

# Traduction humaine et traduction automatique : entre erreur et errance

VALERIA FRANZELLI  
Université de Brescia

## INTRODUCTION

La Toile offre aujourd'hui de nombreux outils de traduction, qui sont de plus en plus employés par les entreprises et les particuliers, surtout pour le transfert vers la langue maternelle, ce qui limite le recours à des professionnels.

La recherche traductologique s'intéresse à la possibilité d'employer ces outils pour améliorer d'un point de vue qualitatif et quantitatif le travail du traducteur, le processus de traduction et ses résultats (cf. Arnold 2003 ; Forcada 2010 ; Gambier 2011). Nous présentons ici quelques réflexions issues d'un projet de recherche de la durée de deux ans, conçu par M. Charles Barone de l'Université de Pise, visant à élaborer une approche linguistique à la traduction automatique (TA).

Dans ce projet, notre contribution a concerné tout spécialement les erreurs humaines : par l'analyse des traductions pédagogiques, exercices de traduction faits en cours d'apprentissage (Ladmiral 1972), nous avons essayé de comprendre quelles sont les unités problématiques qui amènent à l'erreur et quels sont les principaux facteurs qui la déterminent. On définit ici comme traductions pédagogiques les exercices de traduction faits dans des cours de version (traduction du français à l'italien) suivant une approche plus contrastive que traductologique : l'attention des apprenants est portée sur la reconnaissance des phénomènes linguistiques et culturels qui caractérisent le texte de départ et l'étude des

correspondances possibles entre les langues et les cultures impliquées pour les transmettre ensuite dans un texte d'arrivée. Ainsi, la compétence linguistique passe avant tout facteur lié à la pratique professionnelle, comme par exemple les nécessités du donneur d'ouvrage, l'identité du rédacteur du texte source et du récepteur du texte cible (son contexte socioculturel, ses attentes, sa sensibilité ou sa vision du monde) et l'utilisation finale du texte cible, qui ont un rôle fondamental dans les choix traductifs opérés par les professionnels (cf. Nord 2008 ; Reiss 2009). Nous avons choisi ces exercices de version car ils nous paraissent néanmoins révélateurs de la structuration et de l'évolution des processus de transfert qui sont à l'origine de la traduction professionnelle. En outre, les travaux des apprenants présentent un tel degré d'imperfection qu'il est possible de faire des comparaisons fructueuses avec le travail de la machine.

Nous allons ici montrer que les performances des uns et des autres fournissent des instruments utiles pour repenser aussi bien le travail humain que celui de la machine.

## 1. LE CLASSEMENT DES ERREURS HUMAINES

Notre réflexion tire profit d'un corpus constitué de traductions du français vers l'italien faites par des étudiants italophones de FLE de l'Université de Pise, dont les niveaux de connaissance vont du A1 au C1 du CECRL (Cadre Européen Commun de Référence pour les Langues)<sup>1</sup> (cf. Chauvet 2008). Il s'agit d'au moins 600 travaux bien variés : pour chaque niveau de langue nous avons essayé de tester l'influence de certaines variables, qui ne se limitent pas à la complexité linguistique, mais qui incluent aussi les typologies textuelles et les conditions de travail (cf. Hansen 2010). Concernant tout spécialement les conditions de travail, un document qui a été traduit comme épreuve d'examen a été ensuite proposé à d'autres étudiants comme exercice de traduction, à faire soit en classe en deux heures, soit à la maison, tapé à l'ordinateur, pour le lendemain.

Cette variété offre la possibilité de mieux cerner les facteurs déterminant les erreurs et aide en particulier à les classer : quoiqu'il soit assez aisé de repérer des erreurs, leur classement en catégories demande une réflexion approfondie.

Le travail de définition des catégories d'erreur a été effectué à partir de 145 traductions du français vers l'italien, dont 102 faites à la maison d'un jour à l'autre par des étudiants de Master en Langues, Littératures et Civilisations Étrangères (LLCE) de l'Université de Pise, et 43 correspondant aux épreuves d'examen soutenues au cours d'un an par des étudiants de troisième année de Licence en LLCE.

---

<sup>1</sup> Les niveaux A1 et A2 correspondent en gros aux compétences prévues en première année de Licence de Langues, Littératures et Civilisations Étrangères (LLCE) en Italie, le niveau B1 pour la deuxième année et le B2 pour la troisième ; le niveau C1 englobe par contre les compétences du Master LLCE.

Afin de procéder dans notre classement, nous avons tout d'abord exploité la distinction entre « faute » et « erreur », permettant de diviser en deux groupes les phénomènes repérés dans les travaux des étudiants : d'un côté les « erreurs », ou phénomènes systématiques, dictées par un manque de compétences morphosyntaxiques, stylistiques et lexicales ; de l'autre les « fautes », ou erreurs non systématiques, qui sont donc involontaires, telles que les coquilles et souvent les omissions, que l'auteur pourrait corriger de manière autonome (Corder 1980 ; Collombat 2009).

