

ÉVALUATION DE LA PRODUCTION DE QUATRE SYSTÈMES TRADUCTION
AUTOMATIQUE

by

Christine Yen

Submitted in partial fulfilment of the requirements
for the degree of Master of Arts

at

Dalhousie University
Halifax, Nova Scotia
December 2013

© Copyright by Christine Yen, 2013

À ma famille qui n'a pas cessé de m'épauler dès le début.

TABLE DES MATIÈRES

Liste des figures.....	vi
Résumé.....	vii
Remerciements.....	viii
CHAPITRE 1 INTRODUCTION.....	1
1.1 Objectifs et organisation du mémoire.....	3
1.2 Définition et objet de la traduction automatique.....	4
1.3 Critique des publications antérieures.....	7
1.3.1 Popović et Ney.....	9
1.3.2 Balkin.....	11
1.3.3 Olvera-Lobo et García-Santiago.....	12
1.3.4 Hampshire et Salvia.....	13
1.3.5 Aiken et al.....	16
CHAPITRE 2 LA TRADUCTION AUTOMATIQUE : D’HIER A AUJOURD’HUI.....	20
2.1 Historique	20
2.2 Processus de la traduction automatique.....	24
2.3 Deux types de systèmes de traduction automatique.....	29
2.3.1 Systèmes de traduction automatique basés sur le dictionnaire.....	30
2.3.2 Systèmes de traduction automatique basés sur les statistiques.....	32
2.4 Quelques systèmes de traduction automatique existants.....	35
2.4.1 Systèmes de traduction automatique commerciaux.....	36
2.4.2 Systèmes de traduction automatique gratuits.....	38
CHAPITRE 3 EVALUATION DE QUATRE MACHINES DE TRADUCTION AUTOMATIQUE.....	43

3.1	Présentation du corpus.....	43
3.2	Analyse de la performance des machines.....	49
3.2.1	Promt.....	49
3.2.2	Babylon.....	54
3.2.3	Bing	58
3.2.4	Google Translate	61
CHAPITRE 4 ANALYSE DES ERREURS DE TRADUCTION AUTOMATIQUE DANS QUATRE SYSTEMES.		66
4.1	Typologie des erreurs.....	66
4.1.1	Erreurs grammaticales.....	67
4.1.2	Erreurs lexicales.....	69
4.1.3	Erreurs sémantiques.....	71
4.1.4	Erreurs stylistiques.....	74
4.2	Récapitulation des problèmes de traduction automatique rencontrés lors de l'examen de 4 systèmes.....	76
4.2.1	Problèmes liés aux systèmes basés sur les règles	76
4.2.2	Problèmes liés aux systèmes de traduction statistique.....	79
CHAPITRE 5 CONCLUSION ET PERSPECTIVES DE LA TRADUCTION AUTOMATIQUE.....		82
5.1	Suggestions d'amélioration en vue d'un modèle supérieur de traduction automatique.....	84
5.1.1	Recommandations pour le système de traduction basé sur le dictionnaire.....	84
5.1.2	Recommandations concernant le système de traduction statistique.....	87
5.2	Suggestions à l'intention de l'utilisateur	89
5.3	Perspectives d'avenir de la traduction automatique.....	91
RÉFÉRENCES BIBLIOGRAPHIQUES.....		95

APPENDICE A Corpus.....	99
APPENDICE B Erreurs commises par type.....	142
APPENDICE C Erreurs commises par sous-type.....	144

LISTE DES FIGURES

Figure 1 Le processus de traduction dans les systèmes basés sur le dictionnaire.....	31
--	----

RÉSUMÉ

Cette thèse vise à contribuer à l'amélioration des logiciels de traduction automatique en ligne. Nous identifions les erreurs commises dans la traduction de l'anglais vers le français, puis formulons des recommandations. Les systèmes évalués sont *Prompt*, *Babylon*, *Google Translate* et *Bing* et le corpus de référence est tiré de *BankGloss*. *Prompt* a commis le plus d'erreurs, suivi par *Babylon*, *Bing* et *Google*. Les systèmes ensemble ont produit un total de 147 erreurs grammaticales, 74 erreurs sémantiques, 17 erreurs lexicales et 6 erreurs stylistiques. Pour améliorer *Prompt*, nous recommandons l'élargissement de son dictionnaire. Pour *Babylon*, il faut incorporer plus de règles de grammaire et de structure. Afin de réduire le nombre d'erreurs sémantiques commises par *Google* et *Bing*, ces logiciels devraient apprendre à identifier les mots à partir du contexte. La traduction automatique n'est pas une fin en soi, mais une bonne aide qui permet d'accomplir les tâches de traduction.

ABSTRACT

This thesis aims to contribute to the improvement of online machine translation software. We identify errors in the process of translation between English and French and make recommendations. The systems evaluated are *Prompt*, *Babylon*, *Google Translate* and *Bing* and the reference corpus is taken from *BankGloss*. *Prompt* made the most errors, followed by *Babylon*, *Bing* and *Google*. The systems together produced a total of 147 grammatical errors, 74 semantic errors, 17 lexical errors, and 6 stylistic errors. To improve *Prompt*, we suggest expanding its dictionary. For *Babylon*, we advise adding more grammar rules. In order to reduce the number of semantic errors in *Bing* and *Google*, the software should learn to identify words according to context. Machine translation is not an end in itself, but a good aid in accomplishing translation tasks.

REMERCIEMENTS

Je voudrais tout d'abord remercier mon directeur de thèse, Dr Mopoho, pour ses conseils, son encouragement et les nombreuses heures consacrées à l'édition de mon travail. Sa passion pour la traduction m'a donné l'inspiration et la motivation pour continuer mes recherches.

Je tiens également à remercier Dr Frigerio et Dr Milićević qui m'ont préparés pour entreprendre ce projet et m'ont donné des indications utiles sur les méthodes de recherche et les études en linguistique.

Des remerciements spéciaux vont à ma famille et mes amis qui m'ont soutenus jusqu'à la fin.

CHAPITRE 1

Introduction

La demande de technologies d'automatisation des tâches informatiques croît à mesure de l'évolution du monde et de l'avancement de la recherche. Le domaine de la traduction n'échappe pas à cette règle. Il y a deux décennies, Hutchins a établi qu'il n'existait aucune machine capable de produire une traduction parfaite sans assistance humaine (Hutchins, 1992 : 1). Malheureusement, on dirait que c'est encore le cas aujourd'hui, les systèmes de traduction automatique n'étant toujours pas capables de produire des traductions d'un niveau acceptable, sauf sur la base de critères précis qui sont identifiés à l'avance et inscrits dans leur programme (Delavenay, 1960 : 6). Il se pose donc la question de savoir pourquoi, malgré la capacité des machines d'effectuer avec succès des calculs mathématiques complexes, les traductions produites électroniquement sont toujours pleines d'erreurs. Loffler-Laurian décrit les langues comme « une sorte d'organisme vivant » qui peut être manipulée d'une façon créative (Loffler-Laurian, 1996 : 79). La plupart du temps, l'assistance humaine est encore nécessaire pour obtenir une traduction de bonne qualité. Le problème semble résider dans le fait que la demande de traduction est supérieure à l'offre de services de traduction humaine professionnelle, situation qui rend très attrayant le recours à un ordinateur, d'autant plus que les chercheurs sont toujours en train d'améliorer les logiciels de traduction.

Eugene Nida avait formulé dès les années 60 une théorie relative à la traduction automatique. Il a établi que l'évolution de la théorie linguistique avait apporté un éclairage important à la théorie et à la pratique de la traduction. Cette évolution a conduit au constat selon lequel la traduction n'est pas fondamentalement un processus d'appariement des formes de surface au moyen de règles de correspondance, mais plutôt une procédure plus complexe faisant appel à l'analyse du texte source, au transfert de la langue de départ à la langue d'arrivée et à la restructuration de ces langues. Ces procédures linguistiques, telles que les transformations au sens de la grammaire générative et de l'analyse componentielle, fournissent des bases beaucoup plus solides pour la traduction, par rapport à ce qui a été utilisé dans le passé. En même temps, la théorie de la traduction est en mesure d'apporter à la science linguistique de nouvelles connaissances en donnant plus d'informations sur la structure des langues, ce qui rend possible l'amélioration des méthodes de vérification des hypothèses linguistiques (Nida, 1969 : 483).

Selon Henisz-Dostert, le but principal de la plupart des travaux de recherche dans le domaine de la traduction automatique est d'obtenir une traduction rapide de documents techniques. Pour cette raison, l'accent est mis sur les textes de disciplines scientifiques comme la chimie organique, l'aéronautique, l'astrophysique ou la météorologie. Le choix de textes de ce type réduit les problèmes potentiels de traduction, parce qu'il est moins probable que la terminologie scientifique soit ambiguë par rapport à celle de la littérature ou de l'histoire. Il est aussi possible que les textes techniques comportent moins de

spécificités culturelles qui doivent être transférées dans la langue cible (Henisz-Dostert, 1979 : 93). Aujourd'hui, la plupart des textes traduits automatiquement sont des documents scientifiques et techniques, des opérations commerciales et des modes d'emploi, et non des textes ayant un fort contenu culturel ou littéraire. L'utilité pratique d'un système de traduction automatique est déterminée par la qualité de ses données de sortie. Pourtant, ce qu'on considère comme une bonne traduction est difficile à déterminer, tout dépendant des circonstances particulières entourant l'émission du message, et du public visé par ce dernier (Hutchins, 1992 : 2). Selon Loffler-Laurian, une bonne traduction se caractérise par l'harmonie du style, la fluidité de l'expression, le rythme de la phrase, la clarté de l'énoncé et bien d'autres notions encore (Loffler-Laurian, 1987 : 154). D'après Hutchins, les obstacles majeurs entravant la traduction sont d'ordre linguistique plutôt qu'informatique. Il s'agirait de problèmes liés à l'ambiguïté lexicale, à la complexité syntaxique, aux différences de vocabulaire entre deux langues et aux constructions elliptiques. Cependant, ce qui importe dans une bonne traduction, en fin de compte, c'est simplement la préservation du sens entre la langue de départ et la langue d'arrivée (Hutchins, 1992 : 2).

1.1 Objectifs et organisation du mémoire

La présente thèse vise à examiner les différentes approches adoptées jusqu'ici en traduction automatique, et à suggérer des solutions à quelques problèmes qui se pose dans ce domaine, en évaluant quatre systèmes de traduction. Grâce à la technologie moderne, les machines de traduction d'aujourd'hui ont un

fonctionnement beaucoup plus complexe. Toutefois, d'importantes améliorations sont encore nécessaires, en raison de la complexité des difficultés qui continuent d'exister dans ce domaine. Dans la présente étude, le chapitre introductif couvrira la vue d'ensemble du domaine de la traduction, y compris la critique des principales publications antérieures. Le chapitre 2 portera sur l'évolution de la traduction automatique, notamment l'analyse du processus et des systèmes propres à cette traduction. Au chapitre 3, nous allons évaluer quatre traducteurs gratuits actuellement disponibles en ligne, à savoir *Prompt*, *Babylon*, *Google Translate*, et *Bing*, en analysant les erreurs spécifiques qu'ils commettent. L'objectif est de savoir si ces logiciels sont assez efficaces en dépit des erreurs en question. Le chapitre 4 traitera de la typologie des erreurs commises par les quatre systèmes. L'accent sera mis sur l'identification des erreurs, plutôt que sur l'analyse des règles de fonctionnement des systèmes. Un récapitulatif des problèmes de traduction automatique rencontrés dans ces quatre systèmes sera présenté en guise de conclusion du chapitre. Enfin, au chapitre 5, nous traiterons des perspectives d'avenir de l'utilisation des systèmes de traduction disponibles en ligne.

1.2 Définition et objet de la traduction automatique

La traduction automatique sert à produire des traductions d'une langue naturelle vers une autre, avec ou sans assistance humaine (Hutchins, 1992 : 3). D'après Ranchhod, la traduction automatique est « une notion générique qui recouvre l'élaboration d'outils variés d'aide à la traduction (humaine) aussi bien que des

systèmes indépendants de traduction complètement mécanique » (Ranchhod, 2000 : 253). Par ailleurs, le terme *renvoyer* à l'utilisation des ordinateurs pour effectuer des opérations de traduction à grande échelle (ElShiekh, 2012 : 57). Le but principal de la traduction est de formuler des équivalences de signification entre une langue source et une langue cible. Le transfert d'une langue à une autre est effectué au niveau des équivalences de signification et non pas au niveau des mots, ni au niveau syntaxique (Ranchhod, 2000 : 265).

On a recours à la traduction, en particulier à la traduction automatique, pour les trois raisons suivantes : 1) l'assimilation, c'est-à-dire la traduction de textes étrangers dans le but d'en comprendre le contenu ; 2) la dissémination, qui consiste à traduire un texte à des fins de publication dans d'autres langues ; 3) la communication, définie comme la traduction des courriels électroniques, des discussions de clavardage, etc. (Koehn, 2010 : 20). L'utilisation de la traduction automatique est principalement motivée par le désir de saisir le sens général d'un texte. La traduction n'a pas besoin d'être parfaite pour permettre d'accéder au sens de l'original. La traduction automatique est dans ce cas un premier pas vers l'extraction d'information et l'analyse de grands volumes de textes étrangers (Koehn, 2010 : 21).

Dès les années 80, il est devenu évident que la croissance des informations scientifiques et technologiques était beaucoup plus rapide que la capacité de transmettre cette connaissance. Le besoin le plus immédiat, dans le domaine de la traduction, était la diffusion rapide d'informations dans plusieurs langues. La multitude des données à couvrir pose un dilemme : l'insuffisance du

temps, des ressources ou de l'incitation nécessaires pour traduire soigneusement tous les documents par les êtres humains. Les scientifiques sont généralement confrontés davantage à ce dilemme : déjà, on lit plus qu'on ne peut écrire pendant la même période. L'effort requis pour effectuer le travail de traduction sur les textes écrits dans les langues étrangères n'en vaut pas la peine quand il est peu probable que le texte donné revête de l'intérêt (Slocum, 1985 : 111).

La demande de traduction automatique augmente à mesure de l'intensification des connexions au niveau international (ElShiekh, 2012 : 57). Pour le moment, l'anglais est la langue maternelle de la majorité des utilisateurs d'Internet, mais les chercheurs prédisent qu'à l'avenir, ce ne sera plus le cas (ElShiekh, 2012 : 57). Plus précisément, avec la popularité croissante des équipes virtuelles mondiales (des groupes qui travaillent ensemble à travers les réseaux en ligne), les employés internationaux passent un temps considérable à assister à des réunions multinationales. L'anglais est une langue de communication couramment utilisée dans ces réunions, mais un nombre de locuteurs non natifs trouvent difficile la communication dans une langue seconde (Aiken et al., 2010 : 25). En conséquence, quelques professionnels ont du mal à effectuer leur travail dans leurs domaines de spécialisation. Si un moyen très rapide et peu coûteux de traduction était disponible, même une traduction de piètre qualité pourrait être suffisante pour obtenir des informations. Au pire des cas, le lecteur pourrait déterminer si une traduction plus soignée (et plus coûteuse) serait justifiée. Il est davantage probable que sa compréhension du

texte serait suffisante pour exclure la nécessité d'une traduction plus soignée (Slocum, 1985 : 111). Pour toutes ces raisons, l'utilisation des ordinateurs pour la traduction est devenue nécessaire.

1.3 Critique des publications antérieures

Des études récentes envisagent la traduction automatique à partir d'un éventail de points de vue, tels que la qualité des traductions produites par les machines de traduction gratuites dans des langues différentes, l'effet de ces machines sur les études traductologiques et la mesure dans laquelle les traducteurs professionnels les utilisent (Hampshire et al., 2010 : 200). En raison de la popularité croissante des machines de traduction gratuites, plusieurs auteurs ont mené des recherches dans ce domaine. Leurs travaux constituent une source de référence importante pour notre étude. Nous avons pris soin de choisir des études ayant pour centre d'intérêt l'évaluation d'un système de traduction. Nous espérons renforcer les résultats obtenus dans le cadre de ces recherches antérieures, et contribuer à l'avancement des connaissances dans le domaine. Nous concluons notre thèse sur une analyse des similarités et des différences entre nos résultats et ceux des chercheurs évoqués.

Dès les premières années de la traduction automatique, l'évaluation humaine s'est avérée nécessaire. Cependant, comme cette évaluation prend du temps et est donc coûteuse, l'on s'est intéressé à l'évaluation automatique. Des paramètres d'évaluation ont été élaborés pour l'évaluation automatique de la production de la traduction automatique (Anastasiou, 2011 : 645). Grâce au

développement continu des systèmes de traduction automatique, rendu possible par l'utilisation d'ensembles de tests sur des centaines ou des milliers de phrases, les paramètres d'évaluation automatiques sont devenus indispensables pour évaluer les traductions rapidement et de manière économique, et par extension, les machines de traduction automatique elles-mêmes. Le paramètre d'évaluation le plus populaire dans le domaine est *BLEU (Bilingual Evaluation Understudy)*. Ce système a été créé par *IBM* dans le but d'évaluer les textes traduits automatiquement, de manière rapide et peu chère (Ketzan, 2007 : 213). L'idée centrale sous-tendant *BLEU* est que plus qu'une traduction automatique est proche d'une traduction humaine professionnelle, mieux cela vaut (Anastasiou, 2011 :645). Cet indicateur est basé sur une comparaison des sorties d'un système à des traductions de référence (Ketzan, 2007 : 213). Une autre mesure commune est le taux d'erreur par phrase (*sentence error rate*) où le pourcentage de mauvaises correspondances entre les phrases complètes est calculé en comparant les traductions produites par le logiciel et les traductions de référence. Un pourcentage inférieur de ces mauvaises correspondances indique de meilleures traductions. Une troisième mesure, conçue par Levenshtein en 1966, est le taux d'erreur par mot (*word error rate*). Cette mesure calcule les divergences entre les traductions de référence et celles des logiciels candidats en fonction du nombre d'insertions, de substitutions et de suppressions de mots dans les traductions divisées par le nombre de mots de référence corrects (Way, 2009 : 546). Les évaluateurs jouent un rôle important dans l'analyse et l'évaluation des résultats des machines de traduction. Cependant, les chercheurs

ont trouvé que ces résultats avaient tendance à varier considérablement en fonction du nombre de traductions de référence utilisées dans l'évaluation. Pour cette raison, certains d'entre eux ont douté que les évaluateurs automatiques soient parfaitement ou même nettement fiables (Ketzan, 2007 : 213). Dans le cadre de notre étude, la procédure adoptée pour effectuer l'évaluation est semblable à celle de la troisième mesure. Nous allons manuellement compter le nombre d'erreurs produites par mot dans les traductions candidates, en comparant celles-ci à des documents traduits professionnellement.

1.3.1 Popović et Ney

Dans leur article publié en 2011, Popović et Ney ont remarqué que l'évaluation et l'analyse des erreurs de sortie de la traduction automatique étaient des tâches importantes mais difficiles. Ils ont proposé un cadre d'analyse et de classification automatique des erreurs : il faut d'abord identifier les mots erronés réels en utilisant les algorithmes pour le calcul du taux d'erreurs par mot ainsi que le taux d'erreurs par mot de position indépendante (Popović et al., 2011 : 658) (il s'agit des substitutions, des suppressions et des insertions, compte non tenu de l'ordre des mots) (Popović et al., 2011 : 661). Ces mots erronés peuvent ensuite être utilisés en combinaison avec des formes de base différentes, des mots composés, des suffixes, des préfixes, etc., afin d'obtenir diverses informations sur la nature des erreurs réelles, notamment les types d'erreur (erreurs morphologiques, erreurs de réorganisation et mots manquants) et la contribution des différentes classes de mots, etc. (Popović et al., 2011 : 658). C'est une

première étape vers l'élaboration de mesures automatiques d'évaluation qui fournissent des informations plus précises sur certains problèmes de traduction. L'approche proposée permet l'utilisation de différents types de connaissances linguistiques pour classer les erreurs de traduction de nombreuses façons. Le travail de Popović et Ney se concentrait sur cinq types d'erreurs : erreurs flexionnelles, erreurs d'ordre des mots, erreurs d'omission de mots, erreurs d'ajout de mots et erreurs liées aux mauvais choix lexicaux. Popović et Ney ont analysé la fréquence de chacun des types d'erreurs dans les différentes parties du discours. Ils ont comparé les résultats de l'analyse automatique des erreurs aux résultats de l'analyse humaine, afin d'étudier deux applications possibles : l'estimation de la fréquence de chaque type de faute dans une sortie de traduction donnée et la comparaison des différentes sorties de traduction, au moyen des types d'erreurs introduits. Le but était d'identifier les principales sources d'erreurs de chaque système de traduction et d'obtenir plus d'information sur les avantages et les inconvénients des différents systèmes. Les auteurs ont aussi proposé les possibilités d'amélioration, et décrit les avantages et les inconvénients des méthodes appliquées pour l'amélioration. Popović et Ney ont effectué l'évaluation en utilisant des sorties comme celles de *Newswire* et *Broadcast News* pour les paires de langues arabe-anglais et chinois-anglais. Les combinaisons allemand-anglais et espagnol-anglais ont fait l'objet d'évaluation elles aussi. Popović et Ney ont montré que leurs résultats concordaient très bien avec ceux d'une analyse humaine d'erreurs, et que tous leurs paramètres, à l'exception de l'ajout de mots, reflétaient bien les

divergences entre les différentes versions d'un même système de traduction ainsi que les divergences entre les différents systèmes de traduction (Popović et al., 2011 : 657). Dans la présente étude, nous adopterons aussi la méthode de l'analyse humaine d'erreurs pour évaluer les logiciels de traduction.

