

Intégration de la traduction automatique neuronale à la formation universitaire des futurs traducteurs : pistes d'exploration

LOÏC DE FARIA PIRES¹

Loic.DEFARIAPIRES@umons.ac.be

1. Introduction

Au cours des dernières décennies, les évolutions technologiques ont effectué une entrée très remarquée dans le monde de la traduction. Ces dernières ont intégralement remis en question l'apprentissage universitaire de la traduction et la pratique professionnelle, mais également la recherche en traductologie. Mémoires de traductions et bases terminologiques ont progressivement fait leur apparition dans la pratique professionnelle de la traduction ; elles ont plus tard été rejointes par la traduction automatique, ressource faisant régulièrement parler d'elle ces dernières années.

Si la qualité des approches de traduction automatique à base de règles ne permettait pas nécessairement une exploitation par post-édition dans tous les contextes professionnels, la décennie 2003-2013 a progressivement été marquée par le développement d'une technique à la qualité prometteuse, la traduction automatique statistique, comme l'illustrent notamment les travaux de Och (2003), Koehn (2010) et Hearne et Way (2010). Toutefois, la traduction automatique neuronale acquiert depuis 2013 un poids toujours plus considérable sur le marché de la traduction, comme en attestent, entre autres, les travaux de Bentivogli, Bisazza, Cettolo et Federico (2016), Toral et Sánchez-Cartagena (2017) et Koehn (2017), et remplace peu à peu la traduction automatique statistique dans de nombreux secteurs du marché. En outre, le développement de moteurs neuronaux de traduction automatique par de grandes institutions internationales telles que la Commission européenne, au sein de sa Direction générale de la traduction (Foti, 2017), ne laisse planer aucun doute : en cet outil réside une grande partie de l'avenir de la traduction professionnelle (exception faite de la traduction littéraire et autres traductions libres et créatives contenant de l'humour, de l'ironie et des références culturelles pour l'heure difficilement interprétables par une machine).

En ce sens, il convient plus que jamais de remettre en question les différentes formations universitaires disponibles en traduction, notamment au sein de l'Union européenne, dans le but d'harmoniser ces dernières et de permettre aux étudiants suivant ce type de cursus d'avoir, lors de l'obtention de leur diplôme, toutes les cartes en main pour entrer le plus sereinement possible sur le marché professionnel de la traduction, marché ô combien rude et compétitif.

Dans le présent article, nous nous attacherons en premier lieu à décrire brièvement le réseau EMT, sa raison d'être et ses principaux objectifs. Ensuite, nous décrirons les propositions d'intégration de la traduction automatique aux cursus universitaires déjà réalisées par d'autres auteurs. Nous détaillerons ensuite les avancées que représente la traduction automatique neuronale par rapport aux approches précédentes et proposerons des conseils d'intégration de ces nouvelles technologies aux formations universitaires.

¹ Titulaire d'un Master en traduction multidisciplinaire (2016) et d'une Agrégation de l'enseignement secondaire supérieur (2017). Doctorant au sein du service de Traduction spécialisée et Terminologie de l'Université de Mons depuis janvier 2017, sa recherche doctorale en cours a pour objet l'étude des effets de la post-édition de traduction automatique sur la qualité des textes institutionnels traduits à la Direction générale de la traduction de la Commission européenne, en fonction de différents groupes de traducteurs classés selon leur expérience professionnelle.

2. *European Master's in Translation* : un projet prometteur

2.1 Création et objectifs

La majorité des enseignants dispensant des cours dans le cadre des formations universitaires en traduction au sein de l'Union européenne, ainsi que bon nombre des étudiants concernés, auront probablement déjà entendu parler du projet EMT, pour *European Master's in Translation*, sans pour autant connaître de manière détaillée les tenants et aboutissants de ce projet.

L'EMT et son groupe d'experts ont vu le jour entre 2006 et 2007, sous l'impulsion de la Direction générale de la traduction de la Commission européenne (Groupe d'experts EMT, 2009 : 1). Son objectif principal est précisément de faire converger et d'optimiser les formations universitaires en traduction en Europe. Plusieurs facteurs ont été à l'origine de la création de ce projet : citons, entre autres, le développement des technologies qui fait évoluer rapidement les pratiques professionnelles et les critères de qualité, les élargissements de l'UE, l'absence de réglementation de la profession poussant à une recherche d'excellence, une volonté de valoriser les conditions de travail et les salaires des traducteurs, ainsi qu'une volonté de proposer un profil de qualité minimal et un référentiel de compétences communes pour les traducteurs diplômés (Groupe d'experts EMT, 2009 : 1).

Au vu de ces objectifs et de la volonté d'harmonisation émise par le groupe, les universités membres du réseau EMT représentent un formidable terrain d'étude pour la mise en place d'une formation harmonisée en traduction permettant aux étudiants d'entrer en contact, dans le cadre desdits cursus, avec les différentes technologies de la traduction et outils d'aide à la traduction, dont fait partie la traduction automatique. En considérant que la traduction automatique neuronale constituera probablement dans quelques années l'approche dominante, nous nous proposons d'énoncer quelques propositions pour une formation reprenant les technologies de la traduction, lesquelles seraient applicables au niveau des universités membres de l'EMT.

2.2 Le modèle de compétences EMT (versions de 2009 et 2017)

Nous nous permettons de prendre l'exemple de l'EMT dans le cadre de cet article en raison du fait que le réseau a élaboré, pour mener à bien les objectifs pour lesquels il a été créé, un modèle de compétences interdépendantes, reproduit ci-dessous. Le présent volet sera divisé en deux parties : le modèle élaboré en 2009 sera étudié, ainsi que sa mise à jour de 2017. Ensuite, une comparaison entre lesdits deux modèles sera effectuée.