Les conditions de travail ont été les principaux facteurs considérés pour cette première distinction : les travaux tapés à l'ordinateur et à la maison présentent en effet des coquilles (de l'inversion des lettres à l'absence d'espaces de ponctuation) que les épreuves d'examen, écrites à la main et en classe, ne montrent pas ; ces dernières ont par contre des omissions (de mots ou même de phrases) qui sont en général absentes dans les premiers, ceci à cause sans doute d'un manque de temps pour la relecture finale.

Cela dit, la fréquence de manifestation de chaque phénomène dans différentes traductions du même texte oriente la distinction entre erreur et faute, notamment au niveau des omissions qui peuvent être dues à un manque de compétences souvent d'ordre lexical. Lisons des exemples :

- (1) *Ma mère s'était absentée : pas un sourire, pas un signe de connivence.* (texte de départ)<sup>2</sup>
- (1a) *Mia madre si era assentata: [omission] non un segno di connivenza.* (traduction d'étudiant)
- (1c) *Mia madre si era assentata: nessun sorriso, nessun segno di connivenza* (notre traduction)
  
- (2) *Or, si Louis XVI a hérité* (texte de départ)<sup>3</sup>
- (2a) *[Omission] se Luigi XVI ha ereditato* (traduction d'étudiant)
- (2b) *O se Luigi XVI ha ereditato* (traduction d'étudiant)
- (2c) *Orbene, se Luigi XVI ha ereditato* (notre traduction)

Certaines omissions qui se présentent dans une seule copie (1a), bien que rédigée à la maison, peuvent être classées comme des fautes car elles sont manifestement causées par une relecture hâtive du travail : l'unité de traduction « pas un sourire » pour un étudiant italoophone de troisième année de Licence n'est certainement pas problématique. Par contre, sachant que le phénomène de l'omission (2a) se présente dans trois copies sur 19 et que l'erreur de traduction (2b) a été relevée dans une quatrième copie, nous ne parlerons pas de faute, mais d'erreur,

<sup>2</sup> J.-P. Sartre, *Les mots*, 1964. Niveau B2.

<sup>3</sup> F. Furet, M. Ozouf, *Dictionnaire critique de la Révolution française*, Flammarion, 1988, pp. 271-276. Niveau C1.

d'autant plus que le travail a été fait à la maison avec la possibilité de consulter toute ressource papier ou en ligne pour l'éviter. L'unité de traduction « or » devient ainsi problématique car elle génère des erreurs dans 20% des copies examinées et dans des conditions de travail favorables.

Lors de la définition des catégories de classement des erreurs, une autre question assez ardue a été abordée : la maîtrise de la langue d'arrivée, l'italien. Nous avons décidé de classer les phénomènes relevés non seulement en fonction de l'unité de traduction qui était à leur origine (unités grammaticale, stylistique ou lexicale), mais aussi en fonction de la langue. Il a été en effet utile de comprendre si certaines erreurs provenaient d'un manque de compétence de l'étudiant en italien ou bien en français, par exemple dans la traduction des variations : l'apprenant ne sait-il pas reproduire la variation en italien ou bien ne l'a-t-il pas reconnue en français ? Nous avons constaté, en outre, que puisque les étudiants maîtrisent la langue d'arrivée, ils peuvent aisément reformuler les énoncés et que pour cela il devient difficile d'associer certaines erreurs à une unité précise du texte de départ. La catégorie « erreur lexicale » a été donc réajustée en « erreur lexico-sémantique française » : l'étudiant n'a pas compris le sens général de l'énoncé, soit à cause de sa formulation soit à cause d'un lexème en particulier contenu dans le texte de départ. Nous en verrons des exemples dans les pages qui suivent.

Grâce à la comparaison des erreurs repérées dans notre premier corpus, nous avons néanmoins pu structurer une grille de classement définitive (Tab. 1) présentant les étiquettes suivantes :

- *texte de départ*, le segment de la version originale qui contient la source de l'erreur ;
- *grammaire française*, catégorie qui regroupe les erreurs dictées par un manque de compétences grammaticales en langue de départ (temps verbaux, prépositions, adverbes, accents grammaticaux, syntaxe, etc.) ;
- *grammaire italienne*, pour les erreurs dues à des lacunes grammaticales en langue d'arrivée (temps verbaux, prépositions, adverbes, orthographe, syntaxe, etc.), mais aussi pour des mots ou des expressions qui n'ont aucun sens en italien ;
- *lexico-sémantique française*, qui concerne toutes les erreurs relevant d'une interprétation erronée du sens d'un mot, d'une locution ou d'une phrase en langue de départ ;
- *stylistique italienne*, où l'on inclut toutes les erreurs de connotation, registre et ponctuation en langue d'arrivée ;
- *fautes*, pour toute omission ou coquille non intentionnelles que l'auteur pourrait corriger de manière autonome par une relecture attentive de son travail ;
- *fréquence*, qui signale le degré de difficulté d'une unité de traduction donnée, en fonction du nombre d'occurrences dans le même corpus.