1.3.2 Balkin

Jusqu'en 1997, les services de traduction automatique n'étaient disponibles que pour les clients payants. Le système *Babelfish*, un programme de traduction conçu par *Altavista*, marquera le début d'une nouvelle ère de traductions gratuites pour l'internaute. En 1999, Balkin a effectué une étude sur *Babelfish*, en évaluant les traductions entre l'anglais et le français sur neuf sites Web différents. Pour mener cette évaluation, elle a adopté une méthode consistant à faire traduire d'abord le texte du français vers l'anglais, puis à traduire à nouveau vers le français le texte anglais obtenu (Balkin, 1999 : 56). Elle a trouvé qu'à la fin, les traductions étaient incompréhensibles et pleines d'erreurs. Ces erreurs comprenaient des fautes grammaticales, de mauvaises traductions de mots courants (par exemple, le pronom *elle* est traduit par *it* au lieu de *she*) ou des cas de confusion de parties du discours. Les erreurs relatives à l'ordre des mots étaient très fréquentes. Même les noms propres, comme *Jean-Louis Bourdon* étaient traduits de façon incorrecte. Par conséquent, Balkin a conclu que *Babelfish* était un bon outil pour avoir une idée générale de l'article, mais lorsqu'on ne connaissait pas la langue ni le sujet, cela posait beaucoup de problèmes de compréhension. Balkin a recommandé le recours à un traducteur

professionnel pour les traductions plus importantes (Balkin, 1999 : 57).

1.3.3 Olvera-Lobo et García-Santiago

En 2010, Olvera-Lobo et García-Santiago ont évalué les systèmes de traduction automatique *Google Translate*, *Prompt* et *Worldlingo*. L'efficacité des traducteurs en ligne était analysée sur la base d'un échantillon de textes en langue espagnole. Les auteurs ont recouru à des mesures d'évaluation objective et subjective, et une typologie des erreurs produites a été proposée. Il a été observé que les taux d'erreur pour les trois systèmes évalués étaient plus élevés dans les traductions relatives à la combinaison des langues allemand-espagnol. *Prompt* a été identifié comme le traducteur le plus fiable (en moyenne) pour les deux combinaisons linguistiques évaluées. Cependant, pour la combinaison espagnol-allemand, le traducteur *Google* a reçu une bonne évaluation (Olvera-Lobo et al., 2010 : 434). Les erreurs identifiées étaient divisées en cinq types : les erreurs se rapportant aux adverbes interrogatifs ou aux particules, les erreurs de syntaxe, les erreurs lexicales, les erreurs de préposition et les fautes de ponctuation (Olvera-Lobo et al, 2010 : 443). La plupart des erreurs (46,38 %) étaient de nature lexicale. Les auteurs ont aussi constaté que le logiciel de traduction automatique le plus approprié variait selon la langue et le type de phrase utilisée (Olvera-Lobo et al., 2010 : 451).

1.3.4 Hampshire et Salvia

Pendant la même année, Hampshire et Salvia ont conduit une étude sur les importantes contributions apportées au domaine de la traduction en ligne gratuite. Ils se sont rendu compte de l'importance d'un système de classement de ces traducteurs et ils ont évalué la performance de dix machines de traduction gratuites entre l'anglais et l'espagnol. Ces classements constituaient une première étape de l'évaluation, qui visait à amener l'utilisateur à faire un choix éclairé quant au traducteur le plus approprié pour son texte source (Hampshire et al., 2010 : 197). Tout comme Olvera-Lobo et son équipe, Hampshire et Salvia ont remarqué que le logiciel de traduction automatique le plus approprié variait selon le contexte d'utilisation. Le but était la création d'un système de classement pour aider les utilisateurs à choisir le logiciel gratuit de traduction automatique en ligne le plus utile à leurs fins. La décision de Hampshire et Salvia de créer un système de classement était basée sur les constats selon lesquels les classements répondent aux demandes d'informations facilement interprétables des consommateurs sur la qualité des traducteurs, fournissent une justification et un ensemble de critères pour évaluer les traducteurs, et contribuent à une définition de la qualité de la traduction automatique (Hampshire et al., 2010 : 199). Dans leur article, Hampshire et Salvia ont mentionné une étude dirigée par Savoy et Dolamic en 2009, dans laquelle 117 452 textes traduits du français vers l'anglais ont été analysés. La conclusion a été le classement suivant : *Google Translate* en première place, suivie de *Babel Fish* et *Prompt*. En ce qui concerne *Google Translate*, Savoy et Dolamic ont noté des problèmes liés à l'ambiguïté

lexicale et aux idiomes qui sont traduites mot à mot (Hampshire et al., 2010 : 201).

Pour évaluer les traducteurs, Hampshire et Salvia ont suivi l'exemple d'Aiken et son équipe, en utilisant la technique d'« *aller-retour* », processus consistant à traduire un texte vers une autre langue, puis à traduire le résultat vers la langue originale (parfois désignée « *back translation* » en anglais). Les textes originaux et finaux sont ensuite comparés. Hampshire et son équipe ont aussi utilisé l'approche dite de la *ressemblance humaine*, définie par López et Roca en 2006. Leur version de cette méthodologie était basée sur deux concepts relevant de la linguistique appliquée, à savoir la *clarté* et la *fidélité*. La *clarté* est définie comme « la facilité avec laquelle un lecteur peut comprendre la traduction » (Hampshire et al., 2010 : 202). La clarté est synonyme d'intelligibilité, de compréhension ou de ce qui est compréhensible. Autrement dit, moins l'évaluateur comprend, moins bonne est la qualité de la traduction. La *fidélité* est définie comme « la mesure dans laquelle le texte traduit contient les mêmes informations que l'original » (Hampshire et al., 2010 : 202). Les phrases analysées par Hampshire et Salvia visaient à évaluer une caractéristique précise d'une langue, c'est-à-dire que chaque phrase permettait d'évaluer la capacité de la machine gratuite en ligne de traduire un phénomène particulier dans la langue, comme un verbe à particule ou une ambiguïté lexicale. Pour évaluer la qualité des traductions de *Google Translate* et *Reverso*, une échelle de clarté allant de 5 (excellent) à 1 (médiocre) a été adoptée. Toutefois, en ce qui concerne la fidélité, Hampshire et Salvia ont choisi de déduire des points plutôt que d'en ajouter. Par

conséquent, si le texte est écrit parfaitement en anglais, mais le sens du texte ne peut être saisi du tout, cinq points sont déduits. Si la traduction transmet le sens du texte source, aucun point n'est déduit (Hampshire et al., 2010 : 202). Après avoir appliqué la technique *aller-retour* à cinq phrases cibles différentes (de l'anglais vers l'espagnol et de nouveau vers l'anglais), Hampshire et son équipe ont demandé à une petite équipe de cinq traducteurs professionnels d'évaluer le texte final en anglais produit par les dix machines, afin d'en déterminer la clarté et la fidélité et de parvenir à un accord en établissant une note sur cinq pour chaque phrase (Hampshire et al., 2010 : 203). Ils ont trouvé que *Google Translate* s'était classé au premier rang dans la traduction des phrases avec un registre formel, tandis que *Babylon* a dépassé *Google Translate* sur le plan des expressions idiomatiques. Cependant, il n'y avait encore pas de normes ou de critères définis dans les publications pour évaluer la qualité des machines de traduction gratuites. Néanmoins, Hampshire et son équipe ont suggéré que la performance d'un système de traduction automatique dépendait du type de texte à traduire. Par exemple, si le système de *Google* est arrivé en tête de classement pour ce qui est de la note globale, il a occupé la cinquième place en ce qui concerne les expressions idiomatiques. Par conséquent, les utilisateurs devraient décider quelle machine sélectionner en fonction du type de texte qu'ils veulent traduire (Hampshire et al., 2010 : 207).

Une tendance générale a été observée dans tous les classements de l'étude. Dans l'ensemble, *Google Translate* occupait les deux premières places. *WebTrance* et *InterTran* étaient invariablement les deux derniers systèmes. La

performance de *Reverso* était très irrégulière -- troisième dans l'ensemble, deuxième pour les expressions idiomatiques et huitième en ce qui concerne les verbes à particule. Hampshire et son équipe ont tiré la conclusion suivante : à l'aide d'un système multi-classement, les utilisateurs pouvaient utiliser la machine qui convenait à leur fin ; en fin de compte, certaines machines étaient mieux pour traduire les expressions idiomatiques et d'autres mieux indiquées pour le registre formel ; il n'était donc pas toujours nécessaire d'utiliser *Google Translate*, l'utilisateur devant choisir une machine selon le type de texte source à traduire (Hampshire et al., 2010 : 207).

1.3.5 Aiken et al.

Une cinquième étude similaire à celle de Hampshire et al. a été réalisée par Aiken et son équipe en 2012. Une réunion en ligne a eu lieu avec 40 étudiants qui ont échangé des phrases en 40 langues différentes en les saisissant dans l'ordinateur. Cette réunion a été possible grâce à *Polyglot*, un système de soutien qui fournit des traductions automatiques avec *Google Translate*. Environ 83 % du texte a été compris lorsqu'on l'a traduit vers l'anglais, ainsi que 90 % des commentaires dans six réunions en allemand, espagnol, français et italien (Aiken et al., 2012 : 25). Aiken et al. ont souligné que *Google Translate* fournissait des traductions entre le plus grand nombre de combinaisons linguistiques (1 640 combinaisons) et que cette machine était capable de donner des résultats très précis. Dans une évaluation de 22 systèmes de traduction automatique de l'arabe vers l'anglais et du chinois vers l'anglais, *Google* a été classé parmi les

trois premiers dans tous les cas d'essai. Ce niveau élevé de précision était en partie dû à son utilisation de l'apprentissage statistique, par opposition à une approche fondée sur les règles de traduction. Lorsqu'un mot avait plusieurs sens, la traduction correcte était déterminée en découvrant le sens le plus sémantiquement cohérent à partir d'un ensemble de phrases syntaxiquement acceptables. Le système de *Google Translate* fonctionne sur la base de documents des Nations Unies. Le nouveau texte est analysé en permanence, à des fins d'amélioration constante de l'apprentissage (Aiken et al., 2012 : 26). Aiken et al. ont noté que *Google* a obtenu de bons résultats pour la plupart des langues, mais que certains textes étaient mieux traduits que d'autres en raison de similitudes entre les langues utilisées ou à cause de la disponibilité des textes équivalents pour la modélisation (Aiken et al., 2012 : 27).

Les études d'Aiken et son équipe ont montré que certains groupes multilingues (en particulier ceux des pays d'Europe occidentale) devraient pouvoir utiliser *Polyglot* avec succès pour l'échange d'idées lors d'une réunion. Les étudiants ont estimé que le système de *Polyglot* était bon à cet effet et facile à utiliser. Lors d'une réunion en ligne avec des observations en allemand, français, espagnol et italien sur six thèmes différents, les participants pouvaient comprendre environ 90% du texte (Aiken et al., 2012 : 28). Bien que la compréhension ne soit pas parfaite, Aiken et al. étaient d'avis qu'une compréhension à 100 % n'aurait jamais été atteinte, même avec le texte anglais équivalent. Certaines des phrases comprenaient des termes inconnus et des expressions idiomatiques. Le sens global ou « l'essentiel » d'un commentaire

était ce qui avait de l'importance et dans la grande majorité des réunions électroniques, une traduction parfaite de la source de commentaire n'était souvent pas nécessaire où la vitesse et l'informalité étaient la règle. De nombreux membres du groupe préféraient avoir un peu de traduction, même de qualité médiocre, que pas de traduction du tout ou une meilleure traduction fournie trop tard. Ce niveau élevé de compréhension pendant ces réunions internationales était une source de motivation pour continuer à utiliser la traduction automatique à des fins de compréhension mutuelle entre les participants (Aiken et al., 2012 : 29).

En général, les logiciels de traduction sont loin d'être parfaits, mais ils semblent être une bonne solution pour faciliter la communication et la compréhension. Comme dans les travaux antérieurs, nous allons évaluer leurs sorties par le biais de l'analyse humaine. Dans le cadre de notre étude, nous considérons que les erreurs sémantiques et lexicales seront les plus nombreuses et entraîneront des difficultés liées à la désambiguïsation, comme l'ont indiqué Popović et Ney. Nous estimons que malgré ces lacunes, la traduction automatique au moyen des logiciels en ligne gratuits est plus que nécessaire aujourd'hui, en raison de leur disponibilité publique. Nous présumons aussi que chacun des systèmes aura ses propres forces et faiblesses et qu'il revient à l'utilisateur de choisir le système qu'il juge le mieux adapté à ses besoins. Dans le chapitre suivant, nous présenterons l'historique de la traduction automatique et montrerons comment les systèmes de traduction ont évolué au fil

des ans. D'autres systèmes que ceux déjà mentionnés seront également analysés.

CHAPITRE 2

La traduction automatique : d'hier à aujourd'hui

Dans ce chapitre, nous allons discuter de l'histoire de la traduction automatique ainsi que de l'avancement de la linguistique informatique et de l'intelligence artificielle. Le processus de traduction automatique et ses étapes vont être présentés. Ensuite, deux types de systèmes de traduction automatique vont être décrits : ceux qui sont basés sur le dictionnaire et ceux qui sont basés sur les statistiques. Enfin quelques systèmes de traduction automatique existants vont être examinés.

2.1 Historique

L'utilisation des dictionnaires mécaniques ou des dictionnaires lisibles par une machine pour surmonter les difficultés de traduction a été suggérée dès le 17^e siècle. Descartes et Leibniz ont spéculé sur la création de dictionnaires fondés sur des codes numériques universels. L'inspiration venait du concept d'« *universal language* », qui véhiculait l'idée de la création d'une langue sans équivoque, basée sur des principes logiques et des symboles iconiques, et avec laquelle toute l'humanité pourrait communiquer sans crainte de malentendu. Au cours des siècles suivants, d'autres propositions seront formulées pour les langues internationales, peu d'effort étant cependant fourni pour mécaniser la traduction, jusqu'au 20^e siècle. En 1933, deux brevets sont apparus en France et

en Russie, indépendamment l'un de l'autre. George Artsouni, un Français d'origine arménienne, a conçu une mémoire sur bande papier qui permettait de trouver l'équivalent de tout mot d'une autre langue. Un prototype de cette mémoire a été présenté en 1937. La proposition du Russe Petr Smirnov-Troyanskii était plus significative, en rétrospective. Smirnov-Troyanskii a envisagé une traduction mécanique en trois étapes : premièrement, un éditeur connaissant seulement la langue source entreprend l'analyse « logique » des mots dans leurs formes de base et fonctions syntactiques ; deuxièmement, une machine transforme les séquences des formes de base et des fonctions en suites équivalentes dans la langue cible ; enfin, un autre éditeur connaissant seulement la langue cible convertit ces données de sortie en formes normales de cette langue. Smirnov-Troyanskii croyait que le processus de l'analyse logique pourrait lui-même être mécanisé. Cependant, son modèle est resté inconnu hors de la Russie (Hutchins, 1992 : 5).

La possibilité d'utiliser les ordinateurs pour la traduction a été évoquée en premier par Warren Weaver du *Rockefeller Foundation* et Andrew D. Booth en 1947 (Aiken et al., 2010 : 26). Booth a étudié la mécanisation d'un dictionnaire bilingue et s'est mis à collaborer avec Richard H. Richens qui avait indépendamment commencé à utiliser les cartes à perforer pour produire des traductions mot pour mot de résumés d'articles scientifiques. Cependant, c'est un mémoire de Weaver mis au point en 1949 qui a porté l'idée de la traduction automatique à l'attention du public (Hutchins, 1992 : 5-6). Il a suggéré que les méthodes de décodage des messages en temps de guerre pourraient être

appliquées à la langue. À cette époque, l'ordinateur était capable de conserver une quantité de données suffisante pour produire une traduction mot à mot – ce qu'on pouvait obtenir avec l'aide d'un dictionnaire. Aucun compte n'était toutefois tenu de la syntaxe. L'idée était de présenter seulement des mots clés de la langue cible pour faire comprendre un texte. Le lecteur devait dépendre de sa connaissance du sujet pour pouvoir saisir le sens à travers des mots. Weaver a précisé que la traduction mot à mot pouvait être très utile pour la recherche scientifique et technique (Delavenay, 1960 : 28). En quelques années, la recherche a pris de l'essor dans plusieurs centres aux États-Unis et en 1951, Yehoshua Bar-Hillel du MIT a été désigné premier chercheur en traduction automatique (Hutchins, 1992 : 6).

Pendant cette période, la recherche a connu de grandes avancées, non seulement dans le domaine la traduction automatique, mais aussi dans celui de la linguistique informatique et de l'intelligence artificielle – particulièrement l'élaboration de dictionnaires informatisés et de techniques pour l'analyse syntactique (Hutchins, 1992 : 6). Les ordinateurs jouaient un rôle significatif depuis les années 50, lorsqu'ils sont devenus essentiels dans la vie économique et sociale. Ils ont ajouté trois nouvelles caractéristiques aux activités de recherche : une vitesse élevée d'opération, une complexité élevée des programmes et une extension des fonctions essentielles, la logique étant ajoutée à l'arithmétique. Ces trois caractéristiques de base ont créé la possibilité d'imiter des opérations d'intelligence particulières. Le premier changement révolutionnaire dans les machines informatiques a été l'introduction de la

mémoire, c'est-à-dire la capacité de conserver l'information (Delavenay, 1960 : 18). Pourtant, l'objectif fondamental de construire des systèmes capables de produire de bonnes traductions n'était pas encore atteint.

En 1964, le comité *ALPAC* (*Automatic Language Processing Advisory Committee*) a été établi aux États-Unis. Dans leur rapport de 1966, ils ont conclu que la traduction automatique était plus lente, moins précise et deux fois plus chère que la traduction humaine. Ils ont indiqué qu'il n'y avait pas d'avenir prévisible pour une traduction automatique utile. Ils ne voyaient aucun besoin d'investir plus d'argent dans cette recherche et leur rapport a plutôt recommandé la mise au point d'aides automatisées pour les traducteurs, comme les dictionnaires automatiques et le soutien à la recherche fondamentale en linguistique informatique (Hutchins, 1992 : 7).

Cependant, au cours de la décennie suivante, les recherches en traduction automatique ont essentiellement été effectuées hors des États-Unis, au Canada et en Europe occidentale. Un groupe de recherche établi à Montréal a réussi à construire un système pour traduire les manuels d'avion. Ce groupe est devenu célèbre en 1976 pour la création de *Météo*, un sous-langage pour la traduction des bulletins d'information météorologique (Hutchins, 1992 : 7). Un sous-langage est un langage d'un domaine restreint, en particulier un domaine technique. À travers cette importante évolution, la recherche dans ce domaine a gagné en maturité et conduit à la traduction automatique sur ordinateurs personnels, apparue en 1981 (Aiken et al., 2010 : 26). Au début des années 80, les linguistes ont repris la recherche et le développement dans le domaine de la

traduction automatique. À bien des égards, cette renaissance était reliée au développement mondial de l'industrie de l'informatique et plus particulièrement, à un intérêt croissant pour l'intelligence artificielle comme domaine d'application des ordinateurs (Sokolova, 2009 : 4).

L'évolution la plus significative des années 90 a été l'émergence des systèmes de traduction automatique commerciaux. Les produits américains d'*ALP Systems*, *Weidner* et *Logos* ont été suivis de systèmes japonais d'entreprises informatiques comme *Mitsubishi*, *Sanyo* et *Toshiba* (Hutchins, 1997 : 8). Presque tous ces systèmes dépendaient de la post-édition pour produire des traductions acceptables (Hutchins, 1997 : 9). En 1997, *Babel Fish* a été le premier système de traduction automatique gratuite sur Internet (Aiken et al., 2010 : 26). Ce système a ouvert la voie à plusieurs autres, tels que *Google Translate*, *Bing*, *Prompt*, et *Babylon*.

