

MARIANNE REBOUL

marianne.reboul@ens-lyon.fr

SITUATION ACTUELLE

2019-2020 Maître de Conférences en Humanités Numériques (Section 8 – Lettres Classiques)

T H E M E S D E R E C H E R C H E

Littérature française et comparée, Langues anciennes, humanités numériques, traduction (automatique et manuelle)

F O R M A T I O N

2018 **Qualification en sections 8 (18208314413), 9 (18209314413) et 10 (18210314413)**

2012 – 2017 **Paris 4 Sorbonne, Labex OBVIL**

Doctorat de littérature comparée, « Comparaison semi-automatique des traductions en langue française de l'*Odyssée* d'Homère (1547-1955) », sous la direction de Jean-Yves Masson, soutenu le 5 octobre 2017. Mention très honorable avec félicitations du jury.

Jury composé de :

Gregory Crane (Université de Leipzig, Université de Boston)

Paul Demont (président du jury, Université Paris 4 Sorbonne)

Alexandre Gefen (Université Paris 3 Sorbonne Nouvelle, CNRS)

Emiliano Giovannetti (CNR Pise)

Françoise Lavocat (Université Paris 3 Sorbonne Nouvelle)

Jean-Yves Masson (Université Paris 4 Sorbonne)

Cette thèse a été présélectionnée pour le prix SHS 2018, décerné par l'Université PSL.

2011 – 2012 **Paris 4 Sorbonne**

Master 2 de Lettres Classiques : « Étude numérique du chant IX de l'*Odyssée* d'Homère », sous la direction de Paul Demont. Mention très bien.

2010 **Agrégation de Lettres classiques, rang : 35**

2007 – 2008 **Paris 4 Sorbonne**

Master 1 de Lettres Classiques : « Structuration romanesque dans *Daphnis et Chloé* », sous la direction d'Alain Billault. Mention très bien.

2003 – 2006 **Lycée Henri IV, Paris**

Classe préparatoire A/L, option Lettres classiques. Admissibilité au concours d'entrée à l'ENS.

2003 **International Baccalaureate O.I.B., Classics, Georges Duby High School. Mention très bien. Proficiency of Cambridge. Mention A.**

Bourse d'études en Lettres Classiques post-bac proposée par l'Université d'Oxford.

LANGUES

Anglais (bilingue), italien (bilingue), espagnol (lu et parlé), latin et grec ancien, connaissances de base en allemand.
Langages informatiques : Java, JavaScript, Python, PHP, XML/TEI, HTML.

ENSEIGNEMENT

2018 – 2020 **Maître de Conférences en Humanités Numériques à l'École Normale Supérieure de Lyon, UMR 5317 IHRIM**

Les cours sont annuels à raison de 8h par semaine au moins.

« Antiquités Numériques », ENS de Lyon, M2 en Lettres Classiques.

Approche numérique pour les ressources de l'Antiquité, permettant de faire certaines choses plus rapidement, de découvrir des phénomènes qui pourraient corroborer ou invalider des hypothèses de travail, d'ouvrir de nouvelles pistes de réflexion, tant pratiques que théoriques, de réfléchir, finalement à ce que sont les Lettres Classiques, et leur rapport avec la logique du numérique. Nous proposons une connaissance théorique ; l'accès à des outils existants, qui vous facilitent la vie des latinistes et hellénistes, mais qui aussi feront peut-être percevoir un objet d'étude sous un jour nouveau ; la possibilité, à terme, de créer ses propres outils, et aboutir à de nouvelles hypothèses de travail. Outre les bases textuelles libres de droit ou privées, nous analysons les textes avec différents logiciels, dont certains écrits spécifiquement pour ce cours, en Python.

« M2 - S2 Projet personnel en environnement numérique », Sorbonne, Paris-Nanterre, M2 en Lettres Classiques.

Aide individualisée au développement de projets numériques pour étudiants en Lettres Classiques. Aide à la programmation, à l'orientation pour les outils disponibles en Lettres Classiques, à l'élaboration de travaux possibles en fonction des aptitudes et connaissances de chacun (depuis l'analyse automatique à certaines techniques d'Intelligence Artificielle en Lettres Classiques).

« Analyse textuelle automatique, ressources et outils », ENS de Lyon, M2 en Sciences Humaines.

Le cours permet d'apprendre à récupérer des textes (ressources en ligne, numérisation de documents), à les préparer (formats de fichiers, encodage des caractères, notion de balisage) et à utiliser des outils d'analyse (index lexicaux, concordances, cooccurrences) et de recherche (expressions régulières, lemmatisation). Le cours s'appuie notamment sur des carnets Python. Il inclut également une présentation et utilisation de la base Frantext.

« Édition numérique de textes: initiation à la TEI », ENS de Lyon, M2 en Sciences Humaines.

Nous voyons à quoi sert le stylage dans les logiciels d'aide à la publication comme Word, ce qu'est un métalangage et à quoi il sert. Nous prenons l'exemple d'un corpus de romans des XIXe et XXe siècles, pour expliquer quelles sont les exigences des normes TEI (Text Encoding Initiative) et leur utilité pour la diffusion et la fouille de corpus. Utilisation du logiciel Oxygen pour l'édition XML.

« Logiciel bibliographique et documentaire », ENS de Lyon, M1 et pré-Master en Sciences Humaines.

Apprendre à constituer une base bibliographique et à l'utiliser dans un traitement de texte; connaître les différents acteurs et sites institutionnels de diffusion du savoir académique; construire des stratégies de recherche et de veille documentaire.

2016 – 2018 **Post-doctorat pour l'ANR Chapitres, UMR Thalim, Paris 3 Sorbonne Nouvelle**
Les cours sont annuels à raison de 2h par semaine.

« La traduction automatique neuronale dans la pratique », PSL, Ecole des Chartes, M2 Humanités Numériques. Depuis 2018, 12h.

Le cours a pour but de mettre en pratique les principes de la traduction automatique neuronale sur un corpus parallèle donné. Après une explication des principes de la traduction automatique, des réseaux de neurones, des « word embeddings », de l'alignement automatique multilingue et de la traduction automatique neuronale, nous procédons aux installations de base nécessaires à la mise en place d'un réseau neuronale pour la traduction automatique : nous choisissons de mettre en place la machine de Google, « nmt » via « tensorflow » (avec le recours aux outils de « seq2seq » et du « Mosesdecoder »). Une fois le réseau de traduction monté, nous faisons un bilan de ce qui peut engendrer les erreurs, ce qui peut être amélioré et ce qui est susceptible de perdurer.

« Initiation au langage Python », M2 et supérieur.

Cours ouvert aux étudiants de Master, mais aussi aux enseignants de l'Université Sorbonne Nouvelle Paris 3 désirant apprendre les rudiments de la programmation orientée objet. Les cours s'adressaient autant à des personnes n'ayant jamais programmé qu'à ceux qui avaient déjà acquis une certaine expérience de la programmation. Le cours s'orientait dans trois directions. Volet 1 : comprendre les bases de la traduction, la nécessité des corpus enrichis ; prendre conscience de ce qu'est une arborescence textuelle, aborder les différents types de langage, à commencer par la structuration XML. Volet 2 : explorer les bases de données et outils préexistants, commencer à utiliser les programmes à usage littéraire comme TXM (voire, pour des cas plus complexes, des logiciels de représentations de données, comme Gephi). Volet 3 : appréhender les lignes de commande de base et apprendre les rudiments utiles pour les études littéraires en Python.

2012 – 2015 **Contrat doctoral avec charge d'enseignement, Université Paris 4 Sorbonne, Labex OBVIL**

« Hôtes et parasites, dans l'*Odysée* d'Homère, le *Tartuffe* de Molière, les *Contes fantastiques* d'Hoffmann et *Beloved* de Toni Morrison », TD (6h/semaine), L1.