TD [+notre traduction]	GRAMM. FR	GRAMM. IT	LEX-SEM. FR	STYL. IT	FAUTES	FREQ.
<b>Nous venons de visiter la cathédrale</b> [Abbiamo appena visitato la cattedrale]	– visiteremo la cattedrale +1 – abbiamo visitato – siamo andati a visitare+2 – siamo stati a visitare – siamo venuti per visitare					8/21
<b>Oui</b> [sì]		si+15				16/21
Cet appartement est <b>un trois pièces</b> avec une cuisine équipée, [Questo appartamento è un trilocale...]			– ha tre stanze+2 – ha tre piani – è un terzo piano – ha 3 piani – OMISSION+1 – è a 3 piani – è a 3 locali – questo appartamento di tre stanze – è un trivani – è composto da 3 stanze			13/21
<b>Elle</b> peut manger à la cantine [Può mangiare in mensa]				– <b>Lei</b> può mangiare+1 – <b>Ella</b> può mangiare		3/21
<b>Qui</b> est-ce que tu veux inviter ? [Chi vuoi invitare ?]					<b>Qui</b> vuoi invitare	1/21

Tableau 1. Grille de classement des erreurs pour un corpus de niveau A1<sup>4</sup>

<sup>4</sup> Les chiffres accompagnant les exemples indiquent le nombre d'occurrences de l'erreur dans le corpus.

Notre travail a été ensuite de rechercher ou de créer de nouveaux textes afin de recontextualiser et d'analyser certaines unités de traduction soit identifiées comme sources d'erreurs au cours de nos premières analyses, soit employées par les manuels de FLE pour définir les différents degrés de compétence en français langue étrangère.

Notre grille a été appliquée à l'ensemble du corpus pour réfléchir sur les performances des apprenants et les comparer ensuite avec celles des systèmes de TA.

## 2. INCIDENCE ET FRÉQUENCE DES ERREURS HUMAINES

Dans un premier temps, nous avons observé l'incidence de chaque catégorie d'erreur à l'intérieur du corpus (Tab. 2), ceci en comparant le nombre de phénomènes classés sous les différentes catégories sans considérer le nombre d'occurrences de chaque erreur (fréquence) ; dans un deuxième temps nous avons étudié les données liées à la fréquence, c'est-à-dire le nombre d'occurrences des erreurs générées par la même unité de traduction (Tab. 3), ceci dans le but de mieux comprendre si la fréquence enregistrée pour une erreur est strictement liée à l'incidence de sa catégorie d'appartenance. Ce croisement des données nous a paru utile pour approfondir le concept d'unité de traduction problématique.

Dans notre corpus, nous avons observé une haute incidence des erreurs de nature lexico-sémantique qui représentent en moyenne 50% des erreurs repérées sur l'ensemble des travaux, indépendamment des conditions de travail et du niveau de langue (Tab. 2). Cette catégorie se caractérise aussi par une fréquence d'apparition nettement supérieure aux autres : par exemple, pour la même unité de traduction concernée, les erreurs qui se manifestent dans au moins trois copies constituent, pour environ 60%, des erreurs classées « lexico-sémantique fr », tandis que 20% seulement représentent des erreurs de « grammaire fr » (Tab. 3).

Ce qui frappe en outre l'analyste, c'est la maîtrise lacuneuse de la langue maternelle, l'italien, laquelle constitue une source d'erreur assez significative, au niveau grammatical comme au niveau stylistique, pour environ 25% des copies analysées (Tab. 2) et qui représente un problème même pour les niveaux de langue les moins élevés, A1 et A2, dans la statistique relative à la fréquence (Tab. 3).