2.2 Processus de la traduction automatique

Un système de traduction automatique comprend des descriptions linguistiques des langues sources et cibles et un algorithme sur la base duquel la traduction proprement dite est effectuée. Le processus complet de traduction automatique comprend les étapes principales suivantes : 1) l'analyse des textes dans la langue source (la recherche des mots dans le vocabulaire, ainsi que l'analyse morphologique et syntaxique, c'est-à-dire la simulation de la compréhension du texte) ; 2) la conversion (le transfert à partir de la structure du texte dans la

langue source vers la structure du texte dans la langue cible) ; et 3) la synthèse du texte dans la langue cible (la présentation syntaxique et morphologique du texte, c'est-à-dire la simulation de la construction du texte). Dans les systèmes de traduction automatique réels, toutes ces étapes pourraient être intimement liées, et certaines absentes (ElShiekh, 2012 : 57).

La première étape soulignée par ElShiekh est très importante, car sans une analyse adéquate du texte dans la langue source, la conversion et la synthèse ne vont pas entraîner une traduction correcte. D'après L'Homme, afin d'analyser un texte, il faut tenir compte des morphèmes, des unités morphologiques significatives d'une langue qui ne peut être divisés. Les morphèmes aident à identifier la partie du discours d'un mot. Ils sont des éléments qui se combinent à certains mots pour en former de nouveaux (ex. – *ment* dans *traitement*, *-ique* dans *électronique*) (L'Homme, 2000 : 32). Les phrases sont composées d'unités linguistiques fonctionnelles appelées *syntagmes* et qui ont ensemble un sens. Toutes ces unités linguistiques ont une forme différente. Les unités linguistiques peuvent être divisées en catégories plus spécifiques (verbes, noms, adjectifs, etc.) qui peuvent à leur tour être subdivisées. Par exemple, les verbes sont subdivisés en catégories transitives et intransitives. Tout ceci montre que la description des unités linguistiques et le traitement de la langue pourront être des tâches compliquées (L'Homme, 2000 : 33). La nature du problème lié à la traduction est que les caractéristiques individuelles du texte traduit ne peuvent pas être ignorées (Delavenay, 1960 : 9-10). Il faut que la machine soit programmée avec des instructions particulières

concernant la substitution de la syntaxe du français à celle de l'anglais, l'ordre de certains mots, et la conjugaison des verbes. Les machines modernes peuvent faire tout cela, pourvu qu'elles puissent tenir compte de caractéristiques importantes dans une phrase écrite (Delavenay, 1960 : 25-26). Pour réussir à traduire un texte d'une langue à une autre, il faut tenir compte de la combinaison de toutes les unités qui caractérisent les deux langues (L'Homme, 2000 : 33).

Après avoir examiné les morphèmes, il faut tenir compte des mots entiers. Il y a quatre propriétés associées au mot qui sont pertinentes dans la traduction automatique : sa délimitation, son appartenance à une catégorie grammaticale, sa variation flexionnelle et son sens (L'Homme, 2000 : 34). La délimitation du mot est l'unité qui est associée à une suite de lettres comprises entre des espaces, des ponctuations ou d'autres symboles. La phrase suivante comporte six mots (exemples tirés de L'Homme, 2000) :

| Le | chien | assiste | à | la | scène |

Cependant, certaines entités graphiques (ou mot-formes) sont composées de deux mots contractés (ex. *au* → *à le*). Ainsi, la phrase suivante comporte elle aussi six mots (dans sa structure morphologique) :

| Le | chien | assiste | à | le | spectacle |

Les problèmes relatifs à la délimitation des mots varient d'une langue à l'autre. L'anglais utilise parfois des graphies différentes pour le même mot : *set up*, *set-up*, *setup*. Les catégories grammaticales comprennent le nom (*chien*), le verbe (*manger*), l'adjectif (*petit*), l'adverbe (*très*), le pronom (*je*, *lui*), le déterminant (*le*,

sa), la conjonction (*et, mais*) et la préposition (*de, avec*) (L'Homme, 2000 : 34-35).

Dans le traitement automatique de la langue, la catégorisation d'un mot n'est pas toujours claire, comme le montrent les exemples suivants :

Julien ferme (v.) la porte derrière lui.

Denis a la ferme (adj.) intention d'assister à la réunion.

La ferme (n.) de mon oncle se trouve dans le comté de Charlevoix.

La dame a ouvert un compte (n.) pour son chien.

Sophie compte (v.) jusqu'à 100.

Les mots *ferme* et *compte* doivent être interprétés différemment selon les contextes et peuvent produire des ambiguïtés. La première forme d'ambiguïté s'appelle *ambiguïté catégorielle*. Par exemple, lorsqu'une machine traduit la phrase, *Denis a la ferme intention d'assister à la réunion*, l'adjectif *ferme* peut être incorrectement traduit par le nom *farm* en anglais (L'Homme, 2000 : 37).

Avec la variation flexionnelle, le problème concerne les mots qui appartiennent à certaines catégories grammaticales. Un verbe comme *assister* peut apparaître dans des formes différentes selon le contexte :

Julien est allée à Toronto pour assister à la conférence...

...les personnes qui assisteront aux obsèques aujourd'hui...

...parents et amis assisteraient à cette rencontre... (L'Homme,

2000 : 39)

La machine doit être programmée pour reconnaître qu'il existe des formes diverses du même mot (L'Homme, 2000 : 39-40). Le français possède plus de flexions verbales que l'anglais, ce qui a tendance à augmenter l'ambiguïté catégorielle. Les problèmes d'identification de la délimitation, la catégorie grammaticale et la variation flexionnelle pourraient mener à des erreurs syntaxiques dans la traduction.

En ce qui concerne le sens, les mots qui semblent identiques peuvent revêtir des significations différentes.

Les généraux conduisent (v.) les troupes à la défaite.

Les métaux conduisent (v.) très bien l'électricité.

Bien que *conduisent* appartient à la même catégorie grammaticale dans les deux phrases, les deux formes ne peuvent pas être interprétées de la même manière. Dans la première phrase, *conduisent* renvoie à l'action de mener les gens ; dans la deuxième, *conduisent* signifie que le métal laisse passer l'électricité. L'ambiguïté ici concerne le sens des mots et c'est ce qu'on appelle *ambiguïté sémantique*.

Les mots qui ont la même forme mais des significations différentes s'appellent des *homographes sémantiques*, dans lesquels il faut distinguer deux cas : la polysémie et l'homonymie. La polysémie est caractérisée par la présence de composantes communes dans la définition, tandis que dans l'homonymie, il y a une absence de telles composantes. Il existe deux classes de mots ambigus d'homographes sémantiques : les homographes catégoriels et les

homographes sémantiques (L'Homme, 2000 : 42). La première classe concerne la catégorie grammaticale des mots :

Julien note (v.) qu'il manque des dents à ce peigne.

Apportez la note (n.), je vous prie.

Dans le premier exemple, *note* fait partie de la catégorie grammaticale « verbe », mais dans le deuxième exemple, *note* est catégorisé comme un nom.

La seconde classe concerne le sens des mots :

Il manque deux dents à ce peigne.

Les dents de votre chien sont en parfait état.

Dents dans le premier exemple renvoie aux petits éléments pointus d'un instrument qui sert à peigner les cheveux, et dans le deuxième, à des organes durs implantés dans la cavité buccale (TLF : 2013). Si elles ne sont pas résolues, ces ambiguïtés conduisent à des interprétations différentes (L'Homme, 2000 : 44) ainsi que des erreurs grammaticales et lexicales dans la traduction. À l'issue de cette analyse détaillée, il devient possible de faire la conversion et la synthèse d'un texte afin d'obtenir une traduction précise.

2.3 Deux types de systèmes de traduction automatique

Il existe deux types de systèmes de traduction automatique, à savoir les systèmes basés sur le dictionnaire et ceux qui s'appuient sur les statistiques. Les deux types de systèmes fonctionnent d'une manière assez différente, mais ils

sont tous les deux capables de fournir des traductions adéquates.

2.3.1 Systèmes de traduction automatique basés sur le dictionnaire

Presque tous les systèmes de traduction (encore appelés « traducteurs » ci-après) utilisent un dictionnaire pour effectuer la traduction (Nagao, 1989 : 129).

Les systèmes à base de règles sont très populaires dès les premiers jours de recherche de la traduction automatique. Ils s'appuient sur les règles de traduction manuscrites, leur permettant de saisir les relations complexes entre les textes source et cible avec une grande précision au prix d'une mise en œuvre et d'efforts considérables de maintenance (Hardmeier, 2012 : 4). Une des caractéristiques du dictionnaire à laquelle il convient de prêter attention en permanence est le maintien d'un niveau constant de qualité. Le véritable travail de compilation d'un dictionnaire d'un volume suffisant pour servir à la traduction automatique doit être réparti entre plusieurs personnes, mais les capacités du travailleur peuvent varier. Dans les zones nécessitant un certain degré de connaissances linguistiques en particulier, les différences individuelles seront inévitablement grandes. Pour éviter que de telles différences ne soient reflétées dans le dictionnaire, un « manuel de construction de dictionnaire » doit d'abord être mis au point et les données du dictionnaire ensuite saisies par une équipe de travailleurs guidés par ce manuel. Lorsque les travaux d'élaboration du dictionnaire sont achevés, les données lexicographiques peuvent être chargées dans l'ordinateur et utilisées dans les systèmes de traduction automatique à l'aide de « logiciels d'utilisation des dictionnaires ». Les mots qui sont

nécessaires dans un dictionnaire peuvent être divisés en deux grandes catégories : un vocabulaire fondamental des mots requis dans pratiquement tous les types de traduction, et une liste de termes spécialisés et techniques indispensables pour la traduction dans un domaine précis. Il est difficile de faire des généralisations sur le nombre de termes spécialisés dans un domaine donné, mais dans de nombreux cas, ce chiffre est de plusieurs dizaines de milliers, et atteint parfois cent mille (Nagao, 1989 : 129). Les systèmes basés sur des règles utilisent des informations linguistiques de la langue source et de la langue cible extraites de dictionnaires et de grammaires et englobant des règles sémantiques, morphologiques et syntaxiques de chaque langue respectivement. Leur utilisation est donc mieux indiquée pour les domaines spécialisés, tels que la terminologie médicale, qui contient beaucoup de mots aux significations particulières et qui sont déjà programmées dans le système (Geer, 2005 : 19-20).

Dans les systèmes de traduction basés sur des règles, un texte en langue source (LS) est converti par un mécanisme automatique qui utilise des dictionnaires pour le traduire en un texte en langue cible (LC). En général, les logiciels de traduction fonctionnent étapes, comme suit ((Ranchhod, 2000 : 254) :

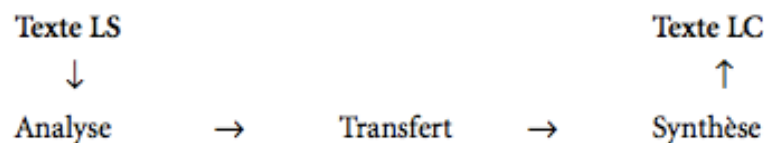


Figure 1. Le processus de traduction des systèmes automatiques

Dans ce processus, comme le montre la figure 1, le texte source est analysé, les unités significatives sont identifiées et leur structure syntaxique est formalisée. Ce processus comporte des dictionnaires et grammaires de la langue source et des programmes qui réalisent leur traitement informatique. Dans l'analyse, les unités significatives de la LS sont associées à des unités correspondantes de la LC. Les dictionnaires bilingues sont utilisés pour cette démarche. Le transfert se fait au niveau syntaxique de représentation dans la plupart des systèmes. Dans l'étape finale de synthèse, les unités de contenu sont organisées de nouveau, selon les règles de construction syntaxique de la LC (Ranchhod, 2000 : 254).

2.3.2 Systèmes de traduction automatique basés sur les statistiques

Contrairement aux systèmes de traduction basés sur un réseau formé d'un ou plusieurs dictionnaires, les systèmes de traduction automatique statistiques se caractérisent par l'utilisation de méthodes d'apprentissage automatique (Lopez, 2008 : 2). Cela implique que lorsque le traducteur génère une traduction, il recherche des modèles dans des centaines de millions de documents pour l'aider à décider quelle est la meilleure version à retenir (*Google*, 2013 : par. 2). Il est donc capable d'apprendre et de stocker de nouvelles informations automatiquement. Actuellement, la traduction automatique statistique est très populaire (Farrús et al., 2011 : 182). En moins de deux décennies, ce type de système a dominé la recherche théorique dans le domaine de la traduction automatique et il a conquis une part importante du marché commercial (Lopez, 2008 : 2). La traduction automatique statistique part du principe que la langue est

si riche et complexe qu'elle ne pourrait jamais être entièrement analysée et réduite en un ensemble de règles qui sont ensuite codées dans un programme de calcul. Au contraire, la nouvelle orientation consiste à formuler un système qui découvre les règles de la traduction automatiquement à partir d'un grand corpus de textes traduits, en combinant l'entrée et la sortie du processus de traduction et en apprenant à partir des statistiques sur les données (Koehn, 2010 : xi). La traduction est traitée comme un problème d'apprentissage de la machine. Cela signifie que l'on applique un algorithme d'apprentissage de langues à un grand nombre de textes déjà traduits, connus sous les noms de corpus parallèle, texte parallèle, bitexte ou multitexte (Lopez, 2008 : 2).

La traduction automatique statistique a pris beaucoup d'importance dans la communauté de recherche et dans le secteur commercial. D'après Koehn, à peu près mille thèses universitaires ont été publiées dans ce domaine, environ la moitié d'entre elles au cours des cinq dernières années seulement. En même temps, les systèmes de traduction automatique statistiques ont commencé à apparaître dans le marché. *Language Weaver* a été la première société établie purement pour la traduction automatique statistique (Koehn, 2010 : xi). Bien que la traduction entièrement automatique soit souvent considérée comme l'objectif à atteindre dans de nombreuses études sur la traduction automatique statistique, elle n'est ni la seule ni même le principal objectif probablement. Comprendre les besoins de traduction des utilisateurs est essentiel à l'amélioration continue des services de traduction automatique. L'intégration dans la reconnaissance vocale, la recherche d'informations, un résumé de documents, les questions et

les réponses, et d'autres applications du traitement des langues naturelles seront sans doute de plus en plus importante à mesure que la traduction automatique statistique prend sa place aux côtés d'un ensemble d'outils permettant d'accéder à l'information à l'échelle mondiale (Lopez, 2008 : 41).

Les systèmes statistiques travaillent généralement en divisant des phrases en n-grammes. Les n-grammes sont des séquences contiguës de n éléments à partir d'une séquence donnée de texte. Les articles peuvent être des phonèmes, des syllabes, des lettres, des mots, etc. L'analyse des n-grammes améliore la précision de la traduction, car si un mot possède plusieurs significations, on peut savoir à laquelle il se réfère en tenant compte de sa séquence donnée dans un texte. La plupart des systèmes traduisent par des trigrammes, d'après Franz Och, directeur du groupe de la traduction automatique de *Google* (Geer, 2005 : 19). Des groupes de trois mots sont assez grands pour rendre le processus efficace et sont susceptibles d'être répétés assez souvent dans les documents pour servir à l'analyse. En revanche, les plus grands ensembles de mots sont moins susceptibles d'être répétés et nécessitent plus de puissance informatique pour être analysés et traduits. Pendant l'entraînement, les systèmes statistiques mémorisent des n-grammes fréquents, apprennent quelles traductions sont les plus souvent utilisées, et les appliquent lorsqu'ils trouvent ces phrases à l'avenir. En outre, ils analysent statistiquement la position des n-grammes par rapport aux autres au sein des phrases, ainsi que les formes grammaticales de mots, pour déterminer la syntaxe correcte. Après leur

formation, les systèmes traitent des phrases et les enchaînent ensemble afin de traduire des documents entiers (Geer, 2005 : 19).

Avec une formation répétée et intensive, les développeurs peuvent rendre les systèmes statistiques plus précis. Par rapport aux systèmes à base de règles, lorsqu'on met le même montant d'argent dans la production des traductions, le système statistique donne de meilleures traductions, selon Stephan Vogel (Geer, 2005 : 20), chercheur associé avec l'institut de technologies langagières à l'université de Carnegie Mellon. Alors que le développement des systèmes complexes à base de règles pour les paires de langues peut durer plusieurs années, les chercheurs peuvent former les systèmes statistiques pour produire des traductions en quelques semaines ou quelques jours. Cela rend l'approche statistique moins laborieuse et plus utile pour les applications gouvernementales et d'entreprise à temps critique. Toutefois, cette approche nécessite également des ordinateurs puissants pour gérer la formation des traductions (Geer, 2005 : 20).

2.4 Quelques systèmes de traduction automatique existants

Il existe de nombreux systèmes compétents de traduction aujourd'hui, la majorité d'entre eux étant payants. Il s'agit de systèmes de traduction automatique commerciaux. D'autres systèmes relativement moins compétents deviennent de plus en plus populaires et sont disponibles gratuitement au grand public.

2.4.1 Systèmes de traduction automatique commerciaux

Trados est le chef de file mondial des solutions de traduction. Il permet aux entreprises multinationales d'améliorer la qualité et d'accélérer la livraison de contenus multilingues sur les marchés mondiaux. *Trados* a permis à 3.500 sociétés comme *Microsoft, IBM, Deutsche Bank, Wal-Mart, Audi, Sony, Rolls Royce* et d'autres de générer plus de recettes, de renforcer les marques et de réduire les coûts de la mondialisation. Basée à Sunnyvale (Californie), *Trados* possède des bureaux dans le monde entier (*Trados, 2005*). D'après *7Brands*, une société de traduction professionnelle, le logiciel est convivial et la plupart des traducteurs peuvent apprendre à l'utiliser sans formation formelle. Les inconvénients majeurs sont des virus informatiques qui restent en dépit des mises à jour fréquentes et coûteuses. Les traducteurs doivent être prêts à faire face à des pannes d'ordinateur conduisant à des heures de travail perdues, des problèmes de mémoire de traduction et un service client médiocre (*7Brands, 2013*).

Wordfast est le premier fournisseur mondial de technologie de mémorisation des traductions indépendante de la plate-forme, et le deuxième fournisseur mondial de solutions logicielles de mémoire de traduction. Les produits et les solutions offerts par *Wordfast* sont classés comme étant conviviaux, avec un des meilleurs soutiens fournis par le service client. Il offre des outils efficaces conçus pour répondre aux besoins des traducteurs, des fournisseurs de services linguistiques, des entreprises et des établissements

d'enseignement à travers le monde (*Wordfast*, 2012). *Wordfast* est aussi bon que *Trados*, avec l'avantage d'être beaucoup moins cher, offrant un excellent rapport qualité-prix, et facile à utiliser. Cependant, comme dans le cas de *Trados*, parfois ses outils ne fonctionnent pas comme ils devraient.

Le logiciel *Déjà Vu X* peut traduire un grand nombre de fichiers en même temps. Un projet peut être constitué de plusieurs centaines de fichiers, organisés suivant une structure de fichier précise. Avec *Déjà Vu X*, la structure du fichier d'origine peut être conservée et le logiciel peut traiter tous les fichiers à la fois. Le logiciel dispose également de nombreuses fonctionnalités qui permettent la révision de la traduction dans les cas où il est nécessaire de la modifier (*7Brands*, 2013). Avec la fonction de balayage, on peut vérifier non seulement comment une phrase ou une expression donnée a été traduite dans le passé, mais si oui ou non une phrase ou une expression similaire a également été traduite. *Déjà Vu X* affiche toutes les correspondances exactes ou floues de la phrase, soulignant les différences entre la phrase recherchée et les correspondances proposées (*Translation.net*, 2013).

Les trois traducteurs susmentionnés ne sont pas utilisés principalement par le grand public, car ils ne sont pas disponibles gratuitement, mais par beaucoup d'entreprises qui ont besoin de services de traduction.

2.4.2 Systèmes de traduction automatique gratuits

La fin des années 90 s'est caractérisée par l'émergence de traducteurs gratuits en ligne (Hampshire, 2010 : 197). Les quatre traducteurs qui font l'objet de notre évaluation offrent gratuitement la traduction instantanée d'expressions, de phrases, de documents et même de sites Web. Pour chacun des traducteurs qui nous intéressent, le processus de traduction s'effectue de manière différente. Pour cette raison, les résultats obtenus varient.