Qu'est-ce qu'un (bon) hôte ? Les quatre textes du programme permettaient aux étudiants de s'interroger sur le dévoiement spécifique des structures de l'hospitalité que constitue le parasitisme : qu'advient-il quand un hôte invité dans une structure d'accueil en vient à mettre en danger la maison qui l'accueille ? Quelles relations s'instaurent entre hospitalité et hostilité ? Entre don et dommage ? Pour ce cours, j'étais notamment chargée de l'étude de l'*Odysée* de Jaccottet, puisque mes compétences en grec me permettaient des comparaisons appropriées avec le texte source.

« Commentaire de traduction anglaise, *Beloved* de Toni Morrison », TD (4-5h/semaine), L1.

Ce cours avait deux buts : enseigner l'anglais et sa traduction aux étudiants, et leur faire découvrir la pratique du commentaire de traduction. La langue traitée était d'un accès complexe et nécessitait une connaissance bien établie de l'anglais. Cet exercice nécessitait aussi d'apprendre à maîtriser le vocabulaire particulier de la traductologie, et de l'intégrer dans un cheminement analytique. Chaque semaine les étudiants étaient invités à exposer leurs réflexions sur un passage donné, et devaient en traduire une portion, en commentant leur choix de traduction.

« Initiation à la structuration textuelle, à la TEI, et première approche de la fouille textuelle », TD (2-3h/semaine), M2.

Les étudiants de M2 participant à ce TD avaient intégré un Master d'édition, avec une option majeure en édition numérique. À chaque étudiant était attribué un texte en sortie de reconnaissance automatique de caractères (O.C.R.), et le but de l'exercice était qu'ils soient capables de rendre à la fin du semestre un texte structuré et utilisable pour la fouille de données. L'enseignement avait pour objectif de permettre aux étudiants la maîtrise d'un usage avancé du traitement de texte (maîtrise des styles, expressions régulières, macros), afin de les former aux outils de base des métiers de l'édition. Il s'agissait également de donner aux étudiants un aperçu du processus de numérisation depuis le livre imprimé jusqu'au livre électronique. L'enseignement avait enfin vocation à faire aborder aux étudiants certains langages de programmation, en commençant par des actions simples intégrées aux éditeurs de textes (macros par exemple), parfois plus abstraites (étude des expressions régulières et autres techniques de fouilles plain text). Les étudiants devaient être capables, en fin de TD, de parcourir une arborescence XML et d'en tirer les données (même de façon simple) en utilisant Python.

2011 – 2012 **Professeur de Lettres Classiques au collège Jules Verne, Rueil-Malmaison.**

Cours de grec et de latin au collège Jules Verne, 3^e et 4^e (9h/semaine).

Cours de grec et de latin au collège Jules Verne de Rueil-Malmaison, pour les classes de 4^e et de 3^e, 9h par semaine. Organisation d'un voyage scolaire à Rome avec les classes CHAM (classes à option Conservatoire de Musique). Les cours de 4^e étaient consacrés à l'étude de la République Romaine et de la démocratie athénienne, et les cours de 3^e étaient consacrés en Latin à l'histoire de l'Empire, en Grec à l'étude des grandes figures mythologiques. J'ai développé un programme d'entraînement aux conjugaisons et déclinaisons latines et grecques, sous forme de quiz dynamique (les élèves participaient en se faisant passer quatre claviers infrarouges).

Cours de littérature et langue française au collège Jules Verne, 5^e (9h/semaine).

Programme portant sur le roman et le théâtre, du Moyen-Âge au XX^e siècle. Les élèves étaient notamment invités à participer à la construction d'un site pour la classe, réunissant l'ensemble de leurs exposés et productions théâtrales.

2010 – 2011 **Professeur de Lettres Classiques à la cité scolaire Hélène Boucher.**

Cours de grec et de latin à la Cité Scolaire Hélène Boucher, 3^e (6h/semaine).

Responsable de l'enseignement du latin au collège Hélène Boucher à Paris, pour les classes de 4^e et de 3^e, 6h par semaine. Les cours de 4^e étaient consacrés à l'étude de la République Romaine et de la démocratie athénienne, et les cours de 3^e étaient consacrés en Latin à l'histoire de l'Empire, en Grec à l'étude des grandes figures mythologiques (mais les cours de grec se faisaient sur la base du volontariat, pour les élèves comme pour moi). J'ai entièrement développé un jeu de rôle interactif sur les voyages d'Ulysse (de type jeu de l'oie) en C++ avec système de questions/réponses.

Cours de littérature et langue française à la Cité Scolaire Hélène Boucher, 5^e (9h/semaine).

Programme portant sur le roman et le théâtre, du Moyen-Âge au XX^e siècle. Les élèves étaient notamment invités à participer à la construction d'un site pour la classe, réunissant l'ensemble de leurs exposés et productions théâtrales.

ACTIVITES DE RECHERCHE

ATTENTION, CETTE PARTIE DU CURRICULUM N'EST PAS A JOUR

A. CHAPITRE DANS UN OUVRAGE COLLECTIF

- A1.** Reboul, Marianne. « Apprendre à traduire, de la lettre au style » dans *L'écrivain et son école : je t'aime moi non plus*, édité par Martine Jey, 165–181. Cultures numériques. Paris: Hermann, 2017.

J'ai été amenée, dans la partie consacrée à la formation des traducteurs, à étudier la façon dont l'enseignement du latin et du grec a conditionné le développement du style de certains auteurs-traducteurs de la première moitié du XVIII^e siècle au début du XX^e siècle. Dans cet ouvrage, j'ai eu recours aux outils numériques que j'avais créés pour mesurer l'impact fondamental de l'émergence de la traduction juxtalinéaire dans les écoles du XIX^e siècle sur les écrivains en formation, comme Leconte de Lisle. J'ai donc dressé un panorama de l'enseignement du grec et du latin en France au XIX^e siècle et montré à quel point le changement dans la perception du grec et du latin, après la Terreur, les progrès archéologiques et la place prépondérante et croissante de l'enseignement de la grammaire latine et grecque ont contribué à former le goût et le style littéraire des écrivains alors en formation. Les techniques de repérages auxquelles j'ai eu recours dans cet article peuvent être ré-exploitées dans le cadre du projet que nous proposons.

B. ARTICLES PUBLIES DANS UNE REVUE A COMITE DE LECTURE

- B1.** Marianne Reboul, Yuri Bizzoni, et Mario Angelo Del Grosso. « **Diachronic Trends in Homeric Translations** », *Digital Humanities Quarterly*, 11, n°2, 2017. <http://www.digitalhumanities.org/dhq/vol/11/2/000297/000297.html>.

Dans cet article nous exposons à la fois les premiers résultats de l'aligneur automatique *Odysseus* que j'ai programmé, et les théories algorithmiques qui le sous-tendent. Nous y expliquons donc en détail ce que signifie aligner automatiquement des séquences, et les problèmes qui se posent lorsque les textes à aligner sont de langues différentes. J'ai écrit *Odysseus* en Java, et j'ai implémenté l'algorithme d'alignement automatique global de Needleman-Wunsch, et mis l'ensemble du code produit en ligne libre d'accès. Nous expliquons aussi le principe de la sémantique distributionnelle appliquée au traitement automatique multilingue, ainsi que les problèmes de post-traitement lors de l'alignement automatique mot-à-mot (via M Giza++) : l'idée est tirée des théories de la sémantique distributionnelle selon lesquelles un mot est défini par son contexte, et de fait peut être situé dans un espace sémantique par rapport aux mots qui l'entourent. Il s'agissait donc d'un travail de réflexion en linguistique computationnelle. Nous avons appliqué le principe de la distribution sémantique monolingue à un corpus bilingue de la façon suivante : après un premier alignement et la création d'un corpus parallèle, nous avons supposé que lorsqu'un mot source et un mot cible apparaissaient régulièrement dans un ensemble de séquences alignées, il était probable qu'ils aient la même signification. Ainsi des termes comme « Ζεὺς », « Zeus » et « Jupiter » pouvaient être automatiquement alignés. Nous détaillons ensuite les calculs statistiques que nous effectuons pour chaque ensemble de séquences alignées, et donnons les résultats concrets de ce que nous trouvons dans l'interface que j'ai écrite en Javascript. La seconde partie de l'article, la plus conséquente, est un panorama de l'histoire des traductions homériques en France depuis le XVI^e siècle, de Salomon Certon à Philippe Jaccottet.