<b>TYPE D'ERREUR</b>	<b>% A1</b>	<b>% A2</b>	<b>% B1</b>	<b>% B2</b>	<b>% C1</b>
<b>Lexico-Semantique Fr</b>	45 %	32%	45%	58%	64%
<b>Grammaire It + Stylistique It</b>	23 %	30%	35%	20 %	16,5%
<b>Grammaire Fr</b>	22%	25%	12 %	11%	10%
<b>Fautes</b>	9%	13%	13%	11,5 %	9,5%

Tableau 2. Incidence des catégories d'erreur dans chaque corpus

<b>TYPE D'ERREUR</b>	<b>% A1</b>	<b>% A2</b>	<b>% B1</b>	<b>% B2</b>	<b>% C1</b>
<b>Lexico-Semantique Fr</b>	55%	39%	56%	76%	76%
<b>Grammaire It + Stylistique It</b>	13%	36%	26%	13%	14%
<b>Grammaire Fr</b>	30%	25%	15%	11%	10%
<b>Fautes</b>	2%	/	3%	/	/

Tableau 3. Fréquence de chaque catégorie d'erreur pour la même unité de traduction

Ces chiffres nous informent en outre sur l'influence relative de certains facteurs pour qu'une traduction soit adéquate : le transfert vers sa langue maternelle, des conditions de travail favorables et un texte de départ dont le niveau de langue est bas ne garantissent pas forcément une traduction adéquate, c'est-à-dire transmettant au plus près le sens convoité par l'original (cf. Franzelli 2013).

Nos considérations peuvent être mieux expliquées par les analyses des traductions de niveau C1. Pour ce niveau nous avons examiné :

- un corpus constitué d'exercices de traduction faits à la maison par des étudiants de Master en LLCE, en cours d'apprentissage ;
- un test proposé à un autre groupe d'étudiants du Master, en salle de cours et en conclusion d'une année de leçons, avec néanmoins la possibilité d'accéder à Internet et de consulter toute ressource disponible en ligne.

Le premier corpus nous a fourni 40 unités de traduction problématiques pour les apprenants, étant donné la haute fréquence avec laquelle ceux-ci se manifestent. Le deuxième a été constitué pour vérifier si ces unités se confirmaient en tant que source d'erreur même dans un autre contexte. Le récit ne comportait pas de nouvelles difficultés et il présentait un sujet familier et actuel, tel que la crise existentielle des nouvelles générations dans un contexte historique de récession économique, auquel faire face tout en luttant contre les idéaux et l'histoire de sa famille d'origine.

Nous avons gardé des conditions de travail semblables à celles fournies pour le premier corpus (traduction avec possibilité de consultation de ressources en ligne), même si nous avons modifié le temps à disposition (deux heures au lieu d'une journée entière) et le lieu de travail (en salle de cours et non pas à la maison), ainsi que le degré de préparation des étudiants, puisque ces derniers ont fait leur traduction en fin d'année.

Résultat : une partie des unités générant une haute fréquence d'erreurs dans le premier corpus se sont révélées moins problématiques dans le deuxième (Tab. 4).

<b>TYPE D'ERREUR</b>	<b>% I<sup>er</sup> corpus C1</b>	<b>% II<sup>ème</sup> corpus C1</b>
<b>Grammaire Fr</b>	24%	15%
<b>Grammaire It</b>	50 %	7%
<b>Lexico-Semantique Fr</b>	65%	26%
<b>Stylistique It</b>	100 %	10%

Tableau 4. Fréquence de chaque catégorie d'erreur pour la même unité de traduction

Ce qui favorise la compréhension du texte de départ, en réduisant ainsi la fréquence des erreurs générées par un élément spécifique, c'est tout d'abord la préparation atteinte par les étudiants en fin d'année, à laquelle s'ajoute le fait que le sujet du récit est familier et qu'il donne un nouveau co-texte aux 40 unités considérées comme problématiques dans le I<sup>er</sup> corpus.

Si l'on observe en effet l'incidence de chaque catégorie sur l'ensemble des erreurs relevées (Tab. 5), dans le II<sup>ème</sup> corpus on enregistre une hausse des unités de traduction qui sont à l'origine d'erreurs de grammaire, soit liées à des lacunes en langue de départ (de 5 à 12%) soit en langue d'arrivée (de 5 à 8%).

<b>TYPE D'ERREUR</b>	<b>% I<sup>er</sup> corpus C1</b>	<b>% II<sup>ème</sup> corpus C1</b>
<b>Grammaire Fr</b>	5%	12%
<b>Grammaire It</b>	5%	8%
<b>Lexico-Semantique Fr</b>	65%	65%
<b>Stylistique It</b>	25%	15%

Tableau 5. Incidence des différentes catégories d'erreur dans chaque corpus

Il est intéressant de remarquer que souvent il s'agit de phénomènes qui ne se sont pas manifestés dans d'autres corpus ou qui sont injustifiés si l'on considère le degré de préparation attendu en Master : par exemple, l'unité « depuis », « da » en italien, dont l'étudiant devrait connaître la traduction déjà en 2<sup>ème</sup> année de Licence, dans le deuxième corpus a été traduite par « dopo » (fr. « après ») si accompagnée par des co-textes spécifiques. Par contre, les erreurs de nature stylistique montrent une baisse d'incidence (de 25 à 15%), alors que l'incidence des erreurs de lexico-sémantique reste telle quelle, quoique leur fréquence dans le même corpus soit devenue moins problématique (Tab. 4).