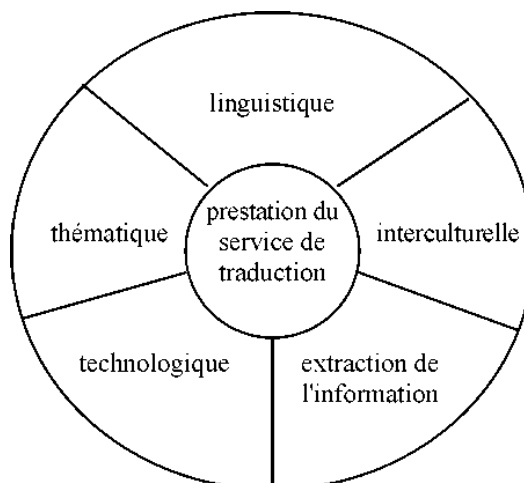


Figure 1. Modèle de compétences EMT 2009 (Groupe d'experts EMT, 2009 : 4).

La compétence technologique représente, comme l'indique le schéma, l'un des cinq piliers du modèle de compétences de l'EMT de 2009. Si la gestion terminologique et le maniement de base de données font l'objet d'une description détaillée dans le document explicatif du modèle, la seule référence à une aptitude liée à la traduction automatique est formulée en ces termes : « Connaître les possibilités et limites de la TA » (Groupe d'experts EMT, 2009 : 7). Certes, il s'agit là d'une aptitude importante, mais devant être à notre sens étoffée et accompagnée d'autres aptitudes liées à la traduction automatique (notamment neuronale) et à son utilisation dans un contexte professionnel.

En raison du caractère assez jeune du groupe EMT et de l'absence de directives claires concernant l'enseignement de l'utilisation de la traduction automatique et de la post-édition, force est de constater que chaque université incluait, selon le modèle de 2009, cet outil comme elle l'entendait dans le cadre de la formation proposée, et que certaines universités n'abordaient (et n'abordent encore) que très peu, voire pas, les principes sous-jacents de la traduction automatique et de la post-édition, bien que cette pratique représente actuellement une grande proportion des tâches de traduction effectuées en contexte professionnel. En ce sens, nous pensons que les universités du réseau EMT pourraient encore relever leurs critères d'excellence si elles bénéficiaient de propositions de lignes directrices (que fournit partiellement le modèle actualisé de 2017 présenté ci-après) à appliquer conjointement en matière de traduction automatique et de post-édition, dans le but de donner à l'ensemble des étudiants du réseau des chances équivalentes lors de leur entrée sur le marché de la traduction professionnelle.

Passons maintenant au modèle de compétences actualisé, proposé par l'EMT en 2017 et figurant ci-dessous, qui définit cinq domaines de compétences pour les cursus universitaires de traduction, à savoir 1) langue et culture, 2) traduction, 3) technologie, 4) compétence personnelle et interpersonnelle et 5) compétence en matière de services.

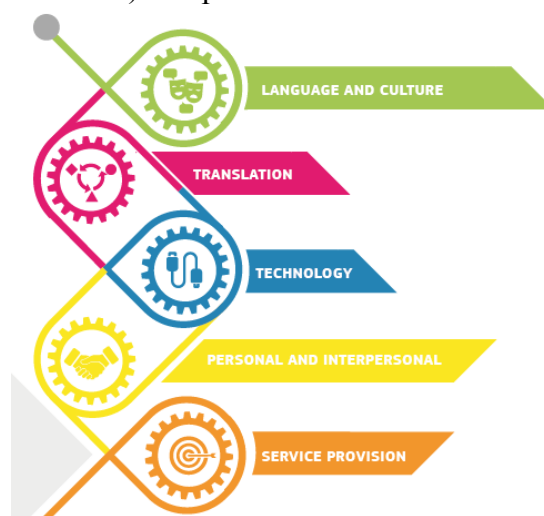


Figure 2. Modèle de compétences EMT 2017 (Groupe d'experts EMT, 2017 : 4).

La compétence technologique est, comme en 2009, présente au sein de ce modèle. Une mention relativement détaillée y est faite de l'inclusion de la traduction automatique dans la formation universitaire des futurs traducteurs. Cependant, la traduction automatique est d'ores et déjà mentionnée dans le cadre de la compétence de traduction, en ces termes : « *This section also recognises that the ability to interact with machine translation in the translation process is now an integral part of professional translation competence* » (Groupe d'experts EMT, 2017 : 7). La traduction automatique ne constitue dès lors plus l'apanage de la compétence technologique du modèle, mais revêt de l'importance dans la compétence de traduction, ce qui démontre son inclusion de plus en plus vaste dans les contextes professionnels de traduction, et

la nécessaire capacité des traducteurs d'utiliser au mieux cette technologie dans le cadre de leurs tâches. Ainsi, la sous-compétence 14 de la compétence de traduction indique que les cursus de traduction doivent former les étudiants à « *Apply post-editing to MT output using the appropriate post-editing levels and techniques according to the quality and productivity objectives, and recognise the importance of data ownership and data security issues* » (Groupe d'experts EMT, 2017 : 8). La sous-compétence 13 souligne par ailleurs l'importance de la préédition de textes en vue de les rendre plus adaptés à un traitement par TA.

