programmes informatiques qui traitent le langage humain. Il introduit toutefois une innovation majeure, comme l'expliquent Ganascia, l'un des coordinateurs du projet. L'algorithme sollicite un quatrième critère, celui du déplacement, qui coordonne les opérations classiques de l'insertion, de la suppression et du remplacement. Voici la manière dont il procède :

The first step of this algorithm detects invariant character blocks (i.e. the “fragments”). Then a second step distinguishes, among the invariant character blocks, those which are displaced, i.e. moved. Lastly, a third step identifies the deleted, inserted and replaced blocks that are between the unmoved invariant blocks. Note that, using this algorithm, a fragment of text may appear to be both displaced and inserted or deleted⁶.

Ce logiciel est ainsi capable, non seulement de relever les coupes, interpolations et remplacements, mais également les déplacements – et donc les remaniements syntaxiques. J'ai voulu le mettre à l'épreuve des traductions françaises de *Gulliver's Travels*, afin d'évaluer son efficacité. Le logiciel tend à repérer des déplacements qui n'en sont pas réellement, notamment pour les auxiliaires et les pronoms et confond parfois les remplacements et les déplacements. Il permet toutefois d'identifier d'un seul coup d'œil les interpolations des traducteurs et les passages censurés. MEDITE souligne ainsi des transformations intéressantes d'une traduction à l'autre. Le logiciel pâtit néanmoins d'un inconvénient majeur dans le cadre de mon projet dans la mesure où il demeure unilingue.

Une troisième voie : l'intégration d'un algorithme de traduction automatique

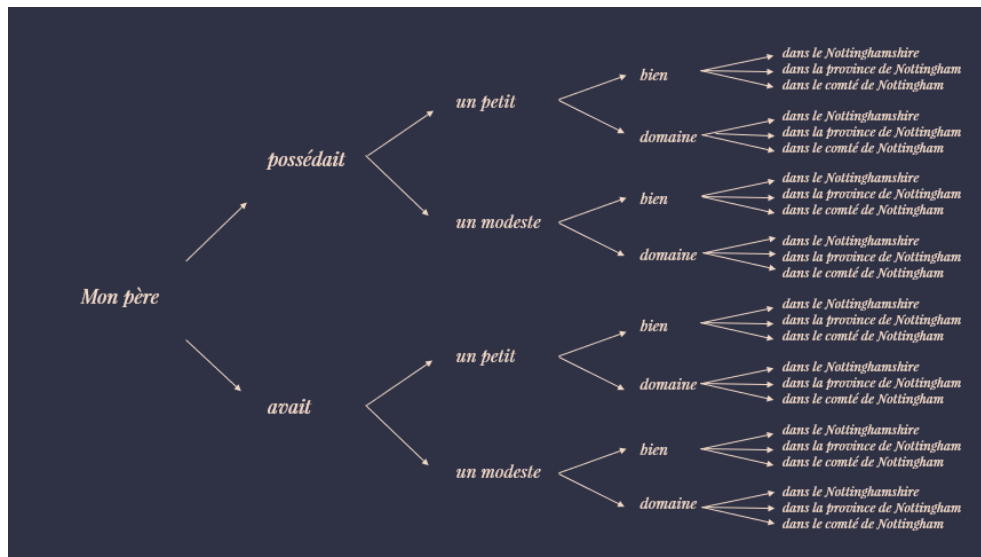
J'ai donc pensé qu'il serait intéressant de le concaténer à d'autres algorithmes de *Natural Language Processing*, ceux qui se cachent derrière les logiciels de traduction automatique. L'ingénieur avec lequel je collabore m'a toutefois expliqué que l'algorithme de MEDITE ne paraissait pas entièrement

⁵ <http://obvil.lip6.fr/medite/>

⁶ Jean-Gabriel Ganascia, « A Unilingual Text Aligner for Humanities Applications to Textual Genetics and to the Edition of Text Variants », 2011.