Une unité de traduction n'est donc pas problématique de manière absolue, puisque cela dépend de nombreux facteurs qui vont au-delà du degré de préparation des apprenants en langue et culture de départ. La performance humaine est déterminée aussi par les compétences en langue d'arrivée (même si elle correspond à la langue maternelle), le contexte (la typologie textuelle, le sujet traité), le co-texte dans lequel certaines unités se manifestent et les conditions (matérielles et psychologiques) de travail.

En partant de ces considérations, nous allons maintenant considérer les similarités et les différences entre les performances de la machine et celles de nos étudiants.

### 3. LEXICO-SÉMANTIQUE ET LANGUE MATERNELLE : DEUX PROBLÈMES POUR LA TRADUCTION HUMAINE ET AUTOMATIQUE

Les traductions effectuées par les systèmes de TA disponibles en ligne ne présentent pas de phénomènes différents des travaux d'étudiants, parce qu'ils sont élaborés en gros comme on instruit un apprenant<sup>5</sup> : acquisition des règles, pour

<sup>5</sup> On pourrait objecter que l'apprentissage et le travail de l'humain ne sont pas superposables et qu'ils ne mettent pas en jeu les mêmes ressources et le même type de raisonnement. Nous

les systèmes *rule-based* comme l'était Systran<sup>6</sup> au début de sa diffusion, et mémorisation d'une série de schémas récurrents, pour les systèmes statistiques et hybrides comme Google Translate<sup>7</sup> ou Bing<sup>8</sup>, à reconnaître et reproduire dans un texte d'arrivée, suivant le co-texte et le contexte (cf. Barone 2011 ; Bersani Berselli 2011). La traduction sera plus ou moins adéquate en fonction de la quantité de règles et de corpus exploités par le système. C'est pourquoi certaines typologies textuelles, ayant une fonction informative ou promotionnelle et pour lesquelles il existe de nombreuses ressources en ligne, ne posent pas trop de problèmes aux systèmes de TA, à base de règles, statistiques ou hybrides qu'ils soient (cf. Franzelli 2013).

Afin de mieux illustrer ces similarités, prenons cet extrait tiré de *Mémoires d'une jeune fille rangée* (S. de Beauvoir 1958) et faisant partie des textes traduits par les étudiants de troisième année de Licence LLCE (niveau B2) :

- (3) *Son regard tranquille me protégeait pendant que je faisais des pâtés au Luxembourg.*

Nous rencontrons ici deux ambiguïtés liées au segment « au Luxembourg » : cette référence culturelle avec sous-entendu à saisir (le jardin situé dans le 6<sup>ème</sup> arrondissement de Paris) est le co-texte qui a une certaine importance pour la restitution italienne du mot français « pâtés ». Le co-texte représente en effet ces unités orientant l'interprétation et la traduction de chaque élément à l'intérieur d'une phrase : la présence de « au Luxembourg » aide à comprendre qu'il ne s'agit pas de pâtés de foie ; on dira donc :

- (3a) Il suo sguardo tranquillo mi proteggeva mentre facevo **formine di sabbia ai giardini del Luxembourg** (notre traduction)

Notre version italienne explicite la référence culturelle (« ai giardini del Luxembourg »), ce qui est peut-être nécessaire pour un destinataire italien ignorant le contexte auquel fait référence le roman, c'est-à-dire la ville de Paris.

Nous avons soumis ce segment aux trois systèmes de TA disponibles en ligne, Google Translate (GT), Systran (SY) et Bing (BI). Voici les résultats :

---

faisons ici allusion au fait que dans l'enseignement d'une langue étrangère, et de la traduction en particulier, on suit un « programme » de règles, d'exercices, de textes, de fonctions à apprendre même par cœur (notamment pour les débutants !). Et pour les machines c'est quand même un être humain qu'il y a derrière la programmation. En outre, les textes alignés et les mémoires de traduction sont des produits humains exploités par la TA.

6 <http://www.systranet.com/it/translate>.

7 <http://translate.google.it>.