Au sein de cette compétence technologique, les données suivantes sont à retenir : « *It also includes basic knowledge of machine translation technologies and the ability to implement machine translation according to potential needs* » (Groupe d'experts EMT, 2017 : 9). Plusieurs sous-compétences de la compétence technologique concernent la TA : il est, selon le modèle, nécessaire de prodiguer une formation aux bases de la TA et à ses répercussions sur le processus de traduction (sous-compétence 18), ainsi qu'à l'évaluation de la pertinence des systèmes de TA dans des contextes de traduction donnés et à l'application des modèles de TA pertinents dans les différents contextes (sous-compétence 19) (Groupe d'experts EMT, 2017 : 9).

Il est aisé de constater que la mise à jour du modèle proposé en 2017 tient, par rapport au modèle initial de 2009, davantage compte de la TA et de sa mise en application toujours plus répandue sur le marché de la traduction actuel. Les universités disposent désormais de davantage d'informations pouvant les aider à enseigner la TA et les éléments afférents, et le groupe d'experts continuera sans nul doute à travailler en collaboration avec les institutions d'enseignement supérieur pour une intégration toujours plus efficace de la TA aux différents cursus de traduction proposés par les universités membres du réseau.

En effet, l'intégration de la TA par les traducteurs professionnels se fait de plus en plus nécessaire pour faire face à la pression du marché actuel de la traduction et, étant donné que les étudiants en traduction d'aujourd'hui seront les professionnels du marché de demain, il apparaît fondamental de doter ces étudiants des armes nécessaires pour affronter l'avenir du marché, sur lequel la productivité requise n'aura de cesse de s'accroître. Ils maîtriseront ainsi des connaissances et des compétences leur permettant d'assister leurs collègues plus expérimentés mais non formés à la traduction automatique, et de disposer d'atouts valorisables lors de leur entrée sur le marché du travail.

3. Enseignement universitaire de la traduction automatique et de la post-édition : suggestions de différents auteurs

Des propositions très pertinentes d'inclusion de la post-édition dans les cursus universitaires ont déjà été proposées. Nous les décrirons brièvement avant de nous lancer dans la formulation de notre propre proposition qui, si elle reprend des éléments des formations décrites au préalable, tentera autant que faire se peut de mettre l'accent sur la post-édition de traduction automatique neuronale, qui représentera une grande partie du marché dans un avenir proche.

3.1 Proposition de Sharon O'Brien (Dublin City University)

Partons d'un fait : Krings et Koby (2001 : 12) ont souligné que les traducteurs devaient recevoir une formation avant de procéder à de la post-édition professionnelle. A notre sens, la proposition de cours de post-édition de traduction automatique la plus étoffée est celle de Sharon O'Brien (2002). Cette proposition de formation était alors divisée en deux volets : un premier volet théorique et un second volet pratique.

En premier lieu, la partie théorique proposée comprenait les modules suivants : une introduction à la post-édition, une introduction aux technologies de traduction automatique, une

introduction à la rédaction au moyen de systèmes de langues contrôlées, une initiation à la gestion terminologique avancée, des notions de linguistique textuelle avancée et une acquisition des compétences de base en programmation.

La partie pratique avait quant à elle pour objectif de permettre aux étudiants d'acquérir le facteur de « confort » en post-édition (étudiant suffisamment formé pour pratiquer la post-édition de manière aisée et efficace), tel que mentionné par Vasconcellos (1986 : 145). L'auteure indique d'abord que la post-édition doit être pratiquée aussi bien dans le cadre des cours qu'à domicile, pour permettre aux étudiants d'acquérir une quantité suffisante d'expérience avant d'entrer sur le marché du travail. Sharon O'Brien indique que les étudiants doivent s'entraîner sur différents types de textes, avec différents systèmes de traduction automatique, en pratiquant les différents types de post-édition (rapide/complète). Les étudiants devraient également apprendre à gérer les changements de paramètres des systèmes utilisés, la terminologie spécifique et les modifications des règles linguistiques utilisées. En outre, l'intégration concrète de la traduction automatique aux outils de traduction assistée par ordinateur devrait être abordée. Les compétences relatives à l'utilisation des langues contrôlées doivent aussi, selon l'auteure, être enseignées et mises en application, tout comme les aptitudes de composition de corpus et d'écriture de macros.

Comme le souligne l'auteure, il apparaît crucial de choisir le moment approprié du cursus pour y insérer un cours approfondi de post-édition de traduction automatique. Ainsi, elle considère que les étudiants doivent au préalable disposer de compétences fondamentales telles qu'une excellente maîtrise linguistique, une aptitude à la gestion terminologique de base, des compétences en informatique et en traduction spécialisée. En outre, ils devraient idéalement avoir suivi une introduction aux technologies du langage. En ce sens, il semble évident qu'un tel cours ne devrait pas être intégré aux premières années de la formation. Au regard des compétences requises, il est clair que ce cours devrait idéalement être donné en fin de parcours universitaire, c'est-à-dire en master, bien que l'auteure évoque la possibilité de le mettre en place dès la fin du *undergraduate programme*.

Cette proposition de cours de post-édition datant de 2002 nous apparaît très pertinente. Bien entendu, 15 ans plus tard, il conviendrait de l'adapter en exploitant tout aussi bien la TA statistique que la TA neuronale, dans le but de permettre aux étudiants de comprendre les avantages et inconvénients inhérents à chaque approche. En outre, l'inconvénient majeur que nous entrevoyons concernant la proposition ci-dessus serait le nombre considérable d'heures de cours requis pour prodiguer tous les enseignements mentionnés. Ainsi, nous considérons qu'il serait pertinent d'étaler le cours de post-édition sur deux années universitaires, tout en conservant une petite proportion de travail à domicile pour que les élèves s'exercent de manière autonome sur les outils proposés.