- B2.** Khan, Anas Fahad, Andrea Bellandi, Giulia Benotto, Francesca Frontini, Emiliano Giovannetti, et Marianne Reboul. « **Leveraging a Narrative Ontology to Query a Literary Text** ». In *OASISs-OpenAccess Series in Informatics*, Vol. 53. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

Il s'agit d'un article qui a été proposé à une revue de linguistique computationnelle, et qui porte sur le développement d'un modèle d'ontologie narrative. Les ontologies sont des entités (théoriquement fixes) numériques qui regroupent toutes les caractéristiques d'un objet, et qui sont mises en réseau (par des rapports hiérarchiques) avec d'autres objets. Ce type d'architecture se prête facilement à la représentation de systèmes (systèmes grammaticaux, systèmes des idées, etc.), mais beaucoup moins à la représentation de concepts mobiles, ou d'entités mobiles. Nous avons pris comme cas d'étude les personnages de l'*Odyssée*, et en particulier Télémaque dans la *Télémaque* : l'ontologie narrative devait prendre en compte l'identité et l'altérité du personnage à lui-même en début et en fin de trame narrative (qui se manifeste de façon syntaxique comme de façon sémantique). Nous avons donc, à partir d'un modèle d'ontologie LEMON, développé un modèle d'ontologie complexe de représentation narrative. Les expériences menées ont trait aux avancées de l'intelligence artificielle, et de l'apprentissage machine en particulier : l'apprentissage machine, via réseaux neuronaux, doit permettre de déterminer automatiquement la nature des changements de chaque ontologie.

- B3.** Yuri Bizzoni, Riccardo Del Gratta, Federico Boschetti, et Marianne Reboul, « **Enhancing the Accuracy of Ancient Greek WordNet by Multilingual Distributional Semantics** ». *CLiC it*, 2015, 47.

Cet article entend contribuer à l'amélioration du seul WordNet disponible pour le grec ancien, nommé AGWN, développé par un des membres du C.N.R. de Pise, professeur à l'Université de Venise, Federico Boschetti. J'ai mis à profit nos recherches en sémantique distributionnelle multilingue pour les appliquer à un corpus bilingue grec ancien/italien, et les résultats mis à jour d'AGWN sont disponibles en ligne. Dans cet article, nous montrons que les résultats obtenus via un WordNet classique, sans le secours de la sémantique distributionnelle, ne peuvent être optimaux puisqu'ils ne permettent pas la désambiguïsation des termes. Seule la représentation de chaque mot dans l'espace sémantique permet de mesurer à sa juste valeur la justesse du choix de telle traduction par rapport à telle autre.

- B4.** Yuri Bizzoni et Marianne Reboul, « **Contextual Distribution for Textual Alignment** ». *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering* 9, n° 8 (2016): 2068–2072.

Dans cet article, nous montrons la nécessité de recourir aux méthodes de la distribution sémantique via l'informatique. Nous montrons, en comparant onze traductions de l'*Odyssée* en italien ainsi que onze traductions de l'*Odyssée* en français, publiées à des périodes similaires, que non seulement il est possible de déceler par analyse de vecteurs multilingues des mouvements, des tendances de traductions similaires entre les deux langues, mais aussi que la constitution automatique de dictionnaires (tant en langues anciennes qu'en langues modernes) pourrait grandement bénéficier de telles techniques.

- B5.** Alexandre Gefen, Marianne Reboul, « **Mesures et savoirs : quelles méthodes pour l'histoire culturelle à l'heure du big data ?** », *Journal of the International Association for Semiotic Studies*, 2017 (à paraître en 2018).

Dans le cadre d'un projet mené entre le Labex Obvil de Paris-Sorbonne et le Literary Lab de Stanford sur l'histoire de l'idée de littéraire (la définition de la littérature comme mot, comme concept et comme champ), et visant à produire une histoire *empirique* de la littérature, nous avons mené depuis deux ans des expériences de fouille d'un corpus de critique littéraire de 1618 titres, 140 millions de mots (dont plus de 50 000 occurrences du lemme « littérature ») de la fin de l'Ancien Régime à la Seconde Guerre mondiale.

En présentant des exemples précis développés dans cette première expérimentation à grande échelle de mesure de l'histoire des idées, nous présentons les méthodes de *text mining* contemporaines en essayant d'éprouver leur pertinence heuristique et de leur capacité à faire remonter des données significatives pour l'histoire et la théorie littéraire. Nous posons que toute enquête quantitative sérieuse mobilise désormais non une échelle intermédiaire standard et immédiatement lisible, mais le maniement d'outils statistiques dont l'interprétation en sciences humaines pose des problèmes particuliers qui, paradoxalement, ne peuvent être résolus que par leur articulation étroite à du *close reading* et à des mesures fines. Nous partons de la configuration des corpus (ceux produits sur des ensembles spécialisés, ceux constitués par des corpus génériques continus, par exemple *La Revue des deux mondes* au XIX^e siècle, et des corpus globaux comme ceux du *Google Ngram* ou de HathiTrust) pour distinguer ensuite les types de fouilles, allant des mesures fermées (fréquentielles) aux analyses machine non dirigées (*topic modeling*) en passant par des mesure semi-ouvertes (collocation) en réfléchissant à la pertinence relative de ces instruments. Nous entrons ensuite dans une typologie et une heuristique plus fine opposant les quantifications simples, les recherches lemmatisées, les mesures de structures syntaxiques ou sémantiques (association mot et étiquette) et les analyses fondées sur les vecteurs de mots. Mais si l'on distingue des modes d'administration de la preuve directs ou indirects selon que l'on cherche de phénomènes sémantiques ayant des échos textuels nets (présence d'un auteur, champs de référence) ou la détection par des analyses stochastiques de changements de paradigme qui seraient individuellement noyés dans le bruit des analyses fréquentielles, ce domaine émergent qu'est l'histoire quantitative des idées ne peut être exploré que par la conjonction de l'esprit de géométrie et de l'esprit de finesse, la capacité à capturer et à représenter des faits mesurables massifs ne pouvant s'appuyant que sur une intuition cynégétique de terrain et une connaissance des corpus permettant la modélisation des évolutions supposées et la détection de phénomènes historiques inattendus.

C. PUBLICATIONS DANS DES REVUES OU ACTES DE COLLOQUE

- C1. Marianne Reboul, « **Homère et le livre augmentable : pour une interface du Traductologue** », actes du colloque international *Digital Intelligence 2014*, s. d. http://www.univ-nantes.fr/medias/fichier/paper_33_homere_et_lelivreaugmentablepdf_1415427760166.pdf?ID_FICHE=717841&INLINE=FALSE

Je m'interroge dans cet article sur la nécessité de concevoir, avant de trouver les moyens pour le faire, la manière dont un traducteur ou un traductologue pouvait envisager de lire simultanément plus de trente textes à la fois. Créer une interface en vue d'une analyse philologique des textes conditionne la perception des textes eux-mêmes. Afin de créer une interface adéquate pour la comparaison numérique de traductions, il faut prendre en compte la masse textuelle à disposition et le mode de lecture que nous souhaitons privilégier, spécifique au traitement numérique des textes. Il faut aussi déterminer le type d'utilisateur qui est susceptible de recourir à l'outil et les libertés que nous choisissons de lui laisser. Enfin, si les résultats obtenus restent soumis aux prédicats du concepteur, ils sont susceptibles d'enrichir considérablement l'activité de la philologie critique.

- C2. Alexandre Gefen, Mark Andrew Algee-Hewitt, David McClure, Frédéric Glorieux, Marianne Reboul, J.D. Porter, et Marine Riguet, « **Vector based measure of semantic shifts across different cultural corpora as a proxy to comparative history of ideas** », *JADH 2017, Proceedings of the 7th Conference of Japanese Association for Digital Humanities*, "Creating Data through Collaboration", 12, 2017.