8 <http://www.microsofttranslator.com>

- (3b) I suoi occhi tranquilli me protetti, mentre stavo facendo **torte Lussemburgo**. (GT)
- (3c) Il suo sguardo calmo lo proteggeva mentre facevo **paté in Lussemburgo**. (SY)
- (3d) I suoi occhi tranquille mi ha protetto mentre stavo facendo la **pasta per il Lussemburgo**. (BI)

Il est évident que les trois systèmes n'ont pas saisi la référence, c'est pourquoi le co-texte n'aide pas à traduire de manière adéquate « pâtés ». Ainsi, Luxembourg devient pour la machine un type de gâteau (3b), ou bien le lieu où le sujet fait du pâté de foie (3c), ou finalement le Pays auquel sont destinées des pâtés (3d). Il est en outre intéressant de remarquer que les systèmes n'ont pas acquis certaines compétences morphosyntaxiques : c'est le participe passé « protetti » (fr. « protégés ») qui traduit l'imparfait « protégeait » (3b) ; « lo » (fr. « le ») remplace « me » modifiant ainsi le COD (3c) ; finalement « occhi » n'est pas reconnu en tant que substantif masculin pluriel, c'est pourquoi il y a une erreur d'accord avec « tranquille », adjectif féminin pluriel, et de conjugaison dans « ha protetto », passé composé à la troisième personne singulière et non pas plurielle (3d).

Voyons comment le même segment a été traduit à la maison par deux étudiants, avec la possibilité de consulter toute source de données :

- (3e) Il suo sguardo tranquillo mi proteggeva mentre **preparavo delle creme del Lussemburgo**.
- (3f) Il suo sguardo tranquillo mi proteggeva mentre **combinavo guai a Lussemburgo**.

Alors que les erreurs de grammaire sont absentes, on ne peut pas en dire de même de l'imagination : des crèmes spéciales deviennent les produits du terroir du Luxembourg (3e) et sa capitale un endroit où l'auteur serait dans le pétrin, « guai » en italien (3f). Dans les deux cas, les étudiants n'ont ni compris la référence ni considéré le co-texte et ils se sont arrêtés au mot « pâtés », jusqu'à donner à la traduction italienne fournie par les dictionnaires (« pasticcio ») le sens figuré de « guaio » (3f).

Les différences entre les deux performances erronées de la machine et de l'homme dépendent ici des compétences en italien, mais non seulement au niveau morphosyntaxique : contrairement aux systèmes de TA, l'étudiant, qui travaille vers sa langue maternelle, a la capacité d'exploiter sa mémoire de natif en reformulant le texte de départ pour lui donner un sens dans la version en langue d'arrivée, bien qu'inadéquat par rapport au texte de départ. La machine n'a malheureusement pas encore la capacité d'ajuster de manière cohérente ses résultats, parce qu'ils sont donnés par une opération statistique agissant en fonction de la quantité des données et non pas de leur qualité.

Dans notre travail de classement des erreurs humaines, l'examen du corpus a mis en évidence deux problématiques liées à la maîtrise de la langue d'arrivée,

l'italien, qui était la langue maternelle des étudiants : d'un côté la fréquence d'erreurs liée aux lacunes en grammaire et en stylistique italiennes, de l'autre les différentes conditions d'erreur et de faute des étudiants étrangers.

Dans les tableaux 4 et 5, nous avons observé que les conditions de travail favorables n'empêchent pas aux apprenants de faire des erreurs de grammaire ou de stylistique italiennes. En particulier, dans le corpus B1, nous avons enregistré un emploi erroné des accents, des apostrophes et des virgules : « sì » (fr. « oui ») apparaît souvent sans accent (59% sur l'ensemble du corpus) ; « po' » (troncation de « poco », « peu » en français) a un accent et non pas une apostrophe dans 14% des copies ; les virgules disparaissent dans 36% des phrases où elles sont dans le texte de départ et elles sont nécessaires dans celui d'arrivée.

Pour ce qui est des étudiants qui n'ont pas l'italien comme langue maternelle, à ces difficultés il faut en ajouter d'autres. Il s'agit en effet d'apprenants migrants ou enfants de migrants, qui sont régulièrement inscrits à l'université italienne, mais qui ont fait une bonne partie de leurs études dans leurs pays d'origine. Observons par exemple la traduction de cet extrait de *Comme un roman* (D. Pennac 1992) traduit, à l'occasion de leur examen, par des étudiants de III<sup>ème</sup> année de Licence (B2) :

- (4) *Lecture d'autant plus délicieuse qu'elle se déroula de nuit* (texte de départ)
- (4a) Lettura tanto più piacevole dal momento che si svolse di notte (notre traduction)
- (4b) Lettura tanto più deliziosa **di quella che** si svolse di notte (traduction d'un étudiant italoophone)
- (4c) Lettura **così** più **deliciosa** che **si farà** di notte (traduction d'un étudiant de langue maternelle non italienne)

Si l'étudiant italoophone (4b) fait une seule erreur sémantique, due à la méconnaissance de « d'autant plus... que », l'étudiant dont l'italien n'est pas la langue maternelle (4c) propose une solution qui contient plusieurs erreurs, non seulement générées par la même unité traductive, mais aussi dictées par un manque de compétences morphosyntaxiques, lexicales et stylistiques en italien, ce qu'il essaie de compenser en recourant sans doute à sa langue maternelle, produisant en outre une sorte d'interférence avec « deliciosa ».