3.2 Autres propositions

Force est de constater que la proposition de O'Brien (2002) reste l'une des seules présentant un niveau de détail à ce point précis. Néanmoins, d'autres auteurs ont également évoqué la formation en post-édition dans leurs travaux académiques.

Prenons l'exemple d'Anne-Marie Robert (2013), formatrice de traducteurs ayant acquis de l'expérience dans l'enseignement des technologies de la traduction dans le cadre de différents masters en France. Cette dernière a mis en exergue la difficulté de mettre en pratique la post-édition dans un contexte universitaire : la traduction automatique devait, dans son cas précis, être associée à d'autres technologies de TAO. Or, selon l'auteure, l'« association intelligente » d'outils de TAO et de la traduction automatique doit être préparée en amont par des spécialistes dans des agences de traduction, sans quoi il s'avère techniquement impossible de mettre en place un tel cours (Robert, 2013 : 37). À défaut, l'auteure proposait lors des cours qu'elle dispensait des exercices de post-édition à partir de traductions automatiques générées

par des moteurs disponibles gratuitement sur le net, moteurs qu'elle considérait comme peu fiables et donc peu pertinents pour une formation professionnelle (Robert, 2013 : 38). Nous considérons qu'en 2018, la qualité de tels moteurs se trouve grandement améliorée par rapport à 2013 et que la post-édition est bien plus répandue en contexte professionnel qu'il y a quatre ans, ce qui rend l'idée d'un cours pratique de post-édition plus pertinente et, à notre sens, plus facilement applicable qu'il y a quelques années.

Citons également Ilse Depraetere (2010), qui a proposé quelques thèmes centraux à aborder. L'auteure insiste sur les différents niveaux de post-édition (rapide, minimale et complète), que les étudiants doivent être en mesure d'appliquer et de distinguer. Elle indique également que la mentalité des étudiants d'aujourd'hui (enclins à accepter des textes à la qualité imparfaite) diffère de celle des traducteurs expérimentés ayant accumulé de nombreuses années d'expérience sur le marché (cherchant davantage à s'approcher de la perfection) : il convient dès lors de tenir compte de cette différence de mentalité lors de la mise en place d'une formation en post-édition. Il est également nécessaire, selon l'auteure, de développer l'esprit critique des étudiants à l'égard de la traduction automatique, de les inciter à se référer au texte source pour remarquer d'éventuelles additions ou omissions d'informations, et de les sensibiliser aux erreurs récurrentes de traduction automatique. En outre, il est intéressant de constater que l'auteure fait également référence à la proposition de O'Brien (2002) tout en doutant de la faisabilité temporelle de cette proposition, au demeurant très convaincante.

Les deux auteures proposent des éléments très pertinents et soulèvent des limites dont il convient de tenir compte dans le cadre de la mise en place d'un cours universitaire de post-édition de traduction automatique. Toutes les pistes évoquées se doivent d'être exploitées dans le cadre de notre proposition.

4. TA neuronale : quelles avancées par rapport aux approches précédentes ?

Pour bien comprendre les enjeux croissants que représente la post-édition de traduction automatique, il apparaît utile de dresser un bref historique de l'évolution des approches en la matière. Dans un premier temps, les approches au mot à mot ayant vu le jour à partir des années 1950 (à la suite du mémorandum sur la traduction automatique de Warren Weaver de 1949) n'ont pas fourni de résultats concluants et ont débouché sur le rapport ALPAC de 1966 (Guidère, 2008 : 148) mettant fin aux subventions gouvernementales des États-Unis à l'égard de la traduction automatique.

Cependant, les années 1980 ont vu le renouveau de la recherche en traduction automatique (Guerberof Arenas, 2008 : 1), avec l'apparition, notamment, de la traduction automatique à base de règles, se basant pour sa part sur l'application d'un transfert syntaxique entre la langue de départ et la langue d'arrivée. Certes, l'impressionnant volume requis des dictionnaires d'équivalents sémantiques utilisés dans le cadre de ces approches à base de règles a rapidement montré leurs limites, mais force a été de constater que les transferts à base de règles ont fourni les premiers résultats intéressants et potentiellement exploitables dans le monde de la traduction automatique.

C'est à partir des années 2000 que la traduction automatique a commencé à se montrer plus largement exploitable en contexte professionnel. Ainsi, l'apparition de la traduction automatique statistique (voir notamment Och, 2003) a contribué au développement de la discipline. Les approches statistiques sont fondées sur l'exploitation de corpus bilingues aussi volumineux que possible (en vue de créer un modèle de traduction) et de corpus monolingues (en vue de créer un modèle de langue visant à assurer la correction linguistique en langue d'arrivée) pour mener à bien le processus de traduction via une probabilisation d'équivalences (calcul, en fonction des données de base introduites dans le moteur, des solutions de traduction les plus probables). Il convient d'admettre qu'en dépit de leurs qualités (réduction drastique des problèmes de sens, adaptabilité au contexte de traduction en fonction des corpus utilisés pour

l'entraînement du moteur...), les approches statistiques supposent encore un certain nombre de lacunes (grandes quantités de données d'entraînement nécessaires, fluidité aléatoire en langue d'arrivée, durée d'entraînement...), bien qu'elles continuent, aujourd'hui encore, à servir de base à une post-édition professionnelle dans certains contextes, comme c'est notamment le cas à la Direction générale de la traduction de la Commission européenne.