Le traducteur *Prompt* peut traduire entre sept langues différentes. Il utilise des dictionnaires pour accomplir la tâche de traduction. Les dictionnaires sont capables de stocker un nombre illimité de variantes de traduction (*Prompt*, 2013 : par. 2). D'après Sokolova, présidente et directrice générale de *Prompt* (Sokolova, 2009 : 2), le premier objectif de ces genres de systèmes est de créer un dictionnaire assez large, car plus le volume du dictionnaire est vaste, plus la traduction sera bonne. Il est aussi nécessaire d'apprendre au système à reconnaître les collocations communes. Enfin, puisqu'il existe des règles particulières pour la langue de départ et pour la langue cible, il faut stocker ces règles ensemble dans un programme. Les méthodes de disposition des grandes bases de données sont bien développées, mais pour la traduction, afin d'assurer la récupération correcte des éléments de base de données, il est plus important de savoir comment configurer les informations attribuées à chaque élément. Par exemple, en russe, jusqu'à 12 formes flexionnelles d'un nom peuvent exister et plus de 30 pour les verbes et les adjectifs. Par conséquent, afin de traduire des phrases avec des mots russes déclinables comme « programme », « à propos

du programme », « programmes », etc., il serait utile de mettre en œuvre une technique de recherche de corrélation entre l'entrée de « programme » contenue dans le dictionnaire et la forme du mot approprié dans le texte. Plus précisément, pour décrire les langues sources et cibles, le système doit utiliser une méthode formelle de description morphologique qui sert comme une base pour la recherche d'entrées lexicales (Sokolova, 2009 : 2). En ce qui concerne la base de données de ses systèmes, *Prompt* contient 300 formes de mots pour le français et 250 types pour l'anglais. La variété des affixes de chaque langue est stockée sous forme de structures arborescentes, ce qui permet d'obtenir non seulement un moyen de stockage efficace, mais aussi un algorithme efficace d'analyse morphologique. Par ailleurs, ce modèle morphologique a été appliqué à l'élaboration du système consultatif pour les utilisateurs qui créent eux-mêmes les dictionnaires (Sokolova, 2009 : 3).

Babylon, établi en 1997 (*Babylon*, 2013), est le premier fournisseur mondial de dictionnaires en ligne et de solutions de traduction (*Canada NewsWire*, 2008). Il est capable de traduire en 30 langues. Comme *Prompt*, il se sert d'entrées de dictionnaire pour réaliser les traductions. Il traduit gratuitement et instantanément les termes, les paragraphes et les documents (*Babylon*, 2013). Le stockage des entrées concerne non seulement les mots, mais aussi les phrases, les expressions et les informations grammaticales similaires à celles des dictionnaires électroniques traditionnels. La base de données de *Babylon* contient plus de 1 400 dictionnaires et glossaires qui proviennent de sources les plus variées, y compris l'*Oxford Dictionary*, le *Merriam Webster Dictionary* et

l'*Encyclopaedia Britannica*. La traduction englobe les termes techniques et l'argot (*Babylon*, 2011).

Google Translate est un service Web gratuit qui se retrouve dans presque toutes les recherches sur l'évaluation de la traduction automatique, notamment parce qu'il offre des traductions à partir et en direction des langues les moins répandues (Sanja, 2012 : 2143). Il fournit des traductions instantanées en 66 langues différentes. Il peut traduire des mots, des expressions, des phrases et des pages Web entre de nombreuses combinaisons de langues. Le but est de rendre l'information universellement accessible et utile, quelle que soit la langue dans laquelle elle est écrite (*Google*, 2013 : par. 1). *Google Translate* utilise le processus de traduction automatique statistique, recherchant des modèles dans des documents pour trouver la meilleure version. Les textes traduits proviennent de livres, d'organismes comme l'Organisation des Nations Unies et de sites Web de partout dans le monde. En détectant les modèles dans ces documents, *Google Translate* peut faire des suppositions intelligentes sur ce qui constituerait une traduction appropriée, en cherchant des modèles statistiquement significatifs. Ceux-ci sont des modèles d'équivalence entre le texte original et le texte cible qui ne sont pas susceptibles de se produire par hasard. Lorsque l'ordinateur trouve un modèle, il peut l'utiliser pour traduire des textes similaires à l'avenir. On peut dire, à première vue, que la traduction entre l'anglais et le français est suffisamment bonne. Cependant, pour d'autres langues où il n'existe pas beaucoup de documents traduits disponibles, il y a moins de modèles détectés par le logiciel. Les traductions étant générées par des machines, il n'est

pas rare que leur qualité laisse à désirer. Plus *Google Translate* analyse des documents humains dans une langue donnée, plus il produit une traduction de meilleure qualité. C'est pour cette raison que l'exactitude de la traduction varie parfois entre les langues (*Google*, 2013 : par. 2). Ce traducteur exige la participation de l'utilisateur. Ce dernier se voit proposer des choix de traductions possibles et il lui faut sélectionner le mot ou la phrase qui correspond à la meilleure traduction. La machine garde cette information dans le système pour la prochaine fois qu'un autre usager saisit un texte identique. Les sorties qui sont choisies le plus souvent comme étant la meilleure traduction par les utilisateurs vont apparaître le plus fréquemment comme résultat.

Le traducteur *Bing*, créé par un groupe de recherche de *Microsoft*, utilise un système de traduction automatique statistique pour fournir des traductions instantanées grâce à un ensemble de technologies flexibles et simples à utiliser. La technologie de traduction automatique de Microsoft Translator est le résultat de plus d'une décennie de travail à *Microsoft Research* (*Microsoft*, 2012 : 1). Le traducteur gratuit est capable de traduire en 41 langues différentes. Comme *Google*, il traduit des mots, des phrases et des sites Web entiers. Bien qu'il adopte un moteur de traduction différent de *Google* (Goldsborough, 2009 : 13), il recourt lui aussi au processus de traduction automatique statistique (*Microsoft*, 2013 : par. 2) et fonctionne également en analysant des documents qui ont déjà été traduits par des humains. Le traducteur demande aussi la participation de l'utilisateur. Ce dernier a le choix d'évaluer la traduction et de déterminer si elle est bonne, mauvaise ou inappropriée. La communication de ces informations aide le

système à savoir quelle est la meilleure traduction. La traduction qui est considérée comme la meilleure apparaîtra le plus souvent à l'avenir.

Tous ces systèmes de traduction en ligne sont utilisés par des millions de personnes chaque jour. Ils sont capables de produire des traductions, mais la question qui se pose est celle de savoir si leur qualité est suffisamment bonne. Chacun d'eux est en compétition pour fournir les meilleures traductions. L'augmentation de l'utilisation des outils gratuits de traduction en ligne a suscité un intérêt accru pour l'évaluation de ces outils (Seljan, 2011 : 332). Dans le chapitre suivant, nous allons procéder à une comparaison des logiciels de traduction en analysant les forces et les faiblesses de leur performance. Cela nous donnera une idée de l'efficacité de chaque système et nous permettra de formuler des recommandations éclairées concernant ces traducteurs.

CHAPITRE 3

Évaluation de quatre machines de traduction automatique

3.1 Présentation du corpus

Beaucoup de machines de traduction (aussi appelés « traducteurs ») sont disponibles gratuitement en ligne aujourd'hui. Une évaluation de ces machines est importante pour déterminer leur degré de fiabilité. Il se peut que certains d'entre eux soient plus efficaces à certains égards que d'autres. Si les publications sur les traducteurs automatiques commerciaux sont nombreuses, il existe seulement quelques études, principalement sous forme de blog, qui sont pertinentes pour évaluer et classer les traducteurs gratuits en ligne (Hampshire, 2010 : 197). Des paramètres d'évaluation automatique ont été récemment élaborés pour comparer la production d'une machine à une traduction de référence (Sanja, 2012 : 2144). Il existe une forte corrélation entre ces paramètres et ceux de l'évaluation humaine. Cette dernière est considérée comme un modèle, mais exige un travail énorme, fastidieux, subjectif et coûteux. Les paramètres d'évaluation automatiques sont généralement rapides et peu chers à mettre en application, car ils nécessitent seulement un système de traduction automatique de bonne qualité et ils ont un minimum d'exigences en matière de travail humain. Il n'est plus nécessaire d'avoir des évaluateurs ou analystes bilingues. Toutefois, les mesures actuellement utilisées ne font pas de distinction significative entre des systèmes de traduction automatique qui sont

très semblables. De plus, elles donnent des résultats plus fiables sur l'ensemble des données en cours d'évaluation que sur les phrases individuelles (Sanja, 2012 : 2144).

Dans le présent travail, nous allons analyser la qualité de la traduction de *Prompt*, *Babylon*, *Google Translate* et *Bing*. Ces machines ont été choisies parce qu'elles sont assez populaires, disponibles en ligne gratuitement et accessibles à des millions de personnes quotidiennement. Nous mesurerons le degré de précision de leur traduction de l'anglais vers le français sur la base du nombre d'erreurs produites, en utilisant comme référence des textes traduits professionnellement par des humains et réunis en un corpus dénommé *BankGloss* (voir l'appendice A). Nous noterons les caractéristiques positives de chaque système, puis nous identifierons leurs erreurs. La proportion de chaque type d'erreur produite va être déterminée, qu'il s'agisse d'une erreur grammaticale, lexicale, sémantique, stylistique ou orthographique. Ces types sont subdivisés en sous-types, comme suit (voir l'appendice C) ;

Les erreurs grammaticales sont notamment :

- Les fautes d'accord (article-nom, sujet-verbe, pronom-antécédent, nom-adjectif ou participe passé)
*ex. Banque Mondial**
- Les mots appartenant à une mauvaise partie du discours
ex. la fourniture au lieu de fournissant

- Les fautes de conjugaison (choix de temps, concordances des temps)
ex. *Nous sommes allé**
- Les omissions de mots appartenant à des parties de discours différentes (déterminant, nom, verbe, préposition etc.)
ex. *IFC accompli** (omission de déterminant)
- Les fautes de ponctuation
ex. *le pouvoir de l'état utilitaire** (majuscule à « État »)
- Le mauvais ordre des mots
ex. *Un pouvoir indépendant projet**
- L'utilisation de prépositions ou d'articles erronés
ex. *dans neuf mois**

Les erreurs lexicales concernent :

- Les choix de termes inappropriés
ex. *la capitale* sur les marchés financiers* (si le nom devrait être au masculin : le capital)
- Les barbarismes (mots inexistants)
ex. *Le chat kharé le chien**
- Les régionalismes
ex. *Pendant l'été, elle veut une coupe glacée**
- Les anglicismes
ex. *développer des produits**

- Les mots de registres inappropriés
ex. « *Professeur, t'a déjà corrigé les examens ?* »*

Les erreurs sémantiques sont :

- Les imprécisions
ex. *La Société de Finance* (au lieu de *La Société financière*)
- les ambiguïtés
ex. *Le pilot ferme la porte*
- les sous-traductions
ex. *The tasty roast beef was tender and juicy* → *Le rosbif était savoureux et tendre**
- les sur-traductions
ex. *He dislikes sports* → *Il déteste les sports**
- Les ajouts
ex. *Nous sommes modifie**
- Les faux-sens
ex. *Il déplace stratégiquement nos affaires**
- Les non-sens
ex. *She participated in the rally* → *Elle est allée au musée**
- Les contresens
ex. *The information was not given to the attendants* → *L'information a été donnée aux assistants**

Dans le présent travail, la définition de « faux-sens », « non-sens » et « contresens » est celle qu'en donne Delisle (Dussart, 2005 : 108), comme suit : un faux-sens est une « faute de traduction qui consiste à attribuer à un mot ou à une expression du texte de départ une acception erronée qui altère le sens du texte, sans pour autant conduire à un contresens ». Un non-sens dénature complètement le sens du texte de départ et n'a rien à voir avec ce texte (Dussart, 2005 : 109). Un contre-sens est « l'interprétation erroné d'un mot, d'une phrase » et transmet un sens contraire au texte de départ (Dussart, 2005 : 110).

Les erreurs stylistiques couvrent principalement :

- Les fautes d'expression
*ex. It's raining cats and dogs → Il pleut des balles**
- l'absence d'idiomaticité
*ex. Tomber en amour**
- Les constructions boiteuses ou peu heureuses
*ex. Pour Noël, il voulait un nouveau chandail, il voulait des nouvelles chaussures, il voulait deux jeux vidéos, etc.**

Enfin, les erreurs d'orthographe couvrent, entre autres :

- les fautes d'épellation, c'est-à-dire les écritures erronées
*ex. Ils ont appris beaucoup sur le sujet**
- Les mauvais accents ou absence d'accent
*ex. Le pays veut développer son secteur prive**

Dans le chapitre suivant, nous discuterons des causes de ces erreurs et proposerons des solutions. Nous procéderons également au décompte du nombre de divergences produites par rapport à la traduction de référence de *BankGloss*. Le nombre de divergences est un indice de la qualité de la traduction de la machine ; un plus grand nombre de divergences signifie que la qualité est médiocre, tandis qu'un plus petit nombre est signe de meilleure qualité. Nous allons calculer la proportion de divergences pour chaque logiciel de traduction en divisant le nombre de divergences commises par 405 (le nombre de mots en total dans les extraits) (voir l'appendice B). Cette proportion en forme de pourcentage nous indiquera le niveau de qualité de la traduction. Les dix extraits sont des textes assez techniques tirés de *BankGloss*, une compilation de traductions alignées par paragraphes avec leur version originale, et qui servent de référence au sein de l'Unité de traduction et d'interprétation de la Banque mondiale. L'Édition de *BankGloss* utilisée est celle de 2003. Le choix de ce genre de textes tient au fait que la demande de la traduction de documents techniques est plus élevée par rapport aux autres types de textes. Nous avons pris soin de réunir des textes comprenant des termes et des structures assez variés, afin d'obtenir des résultats précis et d'effectuer en définitive une analyse plus précise.

3.2 Analyse de la performance des machines

3.2.1 Promt

Le traducteur *Promt* s'est classé dernier avec un total de 65 erreurs et 150 divergences (par rapport à *BankGloss*). Ce système a fait considérablement moins d'erreurs grammaticales que *Babylon* et *Bing*, ce qui indique que le système fonctionne déjà sur la base d'un nombre adéquat de règles de grammaire. Il n'y a eu qu'une faute de conjugaison où l'impératif a été utilisé au lieu de l'infinitif. Un exemple de passage où, contrairement à tous les autres, ce système n'a pas commis d'erreur grammaticale de conjugaison se trouve dans l'extrait 9 :

Extrait 9 :

anglais : *we roll up our sleeves and try to do it*

BankGloss : nous retroussons nos manches et

nous nous attelons à la tâche

Promt : nous enroulons nos manches et essayons de le faire

Babylon : nous retrousser nos manches et de tenter de le faire

Google : nous retroussons nos manches et essayer de le faire

Bing : nous retroussons nos manches et essayez de le faire

On note ici que *Promt* est capable de prendre en considération le contexte du verbe. Il reconnaît que le verbe *essayer* est lié au pronom *nous* tandis que la plupart des autres systèmes identifient le verbe hors de son contexte. En

général, on observe que par rapport aux autres traducteurs, *Prompt* contient plus de règles grammaticales permettant une conjugaison correcte des verbes.

Prompt a commis 32 erreurs grammaticales, 24 erreurs sémantiques, 6 erreurs lexicales, 3 erreurs stylistiques et 0 erreur d'orthographe. Les erreurs produites ont donc été principalement de nature grammaticale, sous forme notamment d'omission de mots. Les erreurs lexicales ont été principalement des anglicismes. On observe des erreurs sémantiques significatives dans les extraits 2, 5 et 9 et 10. Il faut aussi noter deux erreurs stylistiques produites dans l'extrait 9.

Extrait 2 :

anglais : *IFC is strategically shifting our business*

BankGloss : La SFI a pour stratégie d'orienter son action

Prompt : **IFC déplace stratégiquement nos affaires**

Babylon : La SFI est stratégiquement passage notre entreprise

Google : La SFI est stratégiquement passer, dans notre
entreprise

Bing : SFI est stratégiquement changeante de notre
entreprise

Dans cet exemple, l'erreur sémantique produit un faux-sens à cause de l'utilisation du substantif « affaires » pour traduire *business*. S'il est vrai que « business » peut se traduire dans certains contextes par « affaires », tel n'est manifestement pas le cas ici. Les équivalents appropriés dans ce contexte sont plutôt « opération(s) », « activité(s) », « action(s) » ou même « entreprise » (dans

le sens de « mise en œuvre de capitaux et d'une main-d'œuvre salariée en vue d'une production ou de services déterminés » (*TLFi*).

Extrait 5 :

anglais : *while flattening hierarchical structures*

BankGloss : en aplatissant la structure hiérarchique

Prompt : **en écrasant des structures hiérarchiques**

Babylon : en aplatissement structures hiérarchiques

Google : en aplatissant les structures hiérarchiques

Bing : en aplatissant les structures hiérarchiques

L'utilisation d'*écraser* pour traduire l'anglais *flattening* produit un faux-sens dans ce contexte. « Écraser » signifie « aplatir, broyer (une chose, un être vivant) en exerçant une forte pression, sous l'effet d'un choc violent » (*TLFi*). Puisque l'objet direct du verbe est « des structures hiérarchiques », une entité non-physique, l'utilisation de ce verbe n'est pas justifiée. On devrait plutôt recourir au verbe « aplatir », qui signifie « modifier la forme d'une chose de façon à la rendre plate » (*TLFi*).

Extrait 9 (exemple A) :

anglais : *we roll up our sleeves*

BankGloss : nous retroussons nos manches

Prompt : **nous enroulons nos manches**

Babylon : nous retrousser nos manches

Google : nous retroussons nos manches

Bing : nous retroussons nos manches

Ici, l'utilisation du verbe « enrouler » produit une absence d'idiomaticité, car on ne l'utilise pas dans cette expression qui signifie « remonter ou rouler le bas de ses manches pour dégager les bras » (*TLFi*). Pour cette expression, le verbe « retrousser » est utilisé.

Extrait 10 :

anglais : *Ceará reaped \$880 million*

BankGloss : ...a rapporté à l'État 880 millions de dollars

Prompt : **Ceará a moissonné \$880 millions**

Babylon : Ceará récolté 880 millions de dollars

Google : Ceará a récolté 880 millions de dollars

Bing : Ceará engrangé 880 millions \$

L'utilisation du verbe « moissonner » n'est pas appropriée dans cette situation parce qu'il ne s'agit pas d'une moisson au sens littéral. On devrait plutôt utiliser « récolter », comme l'ont fait *Babylon* et *Google*, qui signifie « obtenir quelque chose comme conséquence de l'action, de la conduite, etc. » (*TLFi*) ou « engranger », comme traduit par *Bing*, qui signifie « mettre en réserve pour un usage ultérieur » (*TLFi*). *BankGloss* le traduit par « rapporter », ce qui est le meilleur choix dans ce contexte.

L'extrait 9 contient deux exemples d'erreurs stylistiques :

Extrait 9 (exemple B) :

anglais : *it takes two to tango*

BankGloss : il faut être deux pour avancer

Prompt : **il prend deux au tango**

Babylon : il faut être deux pour danser le tango

Google : il faut être deux pour danser le tango

Bing : il faut être deux pour danser le tango

L'expression idiomatique *it takes two to tango* est traduite incorrectement, d'une manière littérale, ce qui produit une erreur stylistique sous forme d'absence d'idiomaticité. De plus, il s'agit d'une traduction littérale de l'anglais qui est incompréhensible pour un francophone. Une meilleure traduction, comme celle donnée par *BankGloss* est « il faut être deux pour avancer », ce qui renvoie à la même signification que l'expression anglaise.

Dans l'exemple qui suit, l'utilisation du substantif « endroit » au lieu de « place » constitue un autre exemple d'absence d'idiomaticité qui résulte d'une traduction littérale.

Extrait 9 (exemple C) :

anglais : *set in place*

BankGloss : mettre en place

Prompt : **mettre dans l'endroit**

Babylon : mettre en place

Google : mettre en place

Bing : mettre en place

« Mettre en place » est une expression figée, ce qui signifie que le remplacement

d'un de ces éléments par un synonyme aboutit à un barbarisme. Cela explique pourquoi « endroit », bien que synonyme de « place », ne peut être utilisé ici.

Tous ces exemples montrent des cas où *Prompt* était le seul traducteur à commettre une erreur idiomatique et sémantique en particulier. Le nombre élevé d'erreurs et de divergences par rapport à la traduction de *BankGloss* indique non seulement que la traduction ne rend pas le sens du texte original, mais aussi que la qualité de la langue utilisée n'est pas au même niveau que celle des autres traducteurs.