Or, ces étudiants se préparent aux épreuves comme leur collègues de langue maternelle italienne et, par conséquent, ils sont évalués de la même manière, ainsi, un excès de certains phénomènes, qui sont des fautes occasionnelles ou même improbables pour un italoophone, risque de compromettre la réussite des étudiants de langue maternelle non italienne. Pour nos analyses des erreurs, nous avons décidé de ne pas considérer leurs copies (environ 10% de l'ensemble du corpus), même si elles nous invitent à réfléchir non seulement sur la formation des étudiants migrants en Italie, mais aussi sur l'élaboration de systèmes de TA et sur les interférences en traduction.

La performance de ces étudiants est bien plus similaire en effet à celle des systèmes de TA :

- (4d) **Ancora più delizioso lettura** che ha avuto luogo durante la notte. (GT)
- (4e) Lettura **tanto più deliziosa di quanto** si svolse di notte. (SY)
- (4f) Ancor più piacevole **leggere che si è svolto** durante la notte. (BI)

Tout comme pour l'étudiant qui n'est pas de langue maternelle italienne, les résultats donnés par les systèmes montrent leur impossibilité d'élaborer le texte de départ et de comparer celui d'arrivée avec une mémoire suffisamment influente pour éliminer toute sorte d'incohérence à la traduction. Ces segments frappent en outre pour les déplacements et les recatégorisations : l'adjectif « *delizioso* » traduisant « délicieuse » passe avant le substantif « *lettura* » (« lecture » en français) qui était en tête d'énoncé dans l'original (4d) ; le substantif « lecture » devient le verbe « *leggere* », « lire » en français (4f). Ces interventions sur le texte de départ relèvent de la capacité des systèmes de passer de la traduction mot à mot, faite en consultant des textes alignées, à son réajustement en fonction d'un corpus en langue d'arrivée qui instruit le système sur la distribution et les usages les plus fréquents de chaque élément traduit.

On parle en effet de *translation model* et de *language model*, le premier instruisant le système sur le passage d'une langue à une autre, et le deuxième étant le modèle de langue fourni par un corpus en langue d'arrivée que la machine doit exploiter statistiquement afin de garantir des résultats adéquats.

Lorsque le modèle de traduction pour deux langues n'est pas suffisamment alimenté, des systèmes comme celui de Google peuvent présenter des interférences que seulement un modèle de langue bien structuré pourrait corriger. Observons les traductions d'un extrait de niveau B2 tiré du roman *L'enfant et la rivière* (H. Bosco, 1945) :

- (5) *C'était une femme à l'antique avec la coiffe de piqué.* (texte de départ)
- (5a) Era una donna all'antica con la cuffia **di piqué.** (notre traduction)
- (5b) Era una donna all'antica con la cuffia [**omission**] (traduction d'étudiant)
- (5c) Era una donna con copricapo **dive** d'epoca. (GT)

Si l'étudiant ignorant le « piqué » décide de l'omettre dans sa traduction (5b), le système ne peut pas le faire, mais il donne un résultat qu'on peut avoir du mal à interpréter, car le mot italien « *dive* » (en français « *divas* ») est indiqué comme la traduction de ce tissu (5c). Si l'on sélectionne le terme, le système nous donne en outre des alternatives également ambiguës, toutes relevant d'un sport de mer : « *tuffo* », « *immersione* », « *di immersione* », « *immersioni* ». Voilà qu'on découvre la source de « *dive* », c'est-à-dire l'anglais « *to dive* », « plonger » en français, auquel le système est arrivé en donnant à « piqué » une autre de ses possibles significations, c'est-à-dire « plongeon ». On comprend donc que

n'ayant pas de statistiques à consulter pour la traduction du français vers l'italien du terme « piqué », le traducteur automatique emploie l'anglais comme langue pivot qui va donner lieu à cette interférence.

L'anglais reste en effet la langue la plus exploitée par les systèmes de TA, car il existe beaucoup de matériel traduit, c'est pourquoi on peut la considérer comme la langue maternelle à laquelle ils font recours lorsque leur modèle de traduction n'est pas suffisamment alimenté, comme c'est le cas pour le transfert entre l'italien et le français. La machine produit ses résultats en fonction de la précision et de la quantité des informations à sa disposition : elle devra atteindre un certain volume de données correctes en français et en italien, avant de mettre à jour ses statistiques et de diminuer la probabilité d'erreur et de recours à l'anglais.