La dernière révolution en date dans le monde en perpétuelle évolution de la traduction automatique provient de la traduction automatique neuronale, ayant vu le jour entre 2013 et 2014 (voir notamment Kalchbrenner & Blunsom, 2013 ; Cho *et al.*, 2014). Cette fois, la machine est programmée pour simuler l'action du cerveau humain de manière informatique, ni plus ni moins. Pour ce faire, le texte source à traduire est introduit dans un réseau neuronal informatique comportant plusieurs couches, ayant pour but d'encoder les unités de traduction sous la forme de vecteurs numériques représentant des concepts (Toral et Sanchez-Cartagena, 2017 : 1). Des données d'entraînement sont également utilisées lors de l'élaboration du moteur, mais son fonctionnement va bien au-delà d'une probabilisation d'équivalences comme tel était le cas en traduction automatique statistique. En effet, les moteurs neuronaux sont dotés d'une couche algorithmique dite cachée (*hidden layer*) ayant des capacités de réflexion propres sur lesquelles les humains, y compris les concepteurs mêmes du moteur, n'ont aucune emprise. D'où une problématique majeure liée à cette approche : la difficulté d'introduire des corrections pour orienter le moteur dans la bonne direction en cas d'erreur de sa part. Cependant, les approches neuronales disposent de nombreuses qualités, dont les principales sont la qualité du texte d'arrivée, bien meilleure (dans des conditions idéales) que dans le cadre des approches statistiques, ainsi qu'une excellente fluidité en langue d'arrivée (Koehn, 2018). Néanmoins, les approches neuronales ne sont guère parfaites : les résultats laissent à désirer lorsque les données utilisées pour entraîner le moteur ne correspondent pas au domaine des textes traduits, les phrases courtes sont souvent à l'origine de mauvaises traductions et, surtout, le bel « emballage » stylistique et syntaxique des solutions fournies par la traduction automatique neuronale peut mener le post-éditeur à passer à côté d'erreurs de sens évidentes et potentiellement graves, qui existent toujours en dépit du caractère sophistiqué des approches neuronales (Koehn, 2018).

En ce sens, la formation universitaire des futurs traducteurs devrait idéalement tenir compte de l'évolution de la typologie d'erreurs entre les différentes approches. Ainsi, il serait prudent de sensibiliser les étudiants aux problèmes de sens pouvant survenir lors de la post-édition de traduction automatique neuronale, en dépit de l'élégance superficielle en langue d'arrivée, et des problèmes syntaxiques pouvant survenir en traduction automatique statistique, car de nombreuses entreprises de traduction utilisent aujourd'hui des approches hybrides, partiellement statistiques et partiellement neuronales, en vue de tirer le meilleur des deux procédés (Koehn, 2018).

5. Proposition d'intégration de la traduction automatique neuronale aux formations universitaires en traduction

Les propositions décrites ci-dessus, bien qu'elles conservent un caractère très pertinent à l'heure de la traduction automatique neuronale, doivent à notre sens être approfondies et adaptées pour intégrer au mieux cette technologie dans une éventuelle formation à la post-édition dans le cadre des études universitaires en traduction, et notamment au sein des universités membres du réseau EMT. Il convient, dès lors, de poursuivre la réflexion initiée par les différents auteurs en y incluant au mieux la traduction automatique, en particulier neuronale.

5.1 Conditions préalables

Tout comme cela a été dit dans les travaux cités précédemment, nous estimons qu'un certain nombre de compétences préalables sont nécessaires.

Dans la lignée de De Almeida et O'Brien (2010 : 7), nous considérons tout d'abord qu'un bon post-éditeur se doit avant tout d'être un bon traducteur. En ce sens, il convient que tout étudiant tenu de suivre des cours de post-édition soit au préalable familiarisé avec les aspects les plus fondamentaux de la formation en traduction dans son ensemble.

Le premier critère serait une compréhension parfaite de la langue source. Certes, une telle condition relève du bon sens, mais il apparaît fondamental de rappeler que la base de toute traduction (et, par extension, post-édition) correcte repose sur la compréhension de l'intégralité des informations présentes dans le texte source, en vue de rendre fidèlement le sens du texte source.

En lien avec ce premier critère, le second serait une maîtrise complète de la langue cible. En effet, il est également primordial, avant d'aborder l'apprentissage de la post-édition, que les étudiants concernés soient parfaitement capables de s'exprimer avec aisance, élégance et sans erreur en langue cible, et ce dans le but de garantir la fluidité des textes traduits ou post-édités (notamment pour certaines catégories de post-édition, sur lesquelles nous reviendrons plus tard).

En outre, comme le propose Sharon O'Brien (2002 : 103), les étudiants devraient à ce stade déjà disposer de solides compétences en traduction spécialisée, dans la mesure où les textes qu'ils post-éditeront par la suite pourront relever de domaines divers et variés, ce qui rend fondamentaux certains éléments tels que la gestion terminologique et la maîtrise de connaissances aussi approfondies que possible dans de nombreux domaines.

De plus, il semble fondamental que les étudiants maîtrisent, avant de suivre des cours de post-édition, les différents aspects de la qualité en traduction humaine. Ainsi, des connaissances relatives aux notions d'équivalence, de congruence, de fluidité, de style, d'adaptation à la culture cible et de respect de la fonction des textes devraient être autant de cordes à l'arc des étudiants avant d'aborder la post-édition.

Ensuite, les étudiants devraient également connaître le b.a.-ba des outils d'aide à la traduction (ou traduction assistée par ordinateur). La post-édition étant un élément particulier de la vaste gamme des outils de TAO existants, il semble fondamental de pouvoir la distinguer des autres ressources disponibles avant d'apprendre son fonctionnement particulier.