Dans cet article, nous évoquons les méthodes que nous avons développées pour :

- Sélectionner, préparer et comparer des corpus similaires de journaux. Nous avons construit de corpus similaires en français et en anglais, dont les textes ont été publiés entre 1830 et 1880. Nous avons prétraité chacun des corpus en normalisant les textes au format TEI et en les lemmatisant.

- Représenter graphiquement une évolution comparative de certains mots-clés, comme « littérature/literature » en représentant leurs vecteurs dans l'espace sémantique, en utilisant Word2Vec et GloVe.
- Définir une manière fiable de quantifier le degré de polysémie des mots. Pour ce faire, nous construisons pour chaque langue un ensemble de bigrammes pour chacun des mots du vocabulaire présent, avec les dix plus fréquents adjectifs associés. Nous calculons ensuite la similarité cosinus entre chaque bigramme. Nous construisons ensuite une matrice ordonnée en fonction du score de similarité, et nous calculons le rang moyen (« mean rank »). Nous pouvons ainsi déduire le coefficient de « clustering » des bigrammes. Un bigramme de bas rang représenterait, selon cette hypothèse, un terme hautement polysémique, et plus le rang est bas, plus le degré de polysémie est élevé.

Cet ensemble de méthodes nous permet de déterminer les pôles attractifs dans l'évolution de termes littéraires clés, et de représenter leur comportement de façon diachronique. Nous confrontons alors nos résultats avec les théories existantes sur l'histoire de la littérature, confirmant ainsi numériquement la spécialisation du champ des études littéraires.

- C3. Yuri Bizzoni, Marianne Reboul, « **Étude comparative numérique de traductions italiennes et françaises de l'*Odyssée* d'Homère du XVII^e au XX^e siècle** », actes du colloque « La traduction à l'ordre des discours, Manières de traduire, façons d'écrire » des 29 et 30 mai 2015, organisé par Louis Watier, Presses Universitaires de la Sorbonne, « Comparatismes en Sorbonne » (à paraître en 2018).

Nous montrons dans cet article qu'il existe des motifs communs identifiables entre les traductions françaises et italiennes de l'*Odyssée* d'Homère, en utilisant les outils de Traitement Automatique de la Langue que nous avons modifiés et créés, en nous basant sur un macro-corpus allant du XVII^e au XX^e siècle. Nous montrons également de façon statistiquement quantifiable dans cet article qu'a lieu en France un *aggiornamento* de la traduction française à partir des traductions de Madame Dacier qui n'a pas lieu en Italie, et que l'histoire des langues, tant françaises qu'italiennes, ou encore l'histoire de la connaissance du grec, permet de comprendre pourquoi les tendances traductives de l'un et l'autre pays n'ont pas pris le même tour. Nous constatons d'autre part qu'après que les traducteurs français et italiens prennent des chemins totalement divergents à partir de la fin du XVII^e siècle, leurs chemins se rejoignent à nouveau à la fin du XIX^e siècle.

- C4. Marianne Reboul, « **Les chapitres dans les Role Playing Video Games du XXI^e siècle : de la diégèse à l'extension cellulaire d'un univers** », actes du colloque international « Nouvelles aventures du chapitre : division narrative et histoire des supports », du 2 au 4 novembre 2017, aux Presses Universitaires de Montréal (à paraître en 2018).

Cet article concerne un type spécifique de R.P.G., les R.P.V.G., ou « Role Playing Video Games ». Qu'appelle-t-on chapitre dans les jeux vidéo, et particulièrement les R.P.V.G. ? À la différence des chapitres en littérature, deux éléments servent à définir ce qu'est un chapitre, ou tout du moins une séquence qui peut s'apparenter à un chapitre. Un très grand nombre de jeux reprennent ainsi la terminologie littéraire pour qualifier leur fragmentation : l'on retrouve souvent la notion de chapitres, d'actes ou de scènes. Nous pouvons alors penser les chapitres dans les jeux comme des séquences de la diégèse, attendu que la diégèse elle-même n'est pas nécessairement linéaire, mais peut être pensée comme un réseau plus qu'une ligne de trajectoire. Enfin, il est des cas où le chapitre dépasse le jeu lui-même, et est plus souvent appelé « opus » de façon générique. Ces cas sont cependant plus rares dans les jeux en général, mais surreprésentés dans les R.P.V.G. Les R.P.V.G. proposent donc une lecture du phénomène du chapitre de façon singulière : la puissance machine et le public auquel ils s'adressent supposent désormais la

construction d'une trame toujours plus complexe, de moins en moins conçue comme un dévoilement diégétique linéaire que comme une mise en réseau potentiellement infinie.

D. COMMUNICATIONS DANS DES COLLOQUES INTERNATIONAUX

D0. Colloques à venir avant septembre 2018 :

« Les techniques de la traduction automatique appliquées aux textes littéraires », intervention pour la journée d'études de l'ENS Lyon « Traductions et Outils numériques », le 4 avril 2018, suivie d'une « master class » que j'animerai sur les outils d'alignement automatiques écrits en Java et en Python.

« Expériences de topic modeling sur le corpus de romans de l'ANR Chapitres », Colloque international de Cerisy, « Construire le récit, histoire et poétique des chapitres », du 25 juin 2018 au 2 juillet 2018.

« Methods of Meaning: Deciphering the History of "Literature" With Two Word Vector Approaches », « long paper » (confirmed), en collaboration avec le Literary Lab de Stanford, pour la Digital Humanities 2018 Conference, Mexico, du 26 au 29 juin 2018 :

Dans cet article nous comparons deux méthodes pour mesurer les relations qu'entretiennent les mots entre eux, lorsque ceux-ci sont employés dans des corpus de critique littéraire. Plus précisément, nous traçons les changements de sens du terme « littérature » en ayant recours aux vecteurs de mots. Nous comparons la mesure de similarité cosinus avec l'indice de "K-means clustering", sur des modèles de vecteurs de corpus critiques en anglais et en français depuis 1800 jusqu'à 1940. Nos résultats aboutissent à une meilleure compréhension du développement de la littérature comme un champ conceptuel spécifique, illustrent les différences auxquelles les humanistes peuvent s'attendre lorsqu'ils utilisent les deux méthodes susdites, et interrogent le concept de « sens » comme propriété relationnelle des mots.

D1. 4 nov. 2017 « Les chapitres dans les R.P.G. du XXI^e siècle : de la diégèse à l'extension cellulaire d'un univers », Colloque de l'Université de Montréal « Le chapitre avant et après : division narrative et histoire des supports ».

Intervention publiée dans les actes mentionnés ci-dessus.

D2. 11 sept. 2017 « Creating Data through Collaboration », Japanese Association for Digital Humanities, « Vector based measure of semantic shifts across different cultural corpora as a proxy to comparative history of ideas ».

Intervention publiée dans les actes mentionnés ci-dessus.

D3. 13 juillet 2017 « Textual alignment in Translation Studies », colloque international Historical Text Reuse Data Workshop, Université de Leipzig.

Quelles sont les techniques d'alignement déjà en place pour le grec ancien et comment les améliorer ? Notre participation consistait à faire la démonstration directe de notre programme *Odyssens*, non seulement avec notre propre corpus de textes numérisés de l'*Odyssée*, mais aussi avec d'autres textes grecs tirés du corpus de la Perseus Digital Library. Nous avons expliqué en quoi consistait l'algorithme d'alignement de Needleman Wunsch, mais aussi quelles pouvaient être les techniques plus récentes et plus complexes employées pour l'alignement automatique des textes, notamment en « machine learning ». Nous avons repris la théorie du système d'encodage/décodage

des traducteurs automatiques neuronaux en spécifiant quels seraient les prérequis techniques pour qu'ils soient adaptés à l'alignement de textes littéraires, et notamment de textes littéraires antiques.

- D4.** 14 février 2017 « Diachronical vectors and polysemy », Stanford Literary Lab Workshop, en collaboration avec David McClure, MIT.