## CONCLUSION

Par l'analyse de notre corpus, nous avons appris que la définition des unités de traduction problématiques n'est pas absolue, mais qu'elle dépend de nombreux facteurs qui vont au-delà du degré de compétence en une langue et en une culture étrangères. La performance humaine varie aussi selon les compétences en langue d'arrivée (bien que maternelle), les conditions matérielles et psychologiques de travail, le contexte de traduction (la typologie textuelle, le sujet traité), le co-texte à l'intérieur duquel une certaine unité se manifeste. Même dans des conditions de travail propices (avec toute source d'informations à disposition), les étudiants font des erreurs d'origine lexico-sémantique et, surtout, de grammaire et de stylistique italiennes, une langue que l'on présume acquise dans des parcours d'étude centrés sur les langues et les littératures étrangères.

Ces observations montrent que la maîtrise de la L1 est un élément auquel il faut travailler dans la formation en LLCE, en pensant tout spécialement à des étudiants migrants qui ne sont pas de langue maternelle italienne et qui peuplent de plus en plus les bancs des universités. Leur préparation devrait sans doute devenir plus intensive par rapport à celle de leurs collègues italophones, ceci dans le but de réduire le phénomène des interférences et les échecs aux examens.

Si l'on exclut les conditions psychologiques de travail, en comparant les performances de la machine avec celles des étudiants, nous pouvons affirmer que les facteurs cités ci-dessus ont le même poids pour les deux. Ainsi, les erreurs humaines peuvent guider dans le choix des textes servant pour l'alimentation des modèles de traduction et de langue d'un système. Il est donc possible d'améliorer le travail des systèmes de TA en donnant un support linguistique, basé sur la qualité, à l'approche informatique, tendanciellement centrée sur la quantité des données.

Un système de traduction automatique pourrait par exemple apprendre une langue étrangère de manière progressive, comme un apprenant humain, si l'on adaptait les contenus à ceux du CECRL. C'est en effet à partir de cette hypothèse

que Charles Barone a travaillé à la structuration du corpus destiné à l'apprentissage d'un système de TA dérivé de la plateforme logicielle libre Moses.

Il est en outre possible de diminuer la faillibilité de la machine en considérant nos analyses pour la création de corpus ad hoc non seulement selon les niveaux de langue du CECRL (compétences en langue et en culture étrangère), mais aussi en peaufinant la qualité des textes et de la langue d'arrivée notamment pour le transfert du français à l'italien, deux langues qui ne sont pas vraiment exploitées en traduction automatique.

- Arnold Doug (2003) « Why translation is difficult for computers », in Somers Harold (dir.), *Computers and Translation: a Translator's Guide*, Amsterdam, John Benjamins, pp. 119-142.
- Barone Charles (2011), « La traduzione automatica », in Barone Charles, Bruti Silvia, Foschi Albert Marina, Tocco Valeria (a cura di), *Dallo stilo allo schermo: sintesi di teoria della traduzione*, Pisa, Pisa University Press, pp. 113-126.
- Bersani Berselli Gabriele (2011), *Usare la traduzione automatica*, Bologna, CLUEB.
- Chauvet Aude (2008), *Référentiel pour le Cadre Européen Commun. A1-A2-B1-B2-C1-C2*, Paris, Alliance Française/CLE International.
- Collombat Isabelle (2009), « La didactique de l'erreur dans l'apprentissage de la traduction », *The Journal of Specialised Translation*, 12, pp. 37-54
- Corder S. Pit (1980), « Que signifient les erreurs des apprenants ? », *Langage*, 57, pp. 9-15.
- Franzelli Valeria (2013), « Per un approccio linguistico alla traduzione automatica. Fattori d'errore nelle traduzioni automatizzate e umane », in Talone Sandra, Zemella Teresina (a cura di), *Il traduttore visibile*, Parma, MUP.
- Forcada Mikel L. (2010), « Machine translation today », in Gambier Yves, Van Doorslaer Luc (dir.), *Handbook of Translation Studies*, Amsterdam/Philadelphia, John Benjamins, vol. 1, pp. 215-223.
- Gambier Yves (2011), « M4 et tradaptation », in Liénard Fabien, Zlitni Sami (dir.), *La communication électronique : enjeux de langues*, Limoges, Lambert-Lucas, pp. 157-164.
- Ladmiral Jean-Réné (1972), « La traduction dans l'institution pédagogique », *Langages*, 28, pp. 8-39.
- Hansen Gyde (2010), « Translation 'error' », in Gambier Yves, Van Doorslaer Luc (dir.), *Handbook of Translation Studies*, Amsterdam/Philadelphia, John Benjamins, vol. 1, pp. 385-388.
- Nord Christiane (2008), *La traduction : une activité ciblée. Introduction aux approches fonctionnalistes*, traduction de Beverly Adab, Arras, Artois Presses Université.
- Reiss Katharina (2009), *Problématiques de la traduction*, traduction de Catherine A. Bocquet, Paris Economica/Anthropos.