Finalement, la réalité du marché est un contexte dont les étudiants doivent être conscients. En effet, c'est bel et bien ce contexte précis qui se trouve à l'origine de l'explosion de l'utilisation de la post-édition et du développement d'approches de traduction automatique toujours plus innovantes. La productivité est le maître mot régnant sur la majorité des contextes de traduction professionnelle, et, pour cette raison, les étudiants devraient plus que jamais avoir conscience de la réalité professionnelle qui les attend et des raisons pour lesquelles un cours de post-édition leur est dispensé dans le cadre de leur formation en traduction.

5.2 Durée d'enseignement et stade de la formation

Au vu des compétences préalables qui précèdent, nous rejoignons Sharon O'Brien (2002 : 105) lorsqu'elle affirme que la formation de post-édition devrait avoir lieu vers la fin du parcours universitaire des étudiants. De manière à pouvoir organiser sur deux ans cette formation à la post-édition, nous proposons de la mettre en place dans le cadre des deux dernières années de formation, c'est-à-dire de la réserver au master dans le cas des universités dont le premier cycle affiche une durée de trois ans et le deuxième une durée de deux ans, et de la débiter en dernière année de premier cycle dans le cas d'un premier cycle d'une durée de quatre ans et d'un deuxième cycle d'une durée d'un an.

Ainsi, les étudiants disposeraient d'une solide formation à la post-édition, surtout si l'on y ajoute les différents travaux pratiques pouvant être réalisés en dehors des heures de cours (au moyen des laboratoires informatiques au sein des universités, voire de l'ordinateur personnel des étudiants), ainsi que la possibilité d'inclure un volet de post-édition au stage d'insertion professionnelle effectué dans le cadre de la dernière année de master.

5.3 Compétences à inclure

Tentons de répondre à la question suivante : comment mettre en œuvre l'enseignement de la post-édition de traduction automatique neuronale dans les formations en traduction ? En guise de remarque préalable, nous avons conscience du fait que les moteurs de traduction automatique neuronale ne sont à l'heure actuelle pas encore aussi répandus que ceux de traduction automatique statistique. Cependant, leur progression fulgurante et la multiplication du nombre de systèmes disponibles nous laissent penser que cette approche de traduction automatique pourra bientôt se trouver au centre des enseignements prodigués en matière de post-édition. Bien entendu, les points mentionnés ci-dessous n'ont aucune prétention d'exhaustivité. Aussi, il convient de les envisager comme quelques pistes plutôt que comme une vérité complète, absolue et indiscutable.

Le cours serait idéalement divisé en trois volets principaux : un premier volet serait entièrement consacré à la traduction automatique en soi. Celui-ci débiterait par un parcours historique où seraient abordées l'évolution de la traduction automatique et les différentes approches élaborées au fil des années. L'approche neuronale serait enseignée : de sa genèse aux principes sous-jacents régissant son fonctionnement, cette approche pleine d'avenir serait appréhendée dans ses moindres détails, au travers des notions d'algorithme, d'intelligence artificielle, de réseaux neuronaux, d'encodeur et décodeur, de couche cachée, de mécanisme d'attention et de représentation vectorielle (liste non exhaustive). Les principes fondamentaux de la constitution des corpus nécessaires à l'entraînement des moteurs seraient également abordés. L'approche neuronale serait ensuite comparée aux approches statistiques : des exercices pratiques pourraient être réalisés dans ce contexte pour dégager les forces et faiblesses de chaque système et pour les comparer l'un par rapport à l'autre sur des mêmes textes. Ensuite, différents systèmes neuronaux pourraient être au centre de travaux pratiques, avec pour but de permettre aux étudiants de comparer ces systèmes et apprendre à reconnaître une sortie de traduction automatique brute de bonne qualité d'une sortie de mauvaise qualité.

Après cette découverte et cette comparaison des différents systèmes de traduction automatique, l'enseignant pourrait entrer dans le vif du sujet en débutant le deuxième volet de la formation, celui-ci centré sur la discipline de post-édition de traduction automatique neuronale. Dans le cadre de cette partie du cours, les divers niveaux possibles de post-édition, c'est-à-dire la PE rapide, la PE minimale et la PE complète (voir entre autres Depraetere, 2010 : 1) seraient enseignés. Les différents critères de qualité à respecter seraient enseignés de manière théorique, et des exercices pratiques seraient ensuite réalisés par les étudiants (décision du niveau de post-édition requis en fonction des objectifs des textes, application des différents niveaux de post-édition pour un même texte). En parallèle, la notion de productivité serait abordée, dans le but de permettre aux étudiants de saisir l'importance de la maîtrise de compétences de post-édition sur le marché actuel. Les notions de public cible et de fonction des textes occuperaient également une position prépondérante au sein de ce volet de la formation ; les étudiants devant être capables tantôt de s'adapter à un donneur d'ordre exigeant une post-édition de haute volée, tantôt de post-éditer le plus vite possible pour une simple compréhension de l'idée principale du texte.

Finalement, le dernier volet de la formation pourrait englober les dimensions humaines et ergonomiques de la pratique de la post-édition de traduction automatique. En ce qui concerne ces composantes, les aspects plutôt pratiques du travail seraient abordés. Il serait question des

droits et devoirs des post-éditeurs, des tarifs en vigueur sur le marché, des moyens pouvant être employés pour éviter le sentiment de frustration parfois lié à cette discipline, des questions de confidentialité liées à la pratique de la post-édition et du traitement en ligne des textes par traduction automatique, mais aussi de la gestion des données, des notions de *big data*, *cloud*, de travail en collaboration sur un même projet de post-édition... Cette liste est, encore une fois, loin d'être exhaustive, tant les questions abordées dans le cadre d'une formation en post-édition peuvent être variées et diverses en fonction des différents contextes.