Comment mesurer l'évolution des mots dans le temps ? Nous avons, conjointement avec David McClure, du Stanford Literary Lab, élaboré une technique de représentation graphique, non seulement via Gephi, mais aussi avec notre propre implémentation en python (intitulée « hist-vec »), de l'évolution du degré de proximité de mots clés entre eux dans le temps. Nous avons ainsi pu mesurer, entre autres, le rapprochement explicite du terme de « littérature » avec celui de « science » sur un corpus de critique française lemmatisé. Nous avons aussi mis en pratique les différentes techniques de mesure de polysémie des mots et avons comparé leurs résultats. Nous en sommes arrivés à la conclusion que les techniques les plus efficaces étaient celles capables de mesurer le coefficient de « clustering » des mots.

- D5.** 13 juillet 2016 « The *Odysseus* Project », Digital Humanities 2016 Conference, Jagiellonian University & Pedagogical University, Krakow, nominée du prix Fortier des jeunes chercheurs de la conférence.

Le programme *Odysseus* a pour but de proposer aux comparatistes une façon simple de repérer les grandes tendances et les phénomènes notables d'une traduction à une autre, indépendamment de la langue du traducteur. Mais le programme ouvre aussi la voie à la création d'autres outils pour l'étude de corpus multilingues (dictionnaires, détection automatique de paraphrases interlinguistiques), à de nombreuses analyses statistiques (fréquence des mots, « type / token ratio », etc.) qui représentent un apport substantiel à l'étude de l'Histoire des traductions. Les spécialistes en traductologie peuvent bénéficier de cette approche : nul besoin de se fier à leur intuition ou à un décompte manuel des éléments du texte. Il est possible combiner efficacement considérations quantitatives et qualitatives. La brève interprétation diachronique que nous produisons dans l'article est largement facilitée par l'usage de notre programme, et permet de s'affranchir un temps, pour les infirmer ou les confirmer, des hypothèses déjà formulées à l'égard des différents traducteurs, des modes de traduction et de l'évolution de la traduction en général. D'autre part notre travail pourrait aussi largement être utile à un champ plus large que celui de la traductologie, puisqu'il peut permettre de comparer dynamiquement de nombreux textes proches, ou dont on suppose la proximité, dans une seule et même opération, quelle que soit leur langue. Enfin notre étude met en évidence un constat simple trop souvent oublié par les lecteurs des textes classiques : la traduction est un filtre permanent glissé entre le sens authentique perdu dans une langue source souvent inaccessible au public et une compréhension moderne et contextualisée du texte. Autrement dit, le programme met en évidence que la lecture d'une traduction n'est jamais anodine, et qu'il existe en réalité autant de textes qu'il existe de traducteurs : l'*Odysée* que nous lisons n'est pas un seul et même texte, mais une myriade des filtres successifs, qui s'inspirent, s'imitent, se copient ou se rejettent les uns les autres. Nous pouvons, du premier regard, constater que le corps textuel de l'*Odysée*, ou de tout autre texte traduit, est un ensemble dont chaque cellule influence la perception que chacun tire de sa lecture.

- D6.** 18 juin 2016 « Les vecteurs dans la boîte à outils des littéraires : quel instrument pour quels usages ? », colloque international de Cerisy, « Des humanités numériques littéraires ? » (publication à paraître).

Les vecteurs, outils fondés sur une analyse syntaxico-statistique des textes, sont en train de prouver leur utilité dans le domaine, déjà littéraire, de la traduction, mais aussi de l'établissement des textes anciens. Au-delà de ces applications, quelles sont les autres possibilités offertes par ces instruments capables de faire émerger des notions en fonction des contextes syntaxiques scrutés ? Quelles sont leurs limites ? Comment adapter ces logiciels pour les mettre au service de l'étude de la littérature ?

D7. 4 déc. 2016 « De l'intraduisible à l'incalculable : Problèmes de comparaison automatique de traductions », colloque international de l'Université d'Evry, « Intraduisible, les méandres de la traduction ».

Penser l'intraduisible en Traduction Statistique Assistée ne signifie pas nécessairement se confronter à l'échec de la machine. Tout algorithme qui s'applique à un corpus varié est nécessairement confronté à des choix insolubles, qui impliquent qu'une subjectivité entre en compte. En l'état actuel, il n'existe pas de traducteur automatique suffisamment puissant ou suffisamment adapté à des langues autres que l'anglais pour donner des résultats unanimement satisfaisants. Mais il est indéniable que la capacité de calcul d'une machine infiniment supérieure à celle d'un cerveau humain ne peut être mise de côté parce que confrontée, au même titre qu'un humain, à des apories. L'intraduisible en traduction statistique est partie intégrante du calcul statistique. Enfin, l'intraduisible comme incalculable existe aussi parce qu'il est des temps du texte que la machine ne doit pas aligner, parce que le traducteur n'a pas souhaité les traduire. La difficulté de la machine est de pouvoir déterminer si l'incalculable est une erreur ou un choix.

D8. 24 avril 2015 « Traductions homériques et tendances littéraires », Göttingen Dialog in Digital Humanities, Université de Göttingen.

C'est au cours de cette intervention que nous avons évalué les différentes techniques d'alignement automatique de séquence à séquence, et qu'après démonstration nous avons pu montrer que l'alignement par un système de mots ancrés, tels les noms propres ou les ponctuations fortes, étaient des pivots efficaces pour l'alignement automatique de traductions. En effet nous avons comparé les résultats obtenus avec un aligneur classique (mKalign) et ceux obtenus avec un séquençage de noms propres à noms propres (comme le fera par la suite le programme *Odysseus* que nous avons développé), avant de donner les séquences pour un alignement mot-à-mot via MGiza++, et nous avons constaté que l'alignement de MGiza++ était nettement plus performant en présence de séquences définies avec des pivots stables. Nous avons mené l'ensemble de ces expériences sur le corpus encore incomplet des traductions françaises de l'*Odyssee* et de l'*Iliade*, et avons tiré des conclusions préliminaires de quelques comparaisons de traductions automatiquement alignées via MGiza++.

D9. 14 juillet 2014 « Scholars and NLP : issues and perspectives », workshop international organisé par Gregory Crane à l'Université de Leipzig.

Cette intervention avait pour but de montrer qu'il était nécessaire pour les chercheurs en littérature, et spécifiquement pour les spécialistes en langues rares ou mortes, de s'inspirer de l'intérêt croissant des grandes industries pour les fouilles de données et le Traitement Automatique des Langues, pour mettre à profit les outils existants au service de la recherche en littérature, ancienne comme moderne. Nous avons alors procédé à un rappel de l'ensemble des outils disponibles, mettant en avant leurs avantages et insuffisances, et la nécessité de les adapter à des objets spécifiques, notamment en respectant les particularités linguistiques des objets étudiés.

E. COMMUNICATIONS EN COLLOQUES NATIONAUX OU JOURNEES D'ETUDES NATIONALES

- E1.** 18 nov. 2017 « Présentation et compte rendu des travaux de thèse sur la comparaison semi-automatique des traductions de l'*Odyssee* », journée d'études « Sur les traces du traducteur : (re)traductions et édition numérique des textes antiques », Saint-Omer.
- E2.** 3 mars 2017 « Raconter l'aventure en traduction : le *muthos* d'Ulysse dans les traductions françaises de l'*Odyssee* », journée d'études « Les imaginaires de la traduction », Maison de la Recherche, 3 rue des Irlandais, Paris.