3.2.2 Babylon

Le traducteur *Babylon* s'est classé troisième avec 65 erreurs et 138 divergences. Le nombre d'erreurs est le même que chez *Prompt*, mais le premier avait plus de divergences. Globalement, ce traducteur, ainsi que *Google*, a commis le moins d'erreurs sémantiques. Dans certains cas, *Babylon* a fourni la meilleure traduction du type sémantique, comme dans les exemples ci-après :

Extrait 5 :

anglais : *our development effectiveness*

BankGloss : l'efficacité de notre action en faveur du
développement

Prompt : notre efficacité de développement

Babylon : notre efficacité en matière de développement

Google : l'efficacité de notre développement

Bing : notre efficacité de l'aide

Tous les autres traducteurs ont produit un faux-sens lors de la traduction de cette phrase. L'utilisation de l'expression « en matière de » par *Babylon* rend sa traduction correcte.

Extrait 6 :

anglais : *delegation of decision making downward*

BankGloss : déléguer le pouvoir de prise de décision à tous les niveaux

Prompt : la délégation de prise de décision en bas

Babylon : la délégation de la prise de décision vers le bas

Google : la délégation de la prise de décision à la baisse

Bing : la délégation de la prise de décisions à la baisse

Dans cet extrait, les autres traducteurs ont tous produit des erreurs sémantiques en traduisant littéralement l'adverbe anglais *downwards*. *Babylon* est le seul qui a réussi à traduire cet adjectif sans produire un faux-sens. Son utilisation de la préposition « vers » renvoie au sens du texte original.

Dans le même extrait, *Babylon*, comme *Google*, n'a omis aucun mot :

anglais : *IFC is also seeking*

BankGloss : la SFI s'efforce aussi

Prompt : IFC cherche aussi

Babylon : la SFI s'efforce également

Google : l'IFC cherche également

Bing : IFC cherche également

Babylon a traduit *IFC* correctement dans tous les extraits avec son utilisation de l'article défini. Sa traduction est aussi cohérente et constante dans l'ensemble du corpus à cet égard ; il y a des cas où la traduction de *Prompt* a alterné entre *IFC* et *l'IFC*. *Google* et *Bing* avaient tendance à alterner entre les acronymes *SFI* et *IFC*.

Babylon a commis au total 42 erreurs grammaticales, 16 erreurs sémantiques, 6 erreurs lexicales, 1 erreur stylistique et 0 erreur d'orthographe. En ce qui concerne le type grammatical, *Babylon* a surtout eu des omissions de mots ; pour ce qui est du type sémantique, on y relève principalement des faux sens ; s'agissant du type lexical, a surtout commis des fautes d'anglicisme ; eu égard type stylistique, on n'y note qu'une faute de construction boiteuse.

Dans l'extrait 1, *Babylon* a commis plus d'erreurs grammaticales que les autres systèmes. Voici un exemple de faute de conjugaison :

Extrait 1 :

anglais : *fosters economic growth*

BankGloss : favorise la croissance économique

Prompt : encourage économique la croissance

Babylon : encourager économique la croissance

Google : favorise l'économie la croissance

Bing : favorise l'économie croissance

Ici, le verbe « encourager » ne devrait pas être à l’infinitif, mais au présent de l’indicatif.

Extrait 10 :

anglais : *an independent power project*

BankGloss : un projet de production d’énergie indépendante

Prompt : un projet de pouvoir indépendant

Babylon : un pouvoir indépendant projet

Google : un projet énergétique indépendant

Bing : un projet d’électricité indépendants

Il y a trois erreurs ici : une faute grammaticale due à une antéposition du nom et de l’adjectif qualificatif (en supposant que « indépendant se rapporte à projet ») ; un faux sens lié à la traduction de « power » par « pouvoir » au lieu d’« électricité » ; et une ambiguïté au niveau du nom auquel se rapporte l’adjectif qualificatif « indépendant » (il pourrait aussi bien décrire « pouvoir » que « projet »).

Il faut aussi noter quelques exemples d’erreurs sémantiques, comme dans l’extrait 1 :

anglais : *advice to governments and businesses*

BankGloss : des services-conseils aux gouvernements et aux entreprises

Prompt : le conseil aux gouvernements et aux entreprises

Babylon : le conseil aux têtes folles et aux entreprises

Google : des conseils aux gouvernements et aux
entreprises

Bing : conseils aux gouvernements et aux entreprises

Au lieu de « gouvernements », *Babylon* traduit *governments* par « têtes folles », ce qui est un non-sens.

Par rapport aux autres traducteurs, *Babylon* a commis un nombre significatif d'erreurs grammaticales. Ceci peut indiquer que le système ne possède pas assez de règles relatives à la ponctuation, l'utilisation de prépositions ou d'articles et la structure syntaxique des phrases. En général, les erreurs sémantiques ont été minimales, mais l'erreur dans l'exemple ci-dessus était absolument inadmissible.

3.2.3 Bing

Le traducteur *Bing* s'est classé deuxième avec un total de 64 erreurs et 136 divergences. Ce système a fait très peu d'erreurs lexicales (seulement 3). Ces erreurs sont des anglicismes, par exemple « l'International Finance Corporation » dans l'extrait 1. Un exemple où *Bing* ne commet pas la même erreur grammaticale que les autres traducteurs se trouve dans l'extrait 2 :

anglais : "frontier markets"

BankGloss : « marchés pionniers »

Prompt : "marchés de frontière"

Babylon : "marchés frontaliers"

Google : «marchés pionniers»

Bing : « marchés frontières »

Ici, la traduction de *Bing* est la seule qui ne contient pas de faute de ponctuation au niveau des guillemets. Les autres traducteurs ont utilisé des guillemets anglais et n'ont pas d'espace entre le guillemet et le mot.

Le système *Bing* a produit 42 erreurs grammaticales, 18 erreurs sémantiques, 3 erreurs lexicales, 1 erreur stylistique et 0 erreur d'orthographe. En ce qui concerne le type grammatical, il y a eu principalement des omissions de mots ; pour ce qui est du type sémantique, on a relevé surtout des faux-sens ; eu égard au type lexical, le système n'a commis que des fautes d'anglicisme.

Comme *Babylon*, *Bing* a commis un nombre significatif d'erreurs grammaticales. Dans l'extrait 10, le système a commis 13 erreurs grammaticales (voir l'appendice A), dont 7 cas d'omission de mots :

Extrait 10 :

anglais : *privatize the state power utility*

BankGloss : privatiser la compagnie d'électricité de l'État

Prompt : privatisez l'utilité de pouvoir publique

Babylon : privatiser le pouvoir de l'état utilitaire

Google : la privatisation de la puissance publique de l'état

Bing : privatiser l'électricité d'État

Le nom « *utility* » et le déterminant avant « État » constituent deux exemples d'omission de mots. *Bing* semble avoir du mal à savoir quand est-ce qu'il faut ajouter un article. Cette absence de mots peut indiquer que le logiciel ne reconnaisse pas certains mots comme étant des noms.

Bing a commis également plusieurs fautes grammaticales de conjugaison.

L'extrait 8 en donne un exemple :

anglais : *IFC may take*

BankGloss : La SFI peut prendre

Prompt : IFC peut prendre

Babylon : La SFI peut prendre

Google : La SFI peut prendre

Bing : Ceci pourrait prendre

Ici, *Bing* est le seul traducteur qui a commis ce type d'erreur. Il est possible que le système ne contienne pas assez de règles de conjugaison des verbes. *Bing* semble incapable de reconnaître le pronom ou le sujet et de d'effectuer des conjugaisons correctes de manière systématique.

On y relève quelques erreurs de type grammatical, comme dans l'extrait

3 :

anglais : *IFC's role will be to fuel innovation*

BankGloss : Elle devra stimuler l'innovation

Prompt : Le rôle d'IFC devra alimenter l'innovation

Babylon : La SFI aura son rôle à jouer pour stimuler l'innovation

Google : Rôle de la SFI sera de stimuler l'innovation

Bing : Rôle de la SFI sera de carburant innovation

Le système ne reconnaît pas que *fuel* est un verbe ici, et pense plutôt qu'il s'agit d'un nom.

Bing affiche aussi un nombre important de faux-sens (14 en total). On en note un exemple dans l'extrait 7 :

anglais : *investment departments*

BankGloss : départements des placements

Prompt : départements d'investissement

Babylon : services de placement

Google : services de placement

Bing : investissement ministères

S'il est vrai que *department* peut se rendre par « ministère », une analyse contextuelle montre qu'il s'agit ici d'un département ou d'un service. *Bing* commet une double erreur dans cet extrait : un faux-sens avec le choix de « ministère » au lieu de « service » ou « département » ; et une erreur grammaticale, car le déterminant aurait dû être postposé, de sorte qu'on ait « ministères d'investissement ».

3.2.4 Google Translate

Google Translate s'est classé premier avec seulement 49 erreurs et 131 divergences. *Google* a produit le moins de types d'erreurs. L'extrait 1 donne un

exemple où *Google* ne commet pas la même erreur sémantique que les autres logiciels :

anglais : *The International Finance Corporation, part of the World Bank Group*

BankGloss : La Société financière internationale, qui est membre du Groupe de la Banque mondiale

Prompt : La Société de Finance internationale, la partie du Groupe de Banque Mondial

Babylon : La société de finance internationale, la partie du Groupe de la Banque Mondial

Google : La Société financière internationale, qui fait partie du Groupe de la Banque mondiale

Bing : Partie du groupe de la Banque mondiale, l'International Finance Corporation

La manière dont le mot « partie » est utilisé par les autres systèmes produit un faux-sens. La traduction correcte de *BankGloss* utilise le nom « membre », qui signifie « élément constitutif ou portion d'un tout » (*TLFi*). La définition de « partie » est similaire : « élément constitutif ou portion d'un tout, considéré(e) dans ses rapports avec ce tout ». Cependant, « partie » est un faux-sens dans ce contexte s'il est utilisé avec un article défini, car il s'agit d'une expression figée, « faire partie », comme l'a compris *Google*.

Le traducteur *Google* a commis 30 erreurs grammaticales, 16 erreurs sémantiques, 2 erreurs lexicales, 1 erreur stylistique et 0 erreur d'orthographe.

De tous les types de fautes, celui qui apparaît le plus souvent est l'erreur grammaticale, particulièrement les omissions de mots. Sur le plan sémantique, le système a commis le plus d'erreurs de faux-sens. Enfin, les anglicismes sont les erreurs lexicales les plus courantes.

L'extrait 5 contient une erreur grammaticale :

anglais : *We are eliminating*

BankGloss : Nous avons entrepris d'éliminer

Prompt : Nous éliminons

Babylon : Nous éliminons

Google : Nous sommes en éliminant

Bing : Nous éliminons

Google traduit incorrectement la phrase en ajoutant l'auxiliaire « sommes » au lieu d'utiliser simplement le présent de l'indicatif comme le font les autres systèmes. Cet ajout est résultat de la traduction directe du verbe anglais « are ». Le verbe « éliminer » n'est pas conjugué correctement non plus.

On remarque aussi un manque de cohérence de *Google* dans la traduction de *IFC*. Par exemple, ce sigle est traduit par « IFC » dans l'extrait 6, mais par « SFI » dans l'extrait 7.

Le système manque de capacité de reconnaissance ou d'analyse du contexte, comme le montrent les extraits 2 et 4 :

Extrait 2 : *IFC is strategically shifting our business toward “frontier markets” and developing innovative products for these regions of the world.*

La SFI est stratégiquement passer, dans notre entreprise vers les «marchés pionniers» et développer des produits innovants pour ces régions du monde.

Extrait 4 : *An analysis of the Corporation revealed that we have both the capital and the know-how to succeed in these frontier countries.*

Une analyse de la Corporation a révélé que nous avons à la fois la capitale et le savoir-faire pour réussir dans ces pays frontaliers.

Dans l'extrait 2, le système traduit le mot anglais *frontier* par « pionniers », mais dans l'extrait 4, il le rend par « frontaliers ». Dans le premier cas, le terme modifie *markets*, d'où la traduction « marchés pionniers » ; dans le deuxième cas, il modifie *countries* et le système le traduit « pays frontaliers ». Il faut que la machine reconnaisse le contexte des mots, en les identifiant comme appartenant à des groupes de mots ou au paragraphe entier.

Chez *Google* et *Bing*, on observe un manque de cohérence au niveau des signes de ponctuation, en ce qui concerne par exemple les guillemets ainsi que les signes de pourcentage et de dollar. Pour ce qui est des guillemets, les formes utilisées sont tantôt celles de l'anglais, tantôt celles du français (voir l'extrait 9).

Eu égard au signe de pourcentage, parfois « % » est utilisé et d'autres fois on voit la forme écrite « pour cent » (voir les extraits 8 et 10). En ce qui concerne le signe du dollar, on observe un mauvais placement dans l'extrait 7: « de \$3 milliards à 9 milliards \$ depuis 1990 ». Enfin, s'agissant de la traduction d'*IFC*, la plupart du temps il y a un manque d'article, mais dans un seul cas (dans l'extrait 8), l'article défini est présent. On peut ainsi conclure qu'un manque de cohérence au niveau des signes de ponctuation caractérise l'ensemble des systèmes de traduction statistique.

L'évaluation ci-dessus a contribué à déterminer les forces et les faiblesses de chaque machine. En cernant les insuffisances précises de chacun des systèmes, il devient possible de trouver des solutions concrètes permettant d'y remédier. Les constatations de l'évaluation vont donc aider à améliorer les règles qui sont déjà programmées dans les systèmes. Le prochain chapitre traite de la typologie des erreurs. Les différents types d'erreurs vont être analysés plus en détail. Des hypothèses sur les causes des erreurs seront ensuite formulées.

CHAPITRE 4

Analyse des erreurs de traduction automatique dans quatre systèmes

Le chapitre précédent a traité de quatre machines de traduction et des erreurs produites par chacune d'elles. Dans le présent chapitre, nous allons déterminer quels sont les types d'erreurs produits par ces machines de traduction (voir l'appendice C). Cela permettra de mieux connaître les problèmes de traduction automatique dans leur ensemble. Nous noterons aussi quelques différences entre la qualité des traductions effectuées par des systèmes basés sur les dictionnaires et par ceux reposant sur la statistique.

4.1 Typologie des erreurs

Les erreurs de traduction présentes dans les extraits sont divisées en quatre types généraux : grammatical, lexical, sémantique et stylistique. (Il n'y avait pas de fautes d'orthographe produites en traduisant les documents de *BankGloss*.) Certains des sous-types ont été présents seulement chez certains traducteurs.

Les quatre systèmes de traduction utilisés dans cette étude ont produit un total de 243 erreurs : 147 erreurs grammaticales, 17 erreurs lexicales, 74 erreurs sémantiques et 6 erreurs stylistiques (voir l'appendice B). Comme on peut le voir, la plupart des erreurs produites par les systèmes de traduction sont grammaticales de nature. Les différents types d'erreurs et leurs sous-types vont maintenant être examinés, à l'exception des erreurs d'orthographe qui sont absentes, tel que souligné au paragraphe précédent).

4.1.1 Erreurs grammaticales

Les systèmes ont commis au total 147 erreurs grammaticales, dont 39 fautes d'omission, 23 de ponctuation, 21 de conjugaison, 21 de prépositions ou d'article, 19 de mauvaise ordre des mots, 12 d'accord et 12 cas d'utilisation de mots appartenant à une mauvaise partie du discours. La plupart des erreurs commises par les systèmes de traduction sont donc celles d'omission de mot, comme un déterminant ou une préposition. L'extrait 10 nous donne un exemple où tous les traducteurs ont fait la même erreur dans ce sous-type grammatical :

anglais : *an independent power project*

BankGloss : un projet de production d'énergie indépendante

Promt : un projet de pouvoir indépendant

Babylon : un pouvoir indépendant projet

Google : un projet énergétique indépendant

Bing : un projet d'électricité indépendants

Pour tous les traducteurs, il n'y a pas eu d'ajout nécessaire du nom « production », car il ne figure pas dans le texte anglais. Dans cet exemple, il faut que les machines reconnaissent le terme en anglais, *power project*, sinon la traduction donnée va être littérale. Une autre erreur qui est commise par *Babylon*, *Google* et *Bing* est l'omission de la préposition « de ».

Dans le même extrait, tous les systèmes commettent certaines fautes de conjugaison :

anglais : *The governor set his sights high: privatize the state power utility, structure and bid out an independent power project, and set up a regulatory agency*

BankGloss : Le gouverneur s'est fixé d'ambitieux objectifs : il a en effet décidé de privatiser la compagnie d'électricité de l'État, de structurer un projet de production d'énergie indépendante, de lancer un appel d'offres en vue de sa réalisation, et de mettre en place un organisme réglementaire.

Prompt : Le gouverneur avait des visées haut : privatisez l'utilité de pouvoir publique, la structure et offrez un projet de pouvoir indépendant et remettez une agence de contrôle sur pied.

Babylon : Le gouverneur définir ses sites haute: privatiser le pouvoir de l'état utilitaire, la structure et les soumissions à un pouvoir indépendant projet, et de mettre en place un organisme de réglementation.

Google : Le gouverneur jeté son dévolu élevé: la privatisation de la puissance publique de l'état, de la structure et enchérir sur un projet énergétique indépendant, et mettre en place un organisme de réglementation.

Bing : Le gouverneur a jeté son dévolu haute : privatiser l'électricité d'État, la structure et la soumission d'un

projet d'électricité indépendants et de mettre en place un organisme de réglementation.

Dans cet exemple, la plupart des verbes ne sont pas conjugués correctement. Cette erreur est probablement liée à l'incapacité d'identifier le sujet du verbe, comme « le gouverneur ». Il y a aussi l'utilisation de mots relevant d'une de partie du discours erronée, par exemple lorsque le verbe est traduit comme un substantif.

Une erreur d'omission de mots qui est répétée par trois des systèmes tout au long du corpus a trait à la traduction d'*IFC* (voir les extraits 2, 3, 6, 7, 8 et 10 dans l'appendice). *Prompt* oublie toujours l'article défini devant cet acronyme, tandis que *Google Translate* et *Bing* l'omettent parfois. *Babylon* ne commet jamais cette erreur.

4.1.2 Erreurs lexicales

Nous avons dénombré 17 erreurs lexicales au total, dont 15 anglicismes et 2 choix de termes inappropriés.

L'extrait 2 nous donne un exemple où tous les systèmes ont commis la même erreur lexicale :

anglais : *developing innovative products*

BankGloss : créer des produits novateurs

Prompt : développe des produits innovateurs

Babylon : développement de produits innovateurs

Google : développer des produits innovants

Bing : développer des produits innovants

Le mot « développer » n'est pas l'équivalent de « créer » en français. D'après le dictionnaire, « développer » implique la « croissance organique, d'expansion » (TLFi) La définition de « créer » est « donner l'existence à » (TLFi). Sur la base de la définition de ce mot, nous pouvons déduire qu'il y a une erreur sous forme d'anglicisme dans cet extrait.

Un autre exemple où tous les systèmes ont commis la même erreur lexicale se trouve dans l'extrait 10 :

anglais : *the utility*

BankGloss : la compagnie d'électricité

Prompt : l'utilité

Babylon : l'utilitaire

Google : l'utilité

Bing : l'utilitaire

« L'utilité » et « l'utilitaire » ne sont pas équivalents à *utility* dans ce contexte.

Selon le dictionnaire, la définition de « utilité » est « caractère de ce qui est utile » (TLFi). Cet exemple démontre un autre cas d'anglicisme.

L'extrait 1 contient des erreurs lexicales commises par *Prompt* et *Babylon*, mais pas par *Google* et *Bing* :

anglais : *capital*

BankGloss : ressources

Prompt :	capitale
Babylon :	capitale
Google :	aide
Bing :	capital

Ici, les traductions *Prompt* et *Babylon* démontrent un choix de terme inapproprié. « Ressources » signifie « Moyens pécuniaires dont dispose une personne pour assurer son existence » et « Capitale » signifie « Ville principale d'un État, d'une province ou d'une étendue de pays qui est le siège du gouvernement et/ou de l'administration centrale » (*TLFi*). Pour traduire ce mot correctement, il faut un autre mot, comme « capital », qui peut renvoyer aux ressources.

4.1.3 Erreurs sémantiques

Les systèmes ont commis 74 erreurs sémantiques au total. Au niveau de ce type de fautes, nous avons dénombré 62 faux-sens, 7 ajouts non-justifiables, 2 non-sens, 2 imprécisions et 1 ambiguïté. Le sous-type le plus répandu a donc été le faux-sens. D'après Delisle, « le faux sens est une erreur moins grave qu'un contresens ou un non-sens, car il ne dénature pas complètement le sens du texte de départ » (Dussart, 2005 : 108-109).

En regardant les erreurs produites globalement qui figurent en annexe, on peut remarquer aussi que parmi tous les sous-types d'erreurs, les faux-sens ont été les plus répandus dans l'ensemble des systèmes de traduction. Cela indique

que les traductions produites par les quatre systèmes examinés ne sont pas très fiables en matière de traduction de documents techniques.

Les extraits 4, 8 et 9 montrent des cas où tous les systèmes ont commis les mêmes erreurs sémantiques :

Extrait 4 :

anglais : *these frontier countries.*

BankGloss : ces pays.

Prompt : ces pays de frontière.

Babylon : ces pays frontaliers.

Google : ces pays frontaliers.

Bing : ces pays de frontière.

Le sens de « frontière » et « frontalier » n'est pas équivalent à *frontier* dans ce contexte. Ceci produit un faux-sens. La bonne traduction est « pays pionnier »

Extrait 8 :

anglais : *a venture*

BankGloss : une opération

Prompt : une entreprise

Babylon : une coentreprise

Google : une entreprise

Bing : une entreprise

Ici, l'erreur sémantique est aussi un faux-sens parce que le mot « entreprise » altère le sens du texte, car sa définition est « mise en œuvre de capitaux et d'une main-d'œuvre salariée en vue d'une production ou de services déterminés » (TLFi). La bonne traduction dans ce contexte devrait donc être « opération » ou alors l'équivalent plus général « coentreprise ».

Extrait 9 :

anglais : *We can move very quickly in response to a country's evolution.*

BankGloss : Nous pouvons intervenir très rapidement quand la situation évolue.

Prompt : Nous pouvons bouger très vite en réponse à l'évolution d'un pays.

Babylon : Nous pouvons déplacer très rapidement en réponse à une évolution du pays.

Google : Nous pouvons nous déplacer très rapidement en réponse à l'évolution d'un pays.

Bing : Nous pouvons déplacer très rapidement en réponse à l'évolution du pays.

La traduction de *move* est incorrecte pour tous les systèmes. « Bouger » et « se déplacer » sont des traductions littérales. Il faut connaître le sens du texte pour pouvoir traduire correctement le verbe, car il ne s'agit pas d'un déplacement physique.

Le même extrait contient une autre erreur sémantique commise par tous les systèmes:

anglais : *hospitable policies*
BankGloss : un cadre réglementaire favorable
Promt : des politiques hospitalières
Babylon : politiques hospitalière
Google : des politiques hospitalières
Bing : des politiques hospitalier

Comme dans l'exemple précédent, les systèmes effectuent une traduction littérale, ce qui entraîne un faux-sens. *Hospitable* a ici le sens de « favorable » ou « propice »

4.1.4 Erreurs stylistiques

Nous avons dénombré 6 erreurs stylistiques au total, dont 4 constructions boiteuses et 2 cas d'absence d'idiomaticité. Les erreurs stylistiques commises se trouvent dans l'extrait 9 :

anglais : *wants its private sector to develop, wants its financial sector to develop*
BankGloss : veut développer son secteur privé, ou son secteur financier
Promt : veut que son secteur privé se développe, veut que son secteur financier se développe

Babylon : veut son secteur privé à développer, veut que son secteur financier de développer

Google : veut que son secteur privé à développer, veut que son secteur financier pour développer

Bing : veut que son secteur financier afin de développer son secteur privé à développer, veut

Ici, tous les traducteurs commettent une erreur de construction boiteuse sous forme de répétition inutile de mots. Dans ce cas précis, l'anglais permet plus de répétition que le français.

Dans le même extrait, *Prompt* affiche deux erreurs d'absence d'idiomaticité :

anglais : *However, it takes two to tango. It is up to the government to set in place hospitable policies.*

BankGloss : Mais, comme sur la piste de danse, il faut être deux pour avancer. L'État doit mettre en place un cadre réglementaire favorable.

Prompt : Cependant, il prend deux au tango. Il est à la hauteur du gouvernement pour mettre dans l'endroit des politiques hospitalières.

Babylon : Toutefois, il faut être deux pour danser le tango. C'est au gouvernement de mettre en place politiques hospitalière.

Google : Cependant, il faut être deux pour danser le tango. Il appartient au gouvernement de mettre en place des politiques hospitalières.

Bing : Toutefois, il faut être deux pour danser le tango. Il appartient au gouvernement à mettre en place des politiques hospitalier.

Ici, *Prompt* traduit incorrectement deux expressions d'une manière littérale. C'est le seul système qui a commis ces erreurs. Cela montre que *Prompt* manque de capacité de reconnaissance des expressions idiomatiques.

4.2 Récapitulation des problèmes de traduction automatique rencontrés lors de l'examen de 4 systèmes

L'analyse des erreurs nous apprend que les systèmes de traduction automatique commettent le plus souvent des erreurs grammaticales, suivies des erreurs sémantiques, lexicales et stylistiques. Nous examinerons les lacunes des logiciels, afin de pouvoir formuler des recommandations d'amélioration. Nous allons comparer les problèmes liés aux systèmes basés sur des dictionnaires avec ceux des systèmes reposant sur les statistiques.

4.2.1 Problèmes liés aux systèmes basés sur les règles

Globalement, les traducteurs *Prompt* et *Babylon* ont commis plus d'erreurs que *Google* et *Bing*. Les deux premiers sont basés sur des dictionnaires, les deux

derniers sur des équivalents choisis statistiquement à partir d'un corpus de documents déjà traduits. Cela tend à indiquer que les systèmes qui utilisent la statistique sont peut-être plus efficaces. En analysant les documents qui ont été déjà traduits, on constate qu'ils ont une capacité plus élevée de reconnaissance de ces types de structures par rapport aux traducteurs basés sur un dictionnaire. Cependant, tel n'est pas le cas dans tous les types d'erreurs; avec le type grammatical, la somme des erreurs de *Google* et *Bing* est supérieure à celle de *Prompt* et *Babylon*. Ces derniers commettent des fautes récurrentes dans la traduction des formes verbales qui contiennent un auxiliaire. Par exemple, *we are changing* est traduit par « nous sommes modifie » par *Google* au lieu de « nous modifions » ; *Bing* traduit la même expression par « nous sommes remodelage » au lieu de « nous changeons ». Il existe aussi plusieurs autres exemples semblables dans les diverses traductions. Avec tous les autres types, *Google* et *Bing* ont commis moins d'erreurs que *Prompt* et *Babylon*.

En ce qui concerne les autres types d'erreurs (lexicale, sémantique et stylistique), le problème avec les machines de traduction automatique basées sur les dictionnaires est qu'il n'y a pas assez de données. La qualité de la traduction dépend de la qualité de l'information que les programmeurs fournissent à la machine. L'ordinateur peut seulement se référer à un vocabulaire et un ensemble de règles. Or, les langues sont complexes et il y a toujours des exceptions aux règles. Lorsque l'ordinateur essaye d'identifier ces exceptions, la qualité de la traduction commence à se dégrader. Les dictionnaires ne peuvent fournir la matière première qui doit en quelque sorte être convertie en texte

traduit, ou au moins en la meilleure approximation possible d'une traduction (Brown, 1964 : 3). Dans la traduction automatique statistique, la connaissance de la traduction est apprise automatiquement à partir de données d'exemple (Och, 2002 : 9). Ces systèmes doivent simplement disposer d'un réseau de documents traduits qui est assez vaste, pour produire des traductions correctes. Ce type de système est autosuffisant du point de vue de sa capacité d'identifier les modèles et produire des traductions.

En particulier, les problèmes associés uniquement à *Prompt* ont trait aux choix incorrects de termes équivalents. Le système n'est donc pas capable de choisir le bon mot selon le contexte de la phrase. En outre, parfois le contexte n'est pas significatif et le logiciel devrait reconnaître des expressions ou un groupe de termes fixes. *Prompt* ne parvient pas à accomplir cette tâche (ex. *it takes two to tango* ≠ il prend deux au tango). Cela nous indique qu'on n'a pas programmé dans le système assez d'expressions figées et leurs équivalents.

Les erreurs commises exclusivement par *Babylon* portent à croire que ce système ne contient pas assez de règles sur la conjugaison des verbes et sur les temps appropriés à utiliser. Les choix incorrects de termes équivalents montrent que le système est incapable d'identifier le contexte des mots. Par ailleurs, le logiciel donne souvent un ordre incorrect des catégories grammaticales (ex. *independent power project* ≠ un pouvoir indépendant projet). La traduction correcte fournie par *BankGloss* est la suivante :

anglais : ***Independent*** (adj.) ***power*** (nom) ***project*** (nom)

français : **un projet de production** (nom) **d'énergie** (nom)

indépendante (adj.)

Afin de traduire cette expression correctement en français, il faut avoir l'ordre suivant : nom – nom descriptif – adjectif. Cependant, la formule adoptée par *Babylon* est : nom descriptif – adjectif – nom. Cette erreur tient peut-être à une incapacité de distinguer entre le nom et l'adjectif en anglais.

4.2.2 Problèmes liés aux systèmes de traduction statistique

Les systèmes de traduction statistique *Google Translate* et *Bing* accusent eux aussi des lacunes, entre autres parce que les usagers peuvent intervenir dans le processus traductionnel. Le problème avec cette méthode est que tout usager a la possibilité d'ajouter des données au corpus de référence, sans aucun contrôle préalable de qualité. Pour cette raison, la qualité des traductions risque de se dégrader si les données ajoutées sont de piètre qualité. Si par exemple, une mauvaise traduction est sélectionnée plusieurs fois ou si elle est jugée bonne par l'utilisateur, le système de traduction automatique se trompera en proposant une traduction incorrecte la prochaine fois que la même donnée est saisie. Par conséquent, les usagers doivent faire preuve de prudence dans le choix de la bonne traduction, afin de maintenir la qualité du système de traduction. De plus, la langue utilisée dans les données n'est pas restreinte, car les usagers peuvent entrer un texte avec des expressions régionales ou vulgaires. Il s'ensuit que l'utilisateur peut dégrader le système en choisissant des traductions qui ne correspondent pas au sens du texte original, ce qui à son tour rend les futures

traductions incorrectes. Cela diminuerait la qualité de la traduction pour le prochain usager.

Plus spécifiquement, le système de *Google* manque des règles liées à la conjugaison des verbes, comme avec le temps présent progressif (*we are eliminating* ≠ nous sommes modifie). Le logiciel manque également de capacité de reconnaissance des acronymes comme noms et ne place donc pas d'article devant l'acronyme. En outre, *Google* commet des erreurs lexicales en identifiant les termes à traduire hors du contexte de la phrase.

Le système *Bing* n'est pas programmé pour savoir quand il est nécessaire de placer un article après la préposition « de » lorsque le mot qui suit est un nom commençant par une voyelle (ex. « d'État » est donné comme sortie au lieu de « de l'État »). Comme *Google*, *Bing* fait des choix lexicaux incorrects, ce qui implique que les termes sont identifiés hors contexte. Par ailleurs, le logiciel confond les parties du discours (verbe et substantif) et donne un ordre incorrect de ces parties (ex. « investissement ministères » est donné comme sortie au lieu de « ministères d'investissement »). Les règles relatives au placement correct des parties du discours sont donc insuffisantes.

La typologie des erreurs nous aide à identifier les lacunes des traductions automatiques et les causes de leurs erreurs. On a vu que les machines ayant fait l'objet d'évaluation dans le présent travail ne contiennent pas assez d'informations ni de règles programmées dans leurs systèmes aux niveaux grammatical, lexical, sémantique et stylistique. Nous avons examiné aussi les

différences entre les systèmes de traduction basés sur des dictionnaires et ceux qui sont basés sur des statistiques et les problèmes spécifiques rencontrés dans chacun. Dans le chapitre suivant, qui tient également lieu de conclusion, nous formulerons des suggestions pour améliorer les systèmes de traduction et examinerons les perspectives d'avenir de la traduction automatique.

CHAPITRE 5

Conclusion et perspectives de la traduction automatique

Dans ce chapitre, nous nous attacherons à situer les résultats de notre étude par rapport à ceux d'autres travaux de recherche existants. Nous émettrons ensuite des suggestions à l'intention de l'utilisateur. Enfin, nous examinerons les perspectives d'avenir de la traduction automatique.

On note plus de similarités que de différences entre nos résultats et ceux des chercheurs évoqués au début de notre thèse. On peut en déduire que par rapport aux années 90, époque où Balkin a fait le constat que les traductions de *Babelfish* étaient incompréhensibles et pleines d'erreurs, la traduction automatique gratuite s'est beaucoup améliorée. Cependant, dans une certaine mesure, la conclusion de Balkin reste encore valable aujourd'hui : les logiciels de traduction automatique sont de bons outils pour avoir une idée générale d'un texte, mais si on ne maîtrise pas la langue cible ni le sujet traité, on est confronté à des problèmes de compréhension (Balkin, 1999 : 57).

En conformité avec les résultats d'Olvera-Lobo et García-Santiago, nous avons trouvé que les meilleures traductions étaient fournies par *Google Translate* et que le logiciel le plus faible était *Prompt* (Olvera-Lobo et al., 2010 : 434). La conclusion selon laquelle *Google* se classe au premier rang est corroborée aussi par Savoy et Dolamic et Hampshire et al. (Hampshire et al., 2010 : 201). Hampshire et al. ont trouvé que *Babylon* a enregistré une performance supérieure à celle de *Google* seulement sur le plan des expressions

idiomatiques (Hampshire et al., 2010 : 207). Dans notre étude, bien que *Babylon* ait commis moins d'erreurs sémantiques et stylistiques par rapport à la plupart des autres logiciels, il n'a pas été plus performant que Google, qui a commis le même nombre d'erreurs au niveau de ces deux types de fautes. Ainsi, contrairement aux conclusions de Hampshire et son équipe ont dit, nous considérons *Google* comme étant un meilleur système de traduction que *Babylon* à tous égards. Depuis 2010, il est possible que ce logiciel se soit amélioré d'une manière significative. Aiken et son équipe ont également indiqué que ce système fournissait des traductions très précises (Aiken et al., 2012 : 26). Les traductions n'étaient certes pas parfaites, mais c'est bien le sens global du texte qui importait (Aiken et al., 2012 : 29). Par conséquent, comme tous les chercheurs qui ont procédé à une évaluation de *Google*, nous pouvons affirmer que ce système fournit les meilleures traductions.

Pour ce qui est d'Olvera-Lobo et son équipe, la plupart des erreurs commises par les logiciels dans leur étude étaient de nature lexicale (Olvera-Lobo et al., 2010 : 451). De même, nous avons remarqué que la plupart des erreurs commises par les logiciels dans notre évaluation étaient des faux-sens, qui font partie du type sémantique (Olvera-Lobo a précisé le type lexical, mais pas le type sémantique). En outre, Savoy et Dolamic ont noté des problèmes liés principalement à l'ambiguïté lexicale (Hampshire et al., 2010 : 201). Par conséquent, on peut déduire que les erreurs commises par les systèmes de traduction automatique sont toujours attribuables à une mauvaise interprétation du sens.

5.1 Suggestions d'amélioration en vue d'un modèle supérieur de traduction automatique

Sur la base des recherches effectuées par d'autres auteurs et de la typologie des erreurs commises par *Prompt*, *Babylon*, *Google* et *Bing*, on va aussi formuler des recommandations spécifiques pour améliorer ces systèmes de traduction. On a vu que *Prompt* et *Babylon* s'appuient sur des dictionnaires et que *Google* et *Bing* sont basés sur la statistique. Il existe des différences dans les solutions de traduction proposées par les deux types de systèmes. Il existe également des possibilités, pour l'utilisateur, d'adopter certaines mesures pour réduire les erreurs tout au long du processus de traduction.

5.1.1 Recommandations pour le système de traduction basé sur le dictionnaire

L'expérience de la traduction automatique montre que les échecs de la traduction sont souvent causés par des problèmes inhérents aux données contenues dans le dictionnaire. D'après Nagao, il est donc absolument essentiel de construire le système de traduction automatique de telle sorte que la révision immédiate du dictionnaire soit possible. En même temps, il y a lieu de faire preuve d'une grande prudence quand des modifications sont apportées aux données du dictionnaire. Lorsque ces modifications sont opérées en vue de s'assurer qu'une certaine partie d'un texte est traduite correctement, on peut constater que les expressions qui avaient déjà été traduites ne sont plus adéquatement manipulables. À mesure que les systèmes de traduction

automatique augmentent en taille et en complexité, il devient de plus en plus difficile de déterminer quelles seront les influences qui en résultent sur l'ensemble du système, à cause des changements opérés dans une partie du dictionnaire. En conséquence de toutes ces influences, le temps nécessaire pour apporter des améliorations minimales s'étend considérablement. Il y a même des cas où la performance de l'ensemble du système peut se détériorer, aussi les altérations du système ne devraient-elles pas être faites à la légère (Nagao, 1989 : 129).

L'Homme traite du besoin d'enrichissement et d'édition du dictionnaire. L'édition permet d'identifier les termes qui se trouvent dans les textes habituels de celui qui traduit, ainsi que d'extraire les équivalents appropriés. La première proposition de L'Homme est l'ajout d'entrées, cas où l'utilisateur peut saisir de nouveaux mots lui-même, y compris les termes complexes comme les expressions figées. La deuxième est la suppression d'un mot, le système étant conçu pour permettre à l'utilisateur d'éliminer des mots répertoriés qui sont en conflit avec d'autres termes du texte de sortie. La dernière proposition est l'édition d'entrées existantes, qui comprend la suppression d'une catégorie grammaticale pour un homographe catégoriel afin de réduire les mauvaises interprétations. Par exemple, si on veut traduire le mot *bit* dans un contexte informatique, le logiciel doit supprimer l'entrée verbale en tenant compte que seulement la forme nominale peut être utilisée dans les textes qu'il traduit (L'Homme, 2000 : 326).

À la lumière de nos recherches, nous formulons quelques suggestions

précises pour améliorer les systèmes de traduction analysés dans notre étude. Pour améliorer *Prompt*, on peut ajouter plus de termes dans son dictionnaire, afin d'élargir son réseau de vocabulaire et réduire les erreurs sémantiques et idiomatiques. On peut donc ajouter plus d'expressions fixes en vue d'éviter des traductions littérales qui ne correspondent pas sémantiquement au texte du départ. Il serait également bon, pour le logiciel, de tenir compte de l'ensemble du texte au lieu des mots individuels uniquement. Par exemple, on a vu que *Prompt* était le seul logiciel qui a traduit le mot « move » incorrectement par « bouger ». Si le logiciel pouvait analyser l'ensemble du texte, cette erreur n'aurait pas été commise. Il est donc important qu'il puisse reconnaître dans une certaine mesure le contexte du mot donné lors de sa traduction. On a remarqué que le logiciel omet souvent les déterminants, comme dans l'extrait 1 où il est le seul qui fournit « de secteur privé » comme traduction. *Prompt* omet aussi les déterminants avant les acronymes, parce que le système ne les reconnaît pas comme étant des noms. Par conséquent, il faut apprendre au système à identifier les noms communs et à insérer un déterminant avant chacun d'eux.

Pour améliorer la traduction de *Babylon*, il faut ajouter plus de règles de grammaire ; il serait bon pour la machine de reconnaître les cas où il est nécessaire d'insérer des déterminants, des noms, des verbes et des prépositions, afin d'éviter les omissions de catégories grammaticales. En plus des règles d'accord, il faut aussi incorporer plus de règles touchant à la structure dans le système, pour éviter les fautes de ponctuation et l'utilisation de prépositions et d'articles erronés. Il faut apprendre au logiciel que généralement,

l'adjectif suit le nom en français, comme dans l'extrait 1 avec la traduction d'« economic growth » par « la croissance économique » au lieu d'« économique la croissance ». Il convient par ailleurs de lui apprendre les exceptions à ces règles qui décrivent la position correcte des catégories grammaticales, comme dans la traduction de l'expression « the small mouse » qui est rendue par « la petite souris » en français. Tel que déjà suggéré pour *Prompt*, pour réduire le nombre de faux sens, la machine devrait tenir compte du texte dans son ensemble et non seulement des mots individuels.

5.1.2 Recommandations concernant le système de traduction statistique

De même que des améliorations à l'évaluation de la traduction automatique, il est généralement admis que de bonnes connaissances linguistiques peuvent jouer un rôle dans l'amélioration des systèmes statistiques d'aujourd'hui, dans toutes les phases du processus traductionnel (Way, 2010 : 568). Selon Stephan Vogel (Geer, 2005 : 20), chercheur associé à la *Language Technologies Institute* de *CMU* (Université Carnegie Mellon), par rapport aux systèmes à base de règles, lorsqu'on investit le même montant d'argent dans la production des traductions, le système statistique donne de meilleures traductions. Alors que la création de systèmes complexes à base de règles pour les paires de langues nécessite beaucoup d'années et des dépenses considérables, les chercheurs peuvent former de systèmes statistiques utilisables pour produire des traductions en seulement quelques semaines ou même des jours. Cela rend l'approche statistique moins laborieuse et plus utile pour les applications des entreprises et

des gouvernements soucieux du temps. Toutefois, l'approche nécessite également des ordinateurs puissants pour gérer la formation et les nouvelles traductions. Puisque les techniques statistiques n'ont pas besoin de la participation d'experts linguistiques, ils peuvent être efficaces dans la traduction vers et à partir des langues obscures, tant que de nombreux documents et des traductions sont disponibles pour la formation (Geer, 2005 : 20).

Nos suggestions d'amélioration pour *Bing* et *Google* sont semblables, sauf en ce qui concerne les recommandations relatives à l'interaction avec l'utilisateur. Pour améliorer ces logiciels, il faut principalement ajouter plus de règles grammaticales spécifiques. Par exemple, les machines doivent savoir insérer un article avant chaque nom commun, pour éviter les omissions de catégorie grammaticale. Afin de réduire le nombre d'erreurs sémantiques, particulièrement les faux-sens, les logiciels devraient tenir compte du texte dans son ensemble et non seulement des mots individuels. Les systèmes de traduction devraient aussi avoir plus de cohérence dans leurs sorties, pour éviter les correspondances erronées. Pour ce faire, ils devraient avoir la capacité de se souvenir des traductions antérieures.

Il convient en outre de formuler plusieurs autres recommandations particulières pour *Google*. Le logiciel de *Google* permet l'interaction avec l'utilisateur qui peut sélectionner la meilleure traduction d'un mot donné. Le choix de l'utilisateur, quelle que soit sa connaissance de la langue, influe sur les statistiques du système. Parfois, on a l'option de sélectionner un mot qui conduit à des erreurs dans la sortie de traduction. Afin d'éviter ce genre de situation, il

serait bon d'intégrer une fonction de vérification orthographique et sémantique dans le logiciel, de sorte qu'il soit impossible de choisir un mot qui va produire des erreurs de sens et de structure dans la phrase. Ce type de logiciel produit facilement des ambiguïtés catégorielles, comme on l'a vu avec les cas de verbes conjugués à l'impératif. Par exemple, quand la phrase anglaise « Close and secure panel door » est entrée dans le logiciel de *Google*, elle est traduite par « Fermez et une porte d'armoire sécurisée » en français et il n'est pas possible de changer la catégorie grammaticale de *sécurisée* d'adjectif à verbe. Dans ce cas, l'idéal serait que le système permette à l'utilisateur de manipuler la phrase un peu plus. En ce qui concerne les choix lexicaux, le logiciel devrait être capable d'énumérer tous les équivalents possibles du terme du texte de départ. Sur la base du choix effectué par l'utilisateur, le logiciel devrait automatiquement manipuler la suite du texte pour le faire correspondre au contexte donné.

L'Homme appelle ce processus la désambiguïstation des homographes sémantiques (L'Homme, 2000 : 328). Normalement, avec les autres logiciels de traduction, cette désambiguïstation est automatique, mais souvent incorrecte. Il serait également utile que le logiciel fonde une plus grande proportion de ses statistiques sur les documents traduits professionnellement plutôt que sur les choix opérés par les utilisateurs.

5.2 Suggestions à l'intention de l'utilisateur

Pour parvenir à une traduction de haute qualité, les chercheurs ont adressé quelques conseils utiles aux utilisateurs, tels que la nécessité du contrôle de la

langue et de l'édition. On peut rédiger le document de manière très simple pour faciliter l'opération de traduction par les systèmes automatiques (Koehn, 2010 : 21). Cette suggestion concorde avec celle de Hutchins, pour qui la qualité de la traduction peut être améliorée en imposant des limites par rapport aux données d'entrée ; le système peut être programmé pour traiter des textes limités à un sous-langage particulier ou bien les données d'entrées peuvent être écrites dans un langage contrôlé afin de réduire les ambiguïtés (Hutchins, 1992 : 4). Le logiciel de *Google* aide l'utilisateur à faire de la pré-édition, car lorsqu'un mot est mal écrit dans le texte de départ, le système demande si l'utilisateur veut dire quelque chose d'autre. Une suggestion est donnée et l'utilisateur a le choix de l'accepter ou de la refuser. Si tous les systèmes adoptaient cette méthode, les traductions seraient généralement améliorées.

Une deuxième suggestion d'intervention de l'utilisateur proposée par les chercheurs est le processus d'édition. En 2000, L'Homme a défini ce processus comme suit : la pré-édition est l'aménagement du texte de départ avant de le soumettre au logiciel. La post-édition est la correction du texte d'arrivée, ainsi que l'intervention au cours du processus de la traduction (L'Homme, 2000 : 322-323). Une pré-édition lourde ne peut jamais éliminer entièrement la post-édition. La pré-édition simple consiste à corriger les erreurs du texte de départ, comme les fautes d'orthographe ou grammaticales. La pré-édition lourde couvre l'uniformisation terminologique, la réduction des ambiguïtés sémantiques, la réduction des phrases longues et complexes, l'indication explicite des propositions subordonnées ou incises et le remplacement des pronoms par les

noms (L'Homme, 2000 : 323). L'interactivité en cours de traduction permet à l'utilisateur de sélectionner la meilleure traduction d'un mot ou d'une expression donnée (L'Homme, 2000 : 325).

D'après ElShiekh, le degré de précision de la traduction automatique est beaucoup plus faible que celle d'une traduction humaine compétente, mais elle peut être améliorée avec le processus d'édition. Lorsqu'elle est utilisée en conjonction avec des traducteurs humains, l'objectif principal de la traduction automatique est de fournir un premier projet, qui est ensuite donnée à un traducteur humain pour l'édition et le polissage. Dans ce dernier cas, la traduction automatique permet d'économiser beaucoup de temps, d'efforts et de ressources financières (ElShiekh, 2012 : 58). Nous considérons qu'un système de traduction est de grande qualité si sa production ne nécessite presque aucune post-édition. En fait, il s'agit d'une situation relativement rare dans le domaine de la traduction automatique, où la plupart des systèmes offrent des traductions qui doivent subir la post-édition (Farrús et al., 2011 : 182).

5.3 Perspectives d'avenir de la traduction automatique

Les erreurs sémantiques et lexicales étaient les plus nombreuses et ont entraîné des difficultés liées à la désambiguïsation. Bien qu'il existe encore beaucoup de lacunes dans la traduction automatique gratuite, son utilisation est encore essentielle aujourd'hui. Nous avons présumé que chacun des systèmes aura ses propres forces et faiblesses. On a trouvé que même si *Prompt* a commis le plus grand nombre d'erreurs en général, ce système a fourni le plus petit nombre

d'erreurs grammaticales. Cela indique que si on veut qu'une traduction soit correcte grammaticalement, il serait mieux d'utiliser *Prompt* par rapport aux autres logiciels dans notre étude. De la même manière, si on veut une traduction sans erreurs sémantiques et lexicales, il serait mieux d'utiliser le système de *Google*. Par conséquent, l'utilisateur peut choisir le système qu'il juge le mieux adapté à ses besoins.

La question qui reste posée est celle de savoir si oui ou non la traduction automatique va continuer à être un outil essentiel à l'avenir et ce que cela signifie pour les traducteurs humains. D'après Kuras, le processus de traduction effectué par un humain est comparable à celui suivi par un outil de mémoire de traduction, exception faite de l'apport créatif d'un homme (Kuras, 2011 : 239). Dans la traduction, l'accès au système est essentiel, ce qui implique que la traduction automatique est principalement commerciale. En ce qui concerne le contenu, les êtres humains n'éprouvent normalement pas de difficulté à interpréter un terme sémantiquement complexe, alors que dans la traduction automatique, le texte d'entrée doit être restreint. Les êtres humains peuvent généralement traduire des textes de tout domaine de connaissances, alors que la plupart des systèmes de traduction automatique sont efficaces dans des domaines précis. Cela ne signifie pas nécessairement que la traduction humaine est meilleure que celle d'une machine. Par exemple, si un traducteur humain ayant peu d'expérience de la traduction médicale est appelé à traduire un document médical sans utiliser de base de données terminologiques il est possible que sa production soit de moindre qualité que celle d'un système de

traduction automatique basé sur des corpus et des références terminologiques (Anastasiou, 2011 : 648).

Un rapport récent de Dillinger et Laurie (Anastasiou, 2011 : 648) montre que le niveau d'acceptation des articles traduits au moyen de systèmes de traduction automatique, tel mesuré par la rétroaction des utilisateurs, est pratiquement similaire à celui des articles qui sont traduits professionnellement par des êtres humains. En ce qui concerne les délais, les êtres humains doivent soumettre leurs traductions dans un délai relativement long (car s'étendant dans le temps), tandis que la production de la traduction automatique est instantanée. Enfin, la traduction humaine est plus coûteuse que la traduction automatique. Étant donné que la traduction automatique soutient plus d'une paire de langues, l'investissement initial dans un système commercial peut permettre de réaliser des économies à long terme (Anastasiou, 2011 : 648). L'utilisation d'un système de traduction automatique est très profitable lorsqu'un traducteur doit souvent traduire une grande quantité de textes (Kuras, 2011 : 239). Sur la base de cette recherche, on peut affirmer que la traduction automatique est un outil indispensable qui profite de façon considérable au domaine de la traduction.

On peut conclure que la traduction automatique ne vise pas la production d'un texte acceptable en tout point de vue comme dans le cas de la traduction humaine. Il s'agit plutôt d'une bonne aide à la traduction, compte tenu du fait que les sorties de la traduction automatique peuvent permettre d'accomplir certaines tâches (Koehn, 2010 : 217). Les résultats de notre étude confirment cette réalité. Les quatre systèmes de traduction automatique, à savoir *Prompt*, *Babylon*,

Google et *Bing*, commettent encore trop d'erreurs pour que l'on puisse compter sur eux pour produire une bonne traduction de haute qualité. Bien que les systèmes de traduction automatique en ligne ne soient pas toujours le meilleur choix pour une traduction très précise, à grande échelle et adapté au domaine, ils semblent répondre adéquatement aux besoins du grand public, comme en témoignent les millions de visites quotidiennes que ces sites reçoivent, lorsque la nécessité de savoir l'essentiel d'un texte est supérieure au besoin d'avoir une traduction parfaite (Way, 2010 : 558). Comprendre les besoins de traduction des utilisateurs est essentiel à l'amélioration continue des services de la traduction automatique (Lopez, 2008 : 41). Avec la poursuite des recherches, on peut prédire que la traduction automatique continuera à être s'imposer comme une nécessité pour la communication interlinguale dans notre village planétaire.

RÉFÉRENCES BIBLIOGRAPHIQUES

- Aiken, Milam, Mina Park, and Tobin Lindblom. "Integrating Machine Translation With Group Support Systems." *International Journal Of Business & Management* 5.5 (2010): 25-35. *Business Source Complete*. Web. 28 Nov. 2012.
- Anastasiou, Dimitra and Rajat Gupta. "Comparison Of Crowdsourcing Translation With Machine Translation." *Journal of Information Science* 37.6 (2011): 637-659. *Academic Search Premier*. Web. 28 May 2013.
- Babylon. "Babylon Online Translation." *translation.babylon.com*, 2011. Web. 2012.
- Babylon. "Babylon Translator." *www.babylon.com*, 2013. Web. 8 Apr. 2013.
- Balkin, Ruth. "AltaVista's Automatic Translation Program." *Database* 22.2 (1999): 56-7. *ABI/INFORM Global; ProQuest Business Collection; ProQuest Computing; ProQuest Education Journals; ProQuest Nursing & Allied Health Source; ProQuest Research Library; ProQuest Telecommunications*. Web. 6 Dec. 2012.
- Brown, A.F.R. "Automatic Translation of Languages." *Information Storage and Retrieval* 2.1 (1964): 1-28. Web. 6 Dec. 2012.
- "Babylon Presents Babylon 2008 MAC, the First Specially Designed Online Version of its Globally-Known Dictionary and Translation Software for MAC." *Canada NewsWire* Sep 16 2008. *ProQuest*. Web. 8 Apr. 2013.
- Cristinoi-Bursuc, Antonia. « Les erreurs dans la traduction automatique du genre dans les couples français-anglais et anglais-français : typologie, causes linguistiques et solutions ». *Revue française de linguistique appliquée* 14.1 (2009) : 93-108.
- Delavenay, Emile. *An Introduction to Machine Translation*. New York: Praeger, 1960. Print.
- Dussart, André. "Faux sens, contresens, non-sens... un faux débat?." *Meta: Journal des traducteurs* 50.1 (2005): 107-119.
- ElShiekh, Ahmed Abdel Azim. "Google Translate Service: Transfer of Meaning, Distortion or Simply a New Creation? An Investigation into the Translation Process & Problems at Google." *English Language and Literature Studies* 2.1 (2012): 56-68. Web. 12 Dec. 2012.

- Farrús, Mireia, et al. "Overcoming statistical machine translation limitations: error analysis and proposed solutions for the Catalan–Spanish language pair." *Language resources and evaluation* 45.2 (2011): 181-208. Web. 3 Jun. 2013.
- Geer, David. "Statistical Machine Translation Gains Respect." *Computer* 38.10 (2005): 18-21. Web. 4 Jun. 2013.
- Giménez, Jesús, and Lluís Màrquez. "Linguistic measures for automatic machine translation evaluation." *Machine Translation* 24.3-4 (2010): 209-240. Web. 28 Nov. 2012.
- Goldsborough, Reid. "Lost in translation: Internet Multilingualism Can Mean Confusion." *Community College Week* 21.15 (2009): 13. *Academic Search Premier*. Web. 30 Mar. 2013.
- Google. "Google Translate." *translate.google.ca*, 2013. Web. 2012.
- Google. "Find out how translations are created." *translate.google.com/about/intl/en_ALL*, 2013. Web. 3 Dec. 2012.
- Hardmeier, Christian. "Discourse in Statistical Machine Translation." *Discours [En ligne]* 11 (2012): 1-30. Web. 19 Jun. 2013.
- Hampshire, Stephen and Carmen Porta Salvia. "Translation and the Internet: Evaluating the Quality of Free Online Machine Translators." *Quaderns: revista de traducció* 17 (2010): 197-209.
- Henisz-Dostert, Bozena, R R. Macdonald, and Michael Zarechnak. *Machine Translation*. The Hague: Mouton, 1979. Print.
- Hutchins, W J, and H L. Somers. *An Introduction to Machine Translation*. London: Academic Press, 1992. Print.
- Ketzan, Erik. "Rebuilding Babel: Copyright and the Future of Online Machine Translation." *Tulane Journal of Technology & Intellectual Property* 9 (2007): 205-234. Web. 4 Jun. 2013.
- Koehn, Philipp. *Statistical Machine Translation*. New York : Cambridge University Press, 2010. Print.
- Kuras, Marcin. « Les mémoires de traduction : outil nécessaire ou source de problèmes pour les traducteurs ? ». *Romanica Cracoviensia* 11 (2011) : 234-240. Web. 4 Jul. 2013.

- L'Homme, Marie-Claude. *Initiation À La Traductique*. Brossard, Québec: Linguatex, 2000. Print.
- Loffler-Laurian, Anne-Marie. « Enquete stylistique sur la traduction automatique de document techniques ». *IRAL; International Review of Applied Linguistics in Language Teaching* 25:2 (1987) : 154-168. Web. 6 Jun. 2013.
- Loffler-Laurian, Anne-Marie. *La Traduction Automatique*. Presses Universitaires du Septentrion, 1996. Print.
- Lopez, Adam. "Statistical Machine Translation." *ACM Computing Surveys (CSUR)* 40.3 (2008): 1-49. Web. 6 Dec. 2012.
- Microsoft. "Bing Translator." *www.bing.com/translator*, 2013. Web. 2012.
- Microsoft. "Delivering the Power of Instant Translation and Language Services Anytime, Anywhere." (2012): 1-2. Web. 8 Apr. 2013.
- Microsoft. "Microsoft Translator (and Bing Translator) Official Team Blog." *blogs.msdn.com/b/translation*, 2013. Web. 3 Dec. 2012.
- Nagao, Makoto. *Machine Translation: How Far Can It Go?* Oxford: Oxford University Press, 1989. Print.
- Nida, Eugene A. "Science of Translation." *Language*. 45.3 (1969): 483-498. Web. 14 Jun. 2013.
- Och, F. "Statistical Machine Translation: From Single Word Models to Alignment Templates." Ph.D. thesis, RWTH Aachen, Germany. (2002): 1-145. Web. 23 May. 2013.
- Olvera-Lobo, María-Dolores, and Lola García-Santiago. "Analysis Of Errors In The Automatic Translation Of Questions For Translingual QA Systems." *Journal of Documentation*. 66.3 (2010): 434-455. *Academic Search Premier*. Web. 14 Nov. 2011.
- Popović, Maja, and Hermann Ney. "Towards Automatic Error Analysis of Machine Translation Output." *Computational Linguistics* 37.3 (2011): 1-50. *Academic Search Complete*. Web. 26 Nov. 2012.
- Prompt. "Prompt Translator." *www.online-translator.com*, 2013. Web. 2012.
- Ranchhod, Elisabete Marques. "Problèmes de traduction automatique des constructions à verbes supports". *Linguisticae investigations*. 23.2 (2000): 253-267. *ALJC Swetswise*. Web. 14 Nov. 2011.

- Rossi, Laura, and Dion Wiggins. "Applicability and application of Machine Translation quality metrics in the patent field." *World Patent Information*. 35 (2013): 115-125. Web. 16 Apr. 2013.
- Seljan, Sanja, Tomislav Vičić, and Marija Brkić. "BLEU Evaluation of Machine Translated English-Croatian Legislation." *Language Resources and Evaluation* (2012): 2143-2148. Web. 8 Apr. 2013.
- Seljan, Sanja. et al. "Evaluation of Free Online Machine Translations for Croatian English and English-Croatian Language Pairs." *Proceedings of the 3rd International Conference on the Future of Information Sciences: INFUTURE2011 - Information Sciences and e-Society*. (2011): 331-345. Web. 8 Apr. 2013.
- Slocum, Jonathan. "Machine Translation." *Computers and the Humanities* 19.2 (1985): 109-116. Web. 06 Jun. 2013.
- Sokolova, Svetlana. "How the computer translates." www.promt.com/company/technology/pdf/e_how_computer_translates_sokolova.pdf, 2009. St. Petersburg, Russia: PROMT Ltd. Web. 2 Apr. 2013.
- "TRADOS 7 Sets New Industry Standard for Global Information Solutions." *Canada NewsWire*: 1. Jun 22 2005. *ProQuest*. Web. 31 Mar. 2013.
- Translation.net. "Déjà Vu Translation Memory Software for Windows." www.translation.net/deja_vu_x.html, 2013. Web. 1 April. 2013. Université de Lorraine. « Trésor de la langue française ». <http://atilf.atilf.fr>, 2013. Web. 2013.
- Way, Andy. "Machine Translation." *The Handbook of Computational Linguistics and Natural Language Processing* : 19 (2009): 531-573. Web. 3 Jun. 2013.
- Wordfast. "Translation Memory Software from Wordfast." www.wordfast.com/index.html, 2012. Web. 1 April. 2013.
- 7Brands. "Translation Software Review." www.7brands.com/top_translation_software.htm, 2013. Web. 1 Apr. 2013.

APPENDICE A Corpus

Extrait 1 :

The International Finance Corporation, part of the World Bank Group, fosters economic growth in the developing world by financing private sector investments, mobilizing capital in the international financial markets, and providing technical assistance and advice to governments and businesses (BankGloss, 1302).

[1-1] BankGloss

La Société financière internationale, qui est membre du Groupe de la Banque mondiale, favorise la croissance économique des pays en développement en finançant des investissements dans le secteur privé, en mobilisant des ressources sur les marchés des capitaux internationaux et en fournissant une assistance technique et des services-conseils aux gouvernements et aux entreprises.

[1-2] Promt

La Société de Finance internationale, la partie du Groupe de Banque Mondial, encourage économique la croissance dans le monde se développant en finançant des investissements de secteur privé, en mobilisant la capitale sur les marchés financiers internationaux et la fourniture de l'assistance technique et le conseil aux gouvernements et aux entreprises.

Erreurs grammaticales :

[1-1] la Banque mondiale
[1-2] Banque Mondial
Faute d'accord (nom-adjectif)
Omission de mot (déterminant)

[1-1] fournissant
[1-2] la fourniture
Utilisation de mots appartenant à une mauvaise partie du discours (verbe et substantif).

[1-1] La Société financière
[1-2] La Société de Finance
Faute de ponctuation

[1-1] la Banque mondiale
[1-2] Banque Mondial
Faute de ponctuation

[1-1] la croissance économique
[1-2] économique la croissance
Mauvaise ordre des mots (du nom et de l'adjectif qualificatif)

[1-1] des pays en développement
[1-2] dans le monde se développant
Utilisation de préposition erronée

[1-1] dans le secteur privé
[1-2] de secteur privé
Utilisation de préposition erronée

Erreurs lexicales :

[1-1] des ressources
[1-2] la capitale
'ressource' : Moyen permettant de se tirer d'embaras ou d'améliorer une situation difficile.
'capitale' : Ville principale d'un État, d'une province ou d'une étendue de pays qui est le siège du gouvernement et/ou de l'administration centrale.
Choix de terme inapproprié (le nom devrait être au masculin : le capitale)

Erreurs sémantiques :

[1-1] La Société financière
[1-2] La Société de Finance
Imprécision

[1-1] membre
[1-2] la partie
'membre' : élément constitutif ou portion d'un tout. (*Le Trésor de la langue française (TLFi)*)
'partie' : Partie (constituée d'un ou de plusieurs éléments) d'un ensemble. (*TLFi*)
Faux-sens : les deux mots ne sont pas équivalents.

[1-1] favorise
[1-2] encourage
'favorise' : Traiter (quelqu'un) avec une bienveillance particulière.
'encourage' : Traiter (quelqu'un) avec une bienveillance particulière.
Faux-sens : les deux mots ne sont pas équivalents.

[1-1] des services-conseils
[1-2] le conseil
Faux-sens

[1-3] Babylon

La société de finance internationale, la partie du Groupe de la Banque Mondial, encourager économique la croissance dans le monde se développant en finançant des investissements du secteur privé, en mobilisant la capitale sur les marchés financiers internationaux et la fourniture de l'assistance technique et le conseil aux têtes folles et aux entreprises.

Erreurs grammaticales :

[1-1] la Banque mondiale
[1-3] la Banque Mondial
Faute d'accord (nom-adjectif)

[1-1] favorise la croissance économique
[1-3] encourager économique la croissance
Faute de conjugaison (choix de temps)

[1-1] des pays en développement
[1-3] dans le monde se développant
Omission de mot (absence de préposition avant « se développant »)

[1-1] en fournissant
[1-3] la fourniture
Utilisation de mots appartenant à une mauvaise partie du discours (verbe et substantif)

[1-1] la Banque mondiale
[1-3] la Banque Mondial
Faute de ponctuation

[1-1] la croissance économique
[1-3] économique la croissance
Mauvaise ordre des mots (du nom et de l'adjectif qualificatif)

[1-1] des pays en développement
[1-3] dans le monde se développant
Utilisation de préposition erronée avec 'dans'

[1-1] dans le secteur privé
[1-3] du secteur privé
Utilisation de préposition erronée

[1-1] en fournissant
[1-3] la fourniture
Utilisation erronée d'un article

Erreurs lexicales :

[1-1] des ressources
[1-3] la capitale
'ressource' : Moyen permettant de se tirer d'embaras ou d'améliorer une situation difficile.
'capitale' : Ville principale d'un État, d'une province ou d'une étendue de pays qui est le siège du gouvernement et/ou de l'administration centrale.
Choix de terme inapproprié (le nom devrait être au masculin : le capitale)

Erreurs sémantiques :

[1-1] La Société financière
[1-3] La société de finance
Imprécision

[1-1] membre
[1-3] la partie

'membre' : élément constitutif ou portion d'un tout. (*Le Trésor de la langue française (TLFi)*)

'partie' : Partie (constituée d'un ou de plusieurs éléments) d'un ensemble. (*TLFi*)

Faux-sens

[1-1] favorise

[1-3] encourager

'favorise' : Traiter (quelqu'un) avec une bienveillance particulière.

'encourage' : Traiter (quelqu'un) avec une bienveillance particulière.

Faux-sens : les deux mots ne sont pas équivalents.

[1-1] des services-conseils

[1-3] le conseil

Faux-sens

[1-1] gouvernements

[1-3] têtes folles

'gouvernement' : Action d'administrer (quelque chose).

Non-sens

[1-4] Google Translate

La Société financière internationale, qui fait partie du Groupe de la Banque mondiale, favorise l'économie la croissance dans le monde en développement en finançant des investissements du secteur privé, en mobilisant aide à l'investissement dans les marchés financiers internationaux, et fournir une assistance technique et des conseils aux gouvernements et aux entreprises.

Erreurs grammaticales :

[1-1] en mobilisant des ressources

[1-4] en mobilisant aide

Omission de mot (déterminant)

[1-1] fournissant

[1-4] fournir

Faute de conjugaison (choix de temps)

[1-1] la croissance économique
[1-4] l'économie la croissance
Mauvaise ordre des mots (du nom et de l'adjectif qualificatif)

[1-1] dans le secteur privé
[1-4] du secteur privé
Utilisation de préposition erronée

Erreurs sémantiques :

[1-1] en mobilisant des ressources
[1-4] en mobilisant aide
Faux-sens

[1-5] Bing

Partie du groupe de la Banque mondiale, l'International Finance Corporation, favorise l'économique croissance dans les pays en développement en finançant des investissements du secteur privé, mobilisation capital dans les marchés financiers internationaux et fournir une assistance technique et conseils aux gouvernements et aux entreprises.

Erreurs grammaticales :

[1-1] en mobilisant
[1-5] mobilisation
Mot appartenant à une mauvaise partie du discours (verbe et substantif)

[1-1] en fournissant
[1-5] fournir
Faute de conjugaison (choix de temps)

[1-1] La Société financière internationale
[1-5] l'International Finance Corporation
Mauvaise ordre des mots (du nom et de l'adjectif qualificatif)

[1-1] la croissance économique
[1-5] l'économique croissance
Mauvaise ordre des mots (du nom et de l'adjectif qualificatif)

[1-1] des investissements dans le secteur privé
[1-5] des investissements du secteur privé
Utilisation de préposition erronée

Erreurs lexicales :

[1-1] La Société financière internationale
[1-5] l'International Finance Corporation
Anglicisme

Erreurs sémantiques :

[1-1] membre
[1-5] partie
'membre' : élément constitutif ou portion d'un tout. (*Le Trésor de la langue française (TLFi)*)
'partie' : Partie (constituée d'un ou de plusieurs éléments) d'un ensemble. (*TLFi*)
Faux-sens

[1-1] services-conseils
[1-5] conseils
Faux-sens

Extrait 2 :

IFC is strategically shifting our business toward “frontier markets” and developing innovative products for these regions of the world (BankGloss, 1307).

[2-1] BankGloss

La SFI a pour stratégie d'orienter son action vers les « marchés pionniers » et de créer des produits novateurs pour ces régions du globe.

[2-2] Prompt

IFC déplace stratégiquement nos affaires vers les "marchés de frontière" et développe des produits innovateurs pour ces régions du monde.

Erreurs grammaticales :

[2-1] La SFI

[2-2] IFC

Omission de mot (déterminant)

[2-1] « ... »

[2-2] "..."

Faute de ponctuation

Erreurs lexicales :

[2-1] créer

[2-2] développe

‘créer’: Donner l'existence à.

‘développer’ : Ôter ce qui enveloppe (quelque chose) de manière à en étaler le contenu.

Anglicisme (« développer » pour « *developing* »)

Erreurs sémantiques :

[2-1] a pour stratégie d’orienter son action

[2-2] déplace stratégiquement nos affaires

Faux-sens

[2-1] marchés pionniers

[2-2] marchés de frontière

Marché frontière = « border market »

Faux-sens

[2-3] Babylon

La SFI est stratégiquement passage notre entreprise vers "marchés frontaliers" et au développement de produits innovateurs pour ces régions du monde.

Erreurs grammaticales :

[2-1] a pour stratégie d'orienter son action vers
[2-3] est stratégiquement passage notre entreprise vers
Omission de mot (préposition)

[2-1] « ... »

[2-3] "..."

Faute de ponctuation

Erreurs lexicales :

[2-1] créer

[2-3] développement

'créer': Donner l'existence à.

'développer' : Ôter ce qui enveloppe (quelque chose) de manière à en étaler le contenu.

Anglicisme (« développer » pour « *developing* »)

Erreurs sémantiques :

[2-1] marchés pionniers

[2-3] marchés frontaliers

Marché frontalier = « border market »

Faux-sens

[2-4] Google Translate

La SFI est stratégiquement passer, dans notre entreprise vers les «marchés pionniers» et développer des produits innovants pour ces régions du monde.

Erreurs grammaticales :

[2-1] a pour stratégie d'orienter son action

[2-4] est stratégiquement passer, dans notre entreprise

Faute d'accord

[2-1] a pour stratégie d'orienter son action

[2-4] est stratégiquement passer, dans notre entreprise

Mauvaise ordre des mots (du nom et du modificateur)

[2-1] « marchés pionniers »

[2-4] «marchés pionniers»

Faute de ponctuation (absence d'espace entre les guillemets et les mots)

Erreurs lexicales :

[2-1] créer

[2-4] développer

'créer': Donner l'existence à.

'développer' : Ôter ce qui enveloppe (quelque chose) de manière à en étaler le contenu.

Anglicisme (« développer » pour « *developing* »)

Erreurs sémantiques :

[2-1] a pour stratégie d'orienter son action

[2-4] est stratégiquement passer, dans notre entreprise

Faux-sens

[2-5] Bing

SFI est stratégiquement changeante de notre entreprise vers les « marchés frontières » et développer des produits innovants pour ces régions du monde.

Erreurs grammaticales :

[2-1] La SFI

[2-5] SFI

Omission de mot (déterminant)

[2-1] de créer

[2-5] développer

Omission de mot (préposition)

Erreurs lexicales :

[2-1] créer

[2-5] développer

'créer': Donner l'existence à.

'développer' : Ôter ce qui enveloppe (quelque chose) de manière à en étaler le contenu.

Anglicisme (« développer » pour « *developing* »)

Erreurs sémantiques :

[2-1] marchés pionniers

[2-5] marchés frontières*

Marché frontière = « border market »

Faux-sens

Extrait 3 :

IFC's role will be to fuel innovation, economic entrepreneurship, and development in areas of the world that many in the private sector deem high risk.

(BankGloss, 1307)

[3-1] BankGloss

Elle devra stimuler l'innovation, l'esprit d'entreprise et le développement dans des contrées et secteurs jugés très risqués par le secteur privé.

[3-2] Promt

Le rôle d'IFC devra alimenter l'innovation, l'entreprise économique et le développement dans les régions du monde que beaucoup dans le secteur privé jugent le haut risque.

Erreurs grammaticales :

[3-1] Elle devra
[3-2] Le rôle d'IFC devra
Omission de mot (déterminant)

[3-1] très risqués
[3-2] le haut risque
Mots appartenant à une mauvaise partie du discours (adjectif et substantif).

Erreurs sémantiques :

[3-1] stimuler
[3-2] alimenter
Faux-sens (traduction littérale de « fuel »)

[3-1] l'esprit d'entreprise
[3-2] l'entreprise économique
Faux-sens (« entreprenariat économique »)

[3-3] Babylon

La SFI aura son rôle à jouer pour stimuler l'innovation, l'esprit d'entreprise, et le développement dans des régions du monde que de nombreux dans le secteur privé jugent risque élevé.

Erreurs grammaticales :

[3-1] par le secteur privé
[3-3] de nombreux dans le secteur privé
Utilisation de préposition erronée

Erreurs sémantiques :

[3-1] Elle devra
[3-3] La SFI aura son rôle à jouer
Faux-sens

[3-4] Google Translate

Rôle de la SFI sera de stimuler l'innovation, l'entrepreneuriat économique, et le développement dans les régions du monde que beaucoup dans le secteur privé jugent à haut risque.

Erreurs grammaticales :

[3-1] Elle devra
[3-4] Rôle de la SFI
Omission de mot (déterminant)

[3-5] Bing

Rôle de la SFI sera de carburant innovation, l'entrepreneuriat économique et de développement dans les régions du monde que beaucoup dans le secteur privé jugent à risque élevé.

Erreurs grammaticales :

[3-1] Elle devra
[3-5] Rôle de la SFI sera
Omission de mot(déterminant)

[3-1] stimuler l'innovation
[3-5] carburant innovation
Mots appartenant à une mauvaise partie du discours (verbe et substantif)

Erreurs sémantiques :

[3-1] stimuler l'innovation
[3-5] carburant innovation
'stimuler' : Mettre quelqu'un ou quelque chose dans les conditions propres à le faire agir ou réagir; susciter ou renforcer un mouvement.
'carburant' : Qui carbure.
Faux-sens

Extrait 4 :

An analysis of the Corporation revealed that we have both the capital and the know-how to succeed in these frontier countries. (BankGloss, 1307)

[4-1] BankGloss

Selon une analyse de la Société, cette dernière a les ressources financières et l'expertise requise pour mener à bien son action dans ces pays.

[4-2] Promt

Une analyse de la Société a révélé que nous avons tant la capitale que le savoir-faire pour réussir à ces pays de frontière.

Erreurs grammaticales :

[4-1] Selon une analyse de la Société
[4-2] Une analyse de la société a révélé
Faute de ponctuation (« Société » devrait être minuscule)

Erreurs sémantiques :

[4-1] les ressources financières
[4-2] la capitale
Faux-sens (la capitale = la ville principale ; le capital = les ressources financières)

[4-1] ces pays
[4-2] ces pays de frontière
Faux-sens

[4-3] Babylon

Une analyse de la société a révélé que nous avons à la fois le capital et le savoir-faire pour réussir dans ces pays frontaliers.

Erreurs grammaticales :

[4-1] Selon une analyse de la Société
[4-3] Une analyse de la société a révélé
Faute de ponctuation (« Société » devrait être minuscule)

Erreurs sémantiques :

[4-1] dans ces pays
[4-3] dans ces pays frontaliers
Faux-sens

[4-4] Google Translate

Une analyse de la Corporation a révélé que nous avons à la fois la capitale et le savoir-faire pour réussir dans ces pays frontaliers.

Erreurs grammaticales :

[4-1] Selon une analyse de la Société
[4-4] Une analyse de la Corporation a révélé
Faute de ponctuation (« Corporation » devrait être minuscule)

Erreurs sémantiques :

[4-1] les ressources financières
[4-4] la capitale
Faux-sens (la capitale = la ville principale ; le capital = les ressources financières)

[4-1] ces pays
[4-4] ces pays frontaliers
Faux-sens (Il y a aussi une manque de cohérence interne - avant, on a utilisé l'adjectif « pionniers » pour le même adjectif anglais, *frontier*. Ici on a utilisé l'adjectif « frontaliers »)

[4-5] Bing

Une analyse de la société a révélé que nous avons à la fois la capitale et le savoir-faire pour réussir dans ces pays de frontière.

Erreurs sémantiques :

[4-1] les ressources financières

[4-5] la capitale

Faux-sens (la capitale = la ville principale ; le capital = les ressources financières)

[4-1] ces pays

[4-5] ces pays de frontière

Faux-sens

Extrait 5 :

We are reshaping how we work. In response to our clients' criticisms, we are overhauling our business processes to improve our development effectiveness.

We are eliminating bureaucratic procedures while flattening hierarchical structures. (BankGloss, 1309)

[5-1] BankGloss

Nous changeons notre manière de travailler. Face aux critiques de nos clients, nous avons entrepris de modifier nos procédures internes pour accroître l'efficacité de notre action en faveur du développement. Nous avons entrepris d'éliminer les procédures bureaucratiques en aplatissant la structure hiérarchique.

[5-2] Promt

Nous remodelons comment nous travaillons. En réponse aux critiques de nos clients, nous révisons nos processus d'affaires pour améliorer notre efficacité de développement. Nous éliminons des procédures bureaucratiques en écrasant des structures hiérarchiques.

Erreurs sémantiques :

[5-1] procédures internes
[5-2] processus d'affaires
Faux-sens

[5-1] l'efficacité de notre action en faveur du développement
[5-2] notre efficacité de développement
Ajout

[5-1] aplatissant
[5-2] écrasant
Faux-sens

[5-3] Babylon

Nous remodelons comment nous travaillons. En réponse à nos clients des critiques, nous sommes en train de réviser nos processus métier afin d'améliorer notre efficacité en matière de développement. Nous éliminons les formalités bureaucratiques inutiles tout en aplatissement structures hiérarchiques.

Erreurs grammaticales :

[5-1] aplatissant
[5-3] aplatissement
Mots appartenant à une mauvaise partie du discours (verbe et substantif).

[5-1] la structure hiérarchique
[5-3] structures hiérarchiques
Omission de mot (déterminant)

Erreurs sémantiques :

[5-1] critiques de nos clients
[5-3] nos clients des critiques
Faux-sens

[5-1] procédures internes
[5-3] processus métier
Faux-sens

[5-4] Google Translate

Nous sommes modifie la façon dont nous travaillons. En réponse aux critiques de nos clients, nous sommes refonte de nos processus d'affaires afin d'améliorer l'efficacité de notre développement. Nous sommes en éliminant les procédures bureaucratiques tout en aplatissant les structures hiérarchiques.

Erreurs grammaticales :

[5-1] Nous changeons
[5-4] Nous sommes modifie
Faute de conjugaison

[5-1] nous avons entrepris de modifier
[5-4] nous sommes refonte
Faute de conjugaison

[5-1] Nous avons entrepris d'éliminer
[5-4] Nous sommes en éliminant
Faute de conjugaison

Erreurs sémantiques :

[5-1] Nous changeons
[5-4] Nous sommes modifie

Ajout (d'auxiliaire)

[5-1] nous avons entrepris de modifier

[5-4] nous sommes refonte

Ajout (d'auxiliaire)

[5-1] procédures internes

[5-4] processus d'affaires

Faux-sens

[5-1] notre action en faveur du développement

[5-4] notre développement

Faux-sens

[5-1] Nous avons entrepris d'éliminer

[5-4] Nous sommes en éliminant

Ajout (d'auxiliaire)

[5-5] Bing

Nous sommes remodelage comment nous travaillons. En réponse aux critiques de nos clients, nous sommes refonte de nos processus opérationnels afin d'améliorer notre efficacité de l'aide. Nous éliminons les procédures bureaucratiques tout en aplatissant les structures hiérarchiques.

Erreurs grammaticales :

[5-1] Nous changeons

[5-5] Nous sommes remodelage

Faute de conjugaison

[5-1] nous avons entrepris de modifier

[5-5] nous sommes refonte

Faute de conjugaison

[5-1] l'efficacité de notre action

[5-5] notre efficacité de l'aide

Mauvaise ordre des mots (du nom et du modificateur)

Erreurs sémantiques :

[5-1] Nous changeons
[5-5] Nous sommes remodelage
Ajout (d'auxiliaire)

[5-1] nous avons entrepris de modifier
[5-5] nous sommes refonte
Ajout (d'auxiliaire)

[5-1] l'efficacité de notre action
[5-5] notre efficacité de l'aide
Faux-sens

Extrait 6 :

To be more responsive, IFC is also seeking to increase delegation of decision making downward throughout the organization. (BankGloss, 1316)

[6-1] BankGloss

Pour pouvoir mieux répondre aux besoins de ses clients, la SFI s'efforce aussi de davantage déléguer le pouvoir de prise de décision à tous les niveaux de l'institution.

[6-2] Promt

Pour être plus affectueux, IFC cherche aussi à augmenter la délégation de prise de décision en bas partout dans l'organisation.

Erreurs grammaticales :

[6-1] la SFI
[6-2] IFC
Omission de mot (déterminant)

Erreurs sémantiques :

[6-1] pouvoir mieux répondre
[6-2] être plus affectueux
Faux-sens

[6-1] à tous les niveaux
[6-2] en bas partout
Faux-sens

[6-3] Babylon

Pour être plus réactif, la SFI s'efforce également d'accroître la délégation de la prise de décision vers le bas dans toute l'organisation.

Erreurs grammaticales :

[6-1] Pour pouvoir mieux répondre aux besoins de ses clients, la SFI
[6-3] Pour être plus réactif, la SFI
Faute d'accord (nom-adjectif : l'adjectif devrait être au féminin (*réactive*))

[6-4] Google Translate

Pour être plus réactif, l'IFC cherche également à accroître la délégation de la prise de décision à la baisse dans toute l'organisation.

Erreurs sémantiques :

[6-1] à tous les niveaux de l'institution
[6-4] à la baisse dans toute l'organisation
Faux-sens

[6-5] Bing

Pour être plus réactif, IFC cherche également à accroître la délégation de la prise de décisions à la baisse dans l'ensemble de l'organisation.

Erreurs grammaticales :

[6-1] la SFI

[6-5] IFC

Omission de mot(déterminant)

Erreurs sémantiques :

[6-1] à tous les niveaux de l'institution

[6-5] à la baisse dans l'ensemble de l'organisation

Faux-sens

Extrait 7 :

IFC is reorganizing its investment departments to better systematize and oversee management of a portfolio that has grown from \$3 billion to \$9 billion since 1990.

(BankGloss, 1316)

[7-1] BankGloss

Elle a entrepris de