5.4 Modalités concrètes de mise en œuvre

Il convient également de se pencher sur les modalités pratiques de mise en œuvre d'une formation de post-édition de traduction automatique neuronale. En effet, un certain nombre de problématiques à résoudre se présentent, et il sera nécessaire d'y répondre aussi efficacement que possible.

En premier lieu, des enseignants disposant d'une solide formation en technologies de la traduction devront être recrutés dans les universités où aucun spécialiste du domaine n'exerce de fonctions. Autre possibilité : des professeurs volontaires de ces mêmes universités pourraient suivre des formations en technologies de la traduction pour se mettre à niveau et se trouver en mesure de dispenser de tels cours de post-édition.

En outre, si une formation en post-édition venait à être mise en œuvre de manière harmonisée dans les universités du réseau EMT, des réunions auxquelles participerait un représentant EMT provenant de chaque université pourraient être organisées avec pour thème central l'enseignement de la post-édition de traduction automatique. Au sein de cet espace d'échanges et de réflexions, les responsables pourraient partager leurs bonnes pratiques (réussites, difficultés rencontrées, suggestions d'amélioration) tout en œuvrant à un dialogue constructif permettant aux différentes universités d'améliorer leurs pratiques et d'apprendre les unes des autres.

En ce qui concerne les cours à proprement parler, des projets concrets de post-édition par groupes pourraient être envisagés, et pourquoi pas une simulation de contextes professionnels de post-édition (comme tel est déjà le cas avec la traduction dans les universités organisant des *skills labs*). En outre, dans le cadre des cours, des intervenants externes pourraient venir parler aux étudiants de leur expérience professionnelle et de la réalité du marché, pour permettre aux étudiants d'entrer avant l'heure en contact avec des acteurs du monde professionnel de la traduction. Finalement, il conviendrait, dans le cadre du cours de post-édition, de tenir compte des demandes et craintes des étudiants concernant l'influence de ces nouvelles technologies sur leur future profession.

Pour clôturer cette partie, un mot doit également être dit sur l'évaluation du cours. A notre sens, la solution semblant la plus appropriée serait celle d'une évaluation comportant une petite partie théorique, mais surtout une partie principale pratique qui consisterait en une mise en situation professionnelle d'utilisation de traduction automatique neuronale à post-éditer, selon certaines modalités données (destination du texte, public cible, fonction principale)... Le problème majeur de l'évaluation serait, comme dans toute évaluation académique de traduction, la dimension subjective inévitable qui y subsisterait. En ce sens, l'enseignant devrait déterminer la manière la plus juste et la plus objective d'évaluer la mise en application des contenus abordés durant l'année universitaire. Et pourquoi ne pas pousser cette logique un peu plus loin et imaginer une certification EMT en post-édition commune à toutes les universités, test élaboré conjointement par les acteurs du réseau qui permettrait aux étudiants le réussissant d'obtenir une certification supplémentaire à inclure dans leur curriculum vitae ? Ceci représenterait probablement un véritable gage de qualité pour les employeurs potentiels.

6. Conclusion

Nous avons présenté ici les différents outils d'aide à la traduction et retracé brièvement l'histoire de la traduction automatique, pour en arriver aux moteurs neuronaux qui sont de plus en plus populaires. Ensuite, nous avons présenté le réseau EMT, *European Master's in Translation*, décrit le rôle qu'il pourrait jouer dans l'harmonisation des formations en matière de post-édition et présenté le modèle de compétences proposé car ce réseau, modèle au sein duquel les technologies de la traduction occupent une place importante. Nous avons continué en décrivant les propositions existantes de cours de post-édition, et notamment celle de Sharon O'Brien, de Dublin City University. Nous avons enfin élaboré des pistes de mise en œuvre d'un cours harmonisé de post-édition : nous avons en ce sens détaillé les compétences préalables nécessaires, proposé une durée et un moment précis de mise en œuvre au sein des différents cursus de traduction existants, envisagé les compétences à inclure à une telle formation en post-édition et présenté des modalités pratiques de mise en œuvre, telles que des pistes d'évaluation des acquis et différentes réunions et activités possibles.

Pour conclure, nous pensons que les formations en traduction auraient tout à gagner d'une inclusion systématique d'un cours de post-édition harmonisé et adapté aux évolutions technologiques en matière de traduction automatique. Nous pensons en ce sens que, si un tel cours était mis en place dès à présent, il devrait idéalement s'articuler autour de l'approche neuronale occupant le devant de la scène, et qui sera amenée à évoluer et à se perfectionner davantage. Grâce à un tel cours, les traducteurs en formation pourraient disposer des armes nécessaires pour affronter le marché de la traduction tel qu'il est réellement, un marché compétitif où la notion de productivité occupe le devant de la scène. En effet, nous considérons que la profession de traducteur, s'il est improbable qu'elle disparaisse purement et simplement, sera amenée à subir de nombreux changements au cours des années à venir et à devenir une profession de post-éditeur, sauf dans de rares cas. Il apparaît essentiel que les formations universitaires tiennent compte d'un tel changement sur le marché et s'y adaptent au mieux.

Enfin, l'ambitieux et prometteur réseau EMT incarne à la perfection, selon nous, cette volonté d'harmonie et de partage ; il serait donc très pertinent de créer une formation en post-édition dans le cadre d'un tel réseau qui, en outre, inclut la compétence technologique au modèle de compétences que ses responsables ont élaboré, modèle valide pour la majorité des formations universitaires européennes en traduction.

Références

- BENTIVOGLI, Luisa, BISAZZA, Arianna, CETTOLO, Mauro and FEDERICO, Marcello (2016), «Neural versus Phrase-Based Machine Translation Quality: a Case Study», *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, [en ligne] disponible sur <https://arxiv.org/abs/1608.04631> [consulté le 14 septembre 2017].
- CHO, Kyunghyun, VAN MERRIËNBOER, Bart, BAHDANAU, Dzmitry and BENGIO, Yoshua (2014), «On the Properties of Neural Machine Translation: Encoder-Decoder Approaches», *Eight Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8)*, [en ligne] disponible sur <https://www.aclweb.org/anthology/W14-4012> [consulté le 23 avril 2018].
- DE ALMEIDA, Giselle and O'BRIEN, Sharon (2010), «Analysing Post-Editing Performance: Correlations with Years of Translation Experience», *Proceedings of the 14th Annual Conference of the European Association for Machine Translation*, [en ligne] disponible sur <https://www.semanticscholar.org/paper/Analysing-Post-Editing-Performance-Correlations-wi-Almeida-O-Brien/890ec4a0b8a9d7b0f5e7bda3f7467f2cfec0cd4f> [consulté le 8 avril 2017].
- DEPRAETERE, Ilse (2010), «What counts as useful advice in a university post-editing training context? Report on a case study.», *14th Annual EAMT Conference*, [en ligne] disponible sur www.mt-archive.info/EAMT-2010-Depraetere-2.pdf [consulté le 12 octobre 2017].
- FOTI, Markus (2017), *MT@EC and eTranslation*. Communication présentée à la Direction générale de la traduction de la Commission européenne le 9 février 2017, Bruxelles.

GROUPE D'EXPERTS EMT (2009), « Compétences pour les traducteurs professionnels, experts en communication multilingue et multimédia ». Document élaboré en janvier 2009 en vue de la conférence *EMT 2009* (16-17 mars 2009), Bruxelles.

GROUPE D'EXPERTS EMT (2017), « Competence framework 2017 ». European Commission, Brussels.

GUERBEROF ARENAS, Ana (2009), «Productivity and quality in the post-editing of outputs from translation memories and machine translation», *Localisation focus – The International Journal of Localisation*, 7, pp. 11-21, [en ligne] disponible sur https://www.researchgate.net/publication/320467281_Productivity_and_quality_in_the_post-editing_of_outputs_from_translation_memories_and_machine_translation [consulté le 23 avril 2018].

GUIDERE, Mathieu (2008), *Introduction à la traductologie – Penser la traduction: hier, aujourd'hui, demain*, Bruxelles, De Boeck.

HEARNE, Mary and WAY, Andy (2011), «Statistical Machine Translation: A Guide for Linguists and Translators», *Language and Linguistic Compass*, pp. 1-21, [en ligne] disponible sur www.computing.dcu.ie/~away/CA446/SMTforLinguists.pdf [consulté le 8 mai 2017].

KALCHBRENNER, Nal and BLUNSOM, Phil (2013), «Two recurrent continuous translation models», *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1700-1709, [en ligne] disponible sur www.aclweb.org/anthology/D13-1176 [consulté le 23 avril 2018].

KOEHN, Phillip (2010), *Statistical Machine Translation*, Cambridge, Cambridge University Press.

KOEHN, Phillip (2017), *What is Deep Neural Machine Translation and how can it be applied?*, Omniscien Technologies, [en ligne] disponible sur <https://omniscien.com/more/resources/webinar/#-the-state-of-neural-machine-translation--nmt---technology-webinar-series> [consulté le 15 novembre 2017].

KOEHN, Phillip (2018), *Research in Translation – What Is Exciting and Shows Promise Ahead*, Omniscien Technologies, [en ligne] disponible sur <https://omniscien.com/symposium/day3/> [consulté le 23 avril 2018].

KRINGS, Hans P. and KOPY, Geoffrey S. (eds) (2001), *Repairing Texts: Empirical Investigations of Machine-Translation Post-Editing Processes*, Kent, Kent State University Press.

O'BRIEN, Sharon (2002), «Teaching Post-editing: A Proposal for Course Content», *6th EAMT Workshop – Teaching Machine Translation*, pp. 99-106, [en ligne] disponible sur http://www.academia.edu/1160352/Teaching_Post-Editing_A_Proposal_for_Course_Content [consulté le 23 novembre 2017].

OCH, Franz Josef (2003), «Minimum Error Rate Training in Statistical Machine Translation», *41st Annual Meeting on Association for Computational Linguistics*, 1, pp. 160-167, [en ligne] disponible sur <https://dl.acm.org/citation.cfm?id=1075117> [consulté le 18 octobre 2017].

ROBERT, Anne-Marie (2013), «Vous avez dit post-éditrice ? Quelques éléments d'un parcours personnel», *The Journal of Specialised Translation*, 19, pp. 29-39, [en ligne] disponible sur www.jostrans.org/issue19/art_robert.pdf [consulté le 6 juillet 2017].

TORAL, Antonio and SANCHEZ-CARTAGENA, Victor M. (2017), «A Multifaceted Evaluation of Neural versus Phrase-Based Machine Translation for 9 Language Directions», *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, [en ligne] disponible sur <https://arxiv.org/abs/1701.02901> [consulté le 6 octobre 2017].

VASCONCELLOS, Muriel (1986), «Post-editing On-screen: Machine Translation from Spanish into English», *Proceedings of Translating and the Computer*, 8.

WEAVER, Warren (1949), *Memorandum on Machine Translation*, New York, The Rockefeller Foundation.