Lorsqu'Ulysse arrive à la cour d'Alcinoos, après être passé chez Calypso, avoir fait naufrage et avoir été accueilli par Nausicaa, il est d'abord ému par le récit de la guerre troyenne, et, sous les injonctions des assistants, est appelé à raconter ses aventures depuis son départ d'Ilion. Ulysse est souvent, au long des deux récits homériques, un adepte ambigu du *muthos* et de la *peitho* : il sait convaincre les rois et les guerriers, mais il est déjà connu pour sa ruse, voire comme un fieffé menteur. Au début du chant IX de l'*Odyssee*, Ulysse commence son récit. Selon les époques et tendances traductives, la perception du personnage d'Ulysse change, et le crédit à accorder à sa parole également. Nous souhaitons, dans cette communication, voir s'il existe une logique historique au crédit accordé à la parole d'Ulysse à travers les traductions. Dans un premier temps nous analysons le traitement du *muthos* odysseéen dans les traductions du XVI^e et XVII^e siècle, période d'incertitude quant à l'apprentissage et la connaissance du grec, laissant place à une grande variété de traitement et à des traductions parfois très éloignées de leur source. Nous montrons ensuite quelle fut la compréhension du *muthos* d'Ulysse au XVIII^e et XIX^e siècle, moment de redécouverte scientifique du grec et du retour parfois troublant à la source et à l'imitation non seulement sémantique mais syntaxique et phonétique du grec. Enfin nous montrons que si la tendance littéraliste et la compréhension archéologique du texte perdurent au XX^e siècle, les traductions produites laissent beaucoup plus la place à la double lecture du récit d'Ulysse, laissant planer l'ambiguïté sur le crédit à accorder au récit de ses aventures.

- E3.** 29 mai 2015 « Étude comparative numérique de traductions italiennes et françaises de l'*Odyssee* d'Homère du XVII^e au XX^e siècle », colloque « La traduction à l'ordre des discours, Manières de traduire, façons d'écrire ».

Intervention publiée dans les actes mentionnés ci-dessus.

F. COMMUNICATIONS INVITEES EN SEMINAIRE

- F1.** 13 fév. 2018 « Apprendre à une machine à traduire : modéliser les identités multilingues », séminaire « Le littéraire, du papier au numérique » organisé par Michel Bernard, Paris 3 Sorbonne Nouvelle.

Nous avons d'abord fait un rappel de la nécessité de structuration de base des données analysées, de l'histoire de la traduction automatique, puis nous avons abordé les différents modèles algorithmiques disponibles pour la traduction automatique, comme les HMM, et observé en détail les modes de calculs d'encodeurs neuronaux comme ceux de Google, en passant par l'analyse des techniques de « skipgram » et de « negative sampling ». Nous avons enfin montré un cas d'étude, notre aligneur *Odysseus* et la démonstration en ligne sur notre site.

F2. 24 mars 2017 « Comment pense une machine ? », Séminaire #EPHN 2017, organisé par Suzanne Dumouchel.

L'exposé s'est d'abord concentré sur l'histoire de la traduction automatique avant de rendre compte des techniques contemporaines qui sont développées. D'emblée, on s'aperçoit que le questionnement n'est pas nouveau. Bien avant les premiers ordinateurs individuels, le problème était déjà posé. Un des jalons essentiels apparaît en 1947, avec le Translation Memorandum, un texte essentiel qui affirme que le langage est un code et qu'il existe des universaux linguistiques que l'on peut décoder. Il faut aussi avoir à disposition une masse de corpus bilingue qui puisse servir de modèle pour permettre la recomposition dans la langue d'arrivée d'un texte initial. Les techniques de traduction automatique se sont largement développées grâce à Internet et à la publication en ligne. Le grand nombre de données a contribué au progrès mais de façon paradoxale. En effet, l'importance des données a permis d'affiner plus précisément les pratiques de traduction mais la grande diversité des données, c'est-à-dire leur qualité variable, a augmenté la marge d'erreur. Nous avons présenté plusieurs solutions, de la traduction qui s'appuie sur les réseaux neuronaux à celle basée sur la probabilité, et mis en avant les difficultés rencontrées et les points forts et points faibles de chacune des techniques.

F3. 6 avril 2016 « L'Odyssée aujourd'hui. Projets de recherche et d'enseignement », séminaire EHES « Digital Humanities » organisé par Aurélien Berra et Arnaud Laborderie

Présentation en deux temps. Présentation des résultats de thèse, puis ouverture sur les projets pédagogiques en cours envisagés avec la Scuola Normale di Pisa et l'Istituto di Linguistica Computazionale Zampolli du Consiglio Nazionale delle Ricerche (projet soumis au H2020 sur l'élaboration d'une approche didactique des textes grecs via la réalité augmentée).

G. AUTRES TYPES D'INTERVENTIONS

G1. Mars 2017 « Traduction et numérique », conférence « Objet de Recherche » filmée par l'équipe de l'audiovisuel de l'Université Paris 4 Sorbonne, et organisée par Virginie Thibaud, <https://www.franceculture.fr/conferences/universite-paris-sorbonne/objet-de-recherche-marianne-reboul-traduction-et-numerique>

G2. 17 mai 2017 « Ces nouveaux logiciels qui lisent », émission « Les nouvelles vagues » de Maylis Besserie, <https://www.franceculture.fr/emissions/les-nouvelles-vagues/lire-3-ces-nouveaux-logiciels-qui-lisent>

RESPONSABILITES COLLECTIVES ET PROJETS EN COURS

2018 « **Cooperation Project in Localizing Perseus 5.0 for French Language** », contrat de collaboration passé entre Paris III Sorbonne Nouvelle par Marianne Reboul et l'Université de Leipzig par Gregory Crane

Notre objectif est d'adapter le site de Perseus pour le français, mais aussi de fournir des traductions alignées automatiquement. Ainsi, le projet se déroule en cinq étapes successives : 1) la création du français pour la localisation de Scaife DL Viewer (analyse préliminaire des conditions d'internationalisation de Scaife DL Viewer ; création des bundles en français ; traduction des éléments suivants : labels, menus, « tooltips », « buttons », « wizards ») ; 2) la création des alignements (le travail d'alignement des textes présents dans le « repository » avec les textes originaux en grec ancien, « preprocessing » des textes français avec analyse linguistique avec « POStagging » et lemmatisation, choix de la meilleure méthode

d'alignement, adaptation des textes français aux normes « URN CTS » en fonction du texte source correspondant) ; 3) intégration de dictionnaires et grammaires grec-français et latin-français (à partir des résultats des alignements, il est possible de générer une liste de couples de mots grecs-français et latins-français, permettant ainsi d'aider à la création de dictionnaires et grammaires bilingues) ; 4) contribution à l'alignement des traductions anglaises (les instruments développés pour la création des alignements pourront être utilisés et adaptés afin de contribuer à l'amélioration d'alignements des traductions en anglais) ; 5) aide à l'internationalisation de Perseus 5.0 pour le croate (production d'un software open-source pour les alignements, afin de partager le code de programmation avec les équipes croates, service d'aide à la demande pour le « bootstrap » de cette tâche).

2018 **Projet européen en cours d'évaluation HERANET « Aspasia – Historical Languages and Literatures in Translingual Digital Space »** sous la direction de Gregory Crane

Dans ce projet auquel participent des équipes de recherche italiennes, allemandes, croates, estoniennes, brésiliennes, canadiennes, et américaines, nous sommes responsable de la branche française. Le projet Aspasia se concentre sur trois facteurs communs en Europe : la présence d'une très large collection numérique de textes grecs et latins ; l'interaction des individus avec ces textes ; la conception d'internet comme un espace public. Le projet entend mesurer et théoriser les dynamiques de l'inclusion, de l'exclusion, de mesurer les facteurs qui en sont la cause, pour stimuler l'interaction avec les résultats de la recherche sur les textes anciens. Le projet part du principe que ce ne sont pas exclusivement les spécialistes, mais aussi les étudiants et amateurs de la culture antique qui apportent les contributions majeures à cette interaction, et que le cadre officiel de la communauté européenne peut grandement influencer cette interaction et la faciliter. Il s'agit donc d'un projet extrêmement large et conçu à long terme.

2016-... **Projet international « Towards an Empirical History of Literature »** sous la direction de Franco Moretti, Mark Algee-Hewitt (Stanford Literary Lab) et Didier Alexandre (Labex OBVIL).

Projet organisé par le Labex OBVIL et le Stanford Literary Lab, dans le cadre du financement MELLON/FMSH, intitulé « Towards an Empirical History of Literature », dans le cadre du programme « Transatlantic program for collaborative work in the field of digital humanities ». C'est dans le cadre de cette collaboration qu'ont été préparées les communications à Kyoto et Mexico. Mon rôle est charnière dans le projet dans la mesure où je peux à la fois contribuer littérairement et informatiquement aux expériences effectuées.

SEJOURS A L'ETRANGER

2014-2015 **Séjour de trois mois au sein de l'Istituto di Linguistica Computazionale Zampolli** du Consiglio delle Ricerche di Pisa (Mission des Doctorants à l'étranger et Labex OBVIL)

J'ai pu me rendre à Pise grâce à une bourse de mobilité pour les jeunes chercheurs « Mission des Doctorants à l'étranger ». L'équipe avait fait, deux mois auparavant, une présentation de son projet « Greek into Arabic », en montrant une interface dynamique et des protocoles d'alignement de données qui correspondaient parfaitement à mon projet, encore jeune, de recherche. J'ai ainsi rencontré des chercheurs avec qui je travaille encore fréquemment sur les questions d'alignement, comme Angelo Mario del Grosso ou Yuri Bizzoni, et avec qui j'ai écrit plusieurs articles. Lors de mon second séjour, j'ai été envoyée à Pise par le directeur du Labex OBVIL, le Professeur Didier Alexandre, pour travailler sur le développement de modèles d'ontologies. J'ai travaillé intensivement avec d'autres chercheurs dont j'avais fait la connaissance auparavant, notamment Francesca Frontini et Fahad Khan. Le rapport de nos activités écrit par le

responsable du laboratoire souligne mon enthousiasme, ma capacité de travail et ma bonne intégration dans l'équipe.

2016-2017 **Séjour d'un mois au sein du Stanford Literary Lab**, Université de Stanford (d'abord sous la direction de Franco Moretti, puis sous la direction de Mark Algee Hewitt)

Séjour effectué dans le cadre du projet MELLON « Towards an Empirical Literary History of Literature ». J'ai pu rencontrer l'ensemble des chercheurs du laboratoire, ainsi que Franco Moretti, avec qui j'ai eu souvent l'occasion de m'entretenir. J'ai ainsi rencontré en personne David McClure avec qui j'avais déjà travaillé à distance et nous avons pu développer en commun un projet Python sur la représentation et mesure de la polysémie des mots. J'ai eu aussi le privilège de rencontrer Christopher Manning, le concepteur et responsable du projet StanfordCoreNLP et concepteur du vectoriseur GloVe, avec qui j'ai discuté des perspectives de créations de modèles neuronaux multilingues et avec qui je suis toujours en correspondance. J'ai eu aussi la possibilité de travailler avec une autre doctorante Coréenne, Eun Seo Jo, avec qui nous concevons actuellement un projet de détection automatique des genres dans le cadre des études de « Sentiment Analysis ». Nous avons ainsi pu jeter les jalons de ce qui a constitué notre proposition de « long paper » pour la conférence Digital Humanities 2018 à Mexico.

RESUME DE LA THESE

Notre intention dans cette étude a été de donner une approche de l'histoire des traductions de l'*Odyssée* d'Homère depuis le XVI^e siècle jusqu'au XX^e siècle, en retraçant ce qui nous a semblé être la logique évolutive qui a pu conditionner la production de ces traductions, et en montrant que c'est entre la fin du XVIII^e siècle et le début du XX^e siècle que s'opère le tournant dans la façon d'appréhender Homère et de le traduire, en lien direct avec les progrès de la philologie et de l'archéologie.

Cette étude, sur une période temporelle si ample, prenant en compte plusieurs dizaines de textes, a été rendue possible par l'utilisation des techniques numériques. Les techniques que nous employons dans notre analyse permettent d'obtenir des résultats tangibles statistiques fiables et stables pour tous les textes, indépendamment de la période à laquelle ces derniers ont été produits. Elles permettent aussi une visualisation des résultats sans précédent.

Nous avons cherché à produire un outil logiciel capable d'assister la réflexion comparatiste en permettant d'aligner des traductions en français d'Homère sur plus de quatre siècles : outil qui n'existait pas en tant que tel et pour lequel nous avons dû faire appel à des technologies de sémantique distributionnelle avancées capables de rapprocher des traductions d'un même texte dans des langues différentes. Notre recherche s'est donc dédoublée en une réflexion historique et en un développement informatique, les deux dimensions étant étroitement liées dans le cadre de ce qu'on appelle désormais les « humanités numériques ». D'un côté, il fallait inventer des méthodes et des algorithmes capables de mettre en parallèle au mot près des traductions aussi différentes que celle de Peletier (1547) et de Jaccottet (1955), ce qui n'avait jamais été fait et pose des problèmes techniques exposés dans toute une partie, technique, de notre travail. D'un autre, il fallait développer une réflexion méthodologique et épistémologique sur la manière dont on pouvait lire parallèlement 27 textes, mesurer l'évolution de leurs éléments¹, dans cette « forme spécifique de savoir² » que sont les visualisations informatiques. L'ouvrage est donc à la fois une étude informatique et une étude littéraire, l'une étant indissociable de l'autre : si l'esprit humain peine à synthétiser des résultats portant sur au moins trois millions de mots, les machines sont évidemment incapables d'interpréter leurs propres résultats.

1 David Bamman et Gregory Crane, « Measuring Historical Word Sense Variation », *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, New York, ACM, coll. « JCDL '11 », 2011, p. 1-10.

2 Franco Moretti, *Graphs, maps, trees: abstract models for a literary history*, Verso, 2005, p. 1.

Lorsque Peletier du Mans traduit pour la première fois Homère en français, il est confronté à une tension : d'un côté demeure le privilège donné aux Anciens, auquel s'attache une sacralité respectueuse, imprégnée de l'histoire de la transmission du texte et de tous les commentaires qui ont forgé l'image d'Homère avant même qu'il ne soit traduit ; de l'autre se dessine l'affirmation d'une langue française qui n'avait jusque-là pas voix au chapitre, puisqu'elle ne bénéficiait pas encore des lettres de noblesse de ses illustres prédécesseurs, le latin, qui s'est octroyé tout le champ des commentaires érudits, et le grec. Le problème du « génie de la langue française » pose la question de la légitimité historique d'une langue qui jusque-là ne bénéficie pas d'ascendants directs reconnus, et qui doit s'appuyer sur l'héritage des Anciens. C'est cette tension qui condamne nombre de tentatives de traductions à la fin du XVI^e siècle à l'échec : l'imperméabilité de l'héritage à la langue nouvelle parasite encore trop la perception du texte source. L'Homère français ne saurait naître, depuis Jamyn jusqu'à la fin du XVII^e siècle, que d'un arrachement à un héritage trop pesant : Homère doit, pour exister, endosser les habits du temps, et s'accommoder à des mœurs qui ne sont pas les siennes. L'enjeu de la légitimation de la langue française prend le pas sur la recherche d'exactitude et sur la fidélité à l'objet traduit. Grâce au programme *Odysseus*, nous avons pu identifier statistiquement les parentés entre les traducteurs et voir dans quelle mesure leur conception du texte source et de la langue cible a évolué : depuis les premières tentatives de Peletier jusqu'à la traduction de La Valterie, le texte grec a dû se plier aux exigences d'un français de plus en plus poli, précieux. Les traductions d'Homère en français, ainsi conçues, ne pouvaient qu'aboutir à une émancipation totale du grec, rendant bien souvent l'*Odysée* méconnaissable, en lui substituant un second original. Le grec a dès lors pour fonction de se faire l'écho d'une grandeur qui lui est étrangère, celle du « génie de la langue française ». Si Madame Dacier a eu longtemps la réputation d'avoir mis fin au règne de la préciosité dans les traductions homériques, nous avons pu constater, notamment grâce à *Odysseus*, que cette opinion devait être nuancée. Madame Dacier est elle aussi l'héritière d'une culture française dans laquelle Homère n'est pas tout à fait soluble. Elle a cependant le grand mérite de remettre en lumière le sens des mots grecs, qui reprennent un poids indépendant de la langue française. Si Bitaubé et ses héritiers suivent ses pas, ils ne s'affranchissent pas encore de la nécessité de se plier aux attentes d'un lecteur formé à l'éloquence et au bon goût.

La véritable rupture dans l'histoire de la traduction de l'*Odysée* d'Homère est le corollaire du progrès philologique et archéologique à l'échelle européenne qui prend toute son ampleur au XIX^e siècle. La langue française ne cherche plus sa légitimité, puisqu'elle est désormais aussi classée, à la fin du XVIII^e siècle, comme une langue du savoir. L'horizon culturel s'élargit : les Lumières se posent comme étant à la fois le fruit de l'héritage antique et l'ouverture à des horizons plus larges, qui dépassent les frontières nationales. Les nouveaux établissements du texte homérique, l'essor de la philologie allemande et de la philologie française, contribuent à une perception désacralisée du texte antique, qui se dépouille peu à peu du caractère moral que la Révolution lui avait conféré. Le texte est de plus en plus perçu comme un texte étranger, le texte témoin d'une civilisation et d'une langue chronologiquement circonscrites, porteur de ses mœurs propres et de ses traditions. La langue elle-même devient objet de savoir, non plus tant pour l'objet qu'elle transmet et révèle que pour sa valeur linguistique et esthétique intrinsèque. L'enseignement du grec se répand, les hellénistes sont de plus en plus chevronnés, et l'école met l'accent sur la connaissance de la langue autant que sur la connaissance de la culture grecque, qu'il n'est plus question d'adapter à la culture française : connaître la culture grecque, c'est connaître les fondements de la culture française, mais ce n'est plus la confondre avec elle.

Les traductions participent pleinement de cette évolution de la perception de la langue étrangère. Les traductions scolaires foisonnent, et les critères d'évaluation changent : une bonne traduction est une traduction qui sait montrer l'érudition de son auteur. Conspuée moins de deux siècles auparavant, la fidélité au texte devient un gage de qualité. C'est dans ce contexte que sont formés les futurs traducteurs, qui apprennent le grec en ce qu'il est l'écho d'une civilisation ancienne et révolue. La langue française, installée, instruite, balisée, a désormais le champ libre pour se renouveler au creuset de l'étranger. Maintenant qu'Homère est connu, maintenant qu'il est enseigné, il est possible de prendre la pleine mesure de son

étrangeté, et de trouver en lui une nouvelle source d'inspiration esthétique. C'est la redécouverte du texte homérique par la philologie et par la linguistique qui permet le renouvellement de la perception du panthéon antique. Ce processus commence avec la seconde version de Dugas-Montbel et atteint son apogée avec Leconte de Lisle. Le texte grec redevient ce qu'il était encore pour les premiers traducteurs d'Homère : un champ de possibilités nouvelles. Mais la dichotomie entre le couple fidélité-prose et libéralisme-vers se maintient encore : au vers reste attaché l'idée de la suprématie de la langue cible sur la langue source. L'inspiration de l'étranger ne peut encore venir que de l'abdication de la poésie face à l'exactitude la plus stricte. Les traductions versifiées du XIX^e siècle ne survivent pas à cette réputation du vers. Les auteurs ne négligent pas l'apport des variétés sonores du grec, mais ils n'envisagent pas l'*Odyssee* comme une œuvre poétique complète : traduire cette œuvre implique alors encore de se départir de l'espoir d'une traduction qui rende la prosodie originale.

Le premier à oser soutenir qu'une exactitude philologique et la conservation de la forme versifiée sont possibles et indispensables pour reproduire fidèlement l'effet du texte grec est Victor Bérard. Il s'est efforcé, une grande partie de sa vie, à resituer l'*Odyssee* d'Homère dans le monde contemporain. Son érudition et sa connaissance du monde maritime ont été les facteurs prédominants de sa compréhension du texte homérique : l'*Odyssee* est à la fois un document, une assise scientifique qui nous renseigne sur le monde homérique, mais aussi un texte poétique, qui peut dès lors prétendre à l'universel. C'est parce qu'Homère est un poète qu'il a relayé certaines vérités essentielles et immuables de l'espèce humaine : il est donc, pour Bérard, invraisemblable de vouloir restituer, au XX^e siècle, les vérités homériques sans en passer par une restitution de la prosodie. C'est par le vers que parviennent les accents des rituels, c'est par le vers chanté que reviennent les sons de la poésie représentée. Cet auteur parvient ainsi à coupler ce qui paraissait encore moins d'un siècle avant comme une aporie : poésie versifiée et exactitude philologique ne sont que les deux faces d'une même médaille. Est-ce à dire que les traductions de Bérard ont eu un effet immédiat sur ses contemporains et sur ses successeurs ? Probablement pas, mais il a permis de rétablir la perception poétique de l'*Odyssee*, tant dans son contenu que dans sa forme, ce dont tous ses contemporains lui sont redevables, depuis Dufour qui considère avant tout l'*Odyssee* comme un texte qu'il faut traduire exactement tout en se laissant porter par le récit, jusqu'à Jaccottet, qui doit à Bérard le rétablissement de la légitimité du vers. Jaccottet diffère de Bérard en bien des points, puisqu'il ne s'agit jamais pour lui de rétablir Homère au XX^e siècle, mais de capter l'écho de vérités éteintes, ou plutôt, pour le dire avec ses mots, comme une éphémère « fraîcheur d'eau au creux de la main ».

Odysseus, l'outil numérique que nous avons créé de toutes pièces n'est pas seulement un instrument destiné à faciliter le travail du chercheur : il permet de trouver des phénomènes qui sont indécélables à l'œil nu, et d'obtenir des résultats qui ne peuvent pas, même avec la plus grande rigueur, être obtenus par un humain. Nous voyons ainsi la double fonction de l'outil informatique, qui sert à la fois d'outil de vérification et de découverte. D'une part, l'outil informatique permet de rendre vérifiables les intuitions de la philologie qui sont opérationnalisées et parfois visualisables de manière pédagogique. D'autre part, des phénomènes inattendus peuvent être rendus visibles par les expériences, comme des inflexions que la longue durée ou la quantité de textes auraient masquées. Enfin le programme informatique que nous avons développé est né des besoins spécifiques de la recherche philologique qui était la nôtre, mais il répond également à un idéal humaniste : notre outil a été conçu dans un langage informatique commun, souple, et modifiable ; il n'est pas destiné à notre usage exclusif et s'offre à la communauté des chercheurs. Car le grand apport qui peut être celui des humanistes numériques est que le fruit du travail produit est exploitable par d'autres : l'incrémentation potentielle du savoir est un principe fondateur des Humanités numériques.

Notre travail ouvre des perspectives que nous n'avons pas encore pu explorer, parce qu'elles ne faisaient pas partie de notre objet d'étude. Mais nous pensons que de nombreux progrès peuvent encore être faits en mettant le numérique au service de la traductologie, des lettres anciennes, et de la littérature comparée en général. Pour la traductologie tout d'abord, les Humanités numériques, telles que nous entendons les pratiquer, peuvent nous aider à comprendre la genèse historique des traductions, la

transformation génétique des langues, et l'évolution des concepts. Il est par exemple possible de comparer chronologiquement les tendances traductives chez l'ensemble des érudits européens ayant traduit l'*Odyssée*. Pour les lettres anciennes, de très nombreux apports sont possibles, tant du point de vue des techniques d'alignement que des vecteurs de mots : un des champs les plus ouverts et les plus prometteurs est la reconnaissance automatique de caractères des papyri, des stèles et autres sources documentaires. C'est la possibilité de comparer les auteurs antiques sans avoir à en passer par le référent de la langue moderne : il serait possible, grâce aux techniques que nous avons appliquées dans notre étude, de comparer les styles des auteurs antiques, les évolutions graphiques et grammaticales d'une époque à l'autre ou d'un lieu à l'autre. C'est enfin pour la littérature comparée la possibilité d'étudier l'histoire des langues, y compris des langues rares pour lesquelles les données sont peu nombreuses. Nous espérons que ce travail engagera nombre de ceux qui espèrent contribuer non seulement à la survie, mais au progrès des Humanités, à prendre en main des outils qui ont pour vocation de faire avancer la recherche sur des sentiers nouveaux.