compatible avec le *machine learning* qui garantit le bon fonctionnement des programmes de traduction. Ces derniers comparent de vastes corpus de textes, traités à la manière de bases de données où ils piochent les meilleures solutions de traduction possibles d'une unité lexicale, potentiellement polysémique, en fonction du contexte. J'aimerais, ici, illustrer cette façon de procéder par un exemple inspiré du cours de version de M. Jean Viviès que j'ai suivi il y a quelques années. Il évoquait les multiples sens du nom commun bureau : tantôt table, pièce, ou environnement de travail. Le programme de traduction automatique, pour traduire ce terme, consulte les probabilités du sens des occurrences de « bureau » en fonction du contexte où il apparaît. Ainsi, le programme « apprend » que bureau, proche du verbe « poser », par exemple dans la phrase « je posais mon stylo sur le bureau », renvoie probablement au meuble et traduit par « desk ». Si ce procédé optimise la traduction automatique, il ne saurait pourtant s'appliquer à la comparaison des traductions.

En effet, pour comparer les traductions possibles, un tel algorithme concevrait la traduction de *Gulliver's Travels* comme une arborescence. La figure suivante illustre la manière dont l'algorithme traiterai le premier segment de la première phrase de l'œuvre : my Father had a small estate in Nottinghamshire.



Au fil du texte, cette arborescence serait amenée à se déployer exponentiellement, ce qui priverait le programme de fonctionner et le condamnerait, *in fine*, à tourner en boucle. Il faut, pour remédier à cela, donner au programme une voie déterminée qu'il puisse suivre. On peut en effet comparer l'arborescence à un itinéraire en voiture. Si l'on ne donne pas un trajet type au programme, ce dernier ne peut traiter les données qu'on lui fournit. Afin de comparer les différentes routes possibles qui relient Aix-en-Provence à Paris, par exemple, il doit connaître une trajectoire idéale – qui pourrait être la plus rapide ou bien la plus courte – à laquelle les confronter. La nécessité d'une traduction de référence s'est ainsi imposée dans le cadre de mon projet.

Deux possibilités s'ouvrent alors : choisir celle qui me paraît la plus fidèle ou bien en proposer une nouvelle. La question d'une traduction type introduit toutefois un biais dans la comparaison puisqu'elle devient le mètre-étalon à partir duquel le programme apprécie les autres versions. Il faudrait donc que la traduction choisie comporte le moins d'interpolations, de censures, d'aménagements syntaxiques et de fluctuations de registre possibles – en somme, qu'elle corresponde aux normes traductives actuelles. J'envisage également de faire des

traductions infidèles les mètres-étalon des autres traductions afin de percevoir les variations des résultats qui en découleraient.

Quelle que soit la méthode privilégiée, la mise en ligne du comparateur sur un site, qui affichera aussi bien les traductions françaises que les différentes éditions originales de l'œuvre, constituera bien ce lieu où les versions d'un texte peuvent cohabiter, rappelant au lecteur que « l'idée de 'texte définitif' ne relève que de la religion ou de la fatigue⁷ ».

Bibliographie

BERMAN, Antoine, *La Traduction et la lettre ou l'auberge du lointain* (Paris : Seuil, 1999).

BORGES, Jorge Luis, « Les Traductions d'Homère » in *Œuvres I* (Paris : Gallimard, Pléiade, 1993).

GANASCIA, Jean-Gabriel « A Unilingual Text Aligner for Humanities Applications to Textual Genetics and to the Edition of Text Variants », 2011.

Lueur. Association Lueur, Angers, 2000, <http://www.lueur.org/bible/comparer/>, accessed May 2018.

Prismatic Translation. Creative Multilingualism, Oxford University, Oxford, 2016, <https://www.creativeml.ox.ac.uk/research/prismatic-translation>, accessed May 2018.

MEDITE. Item, Ens Ulm et LIP6, Paris 6, Paris, 2015, <http://obvil.lip6.fr/medite/>, accessed May 2018.

⁷ Borges, *op. cit.*, p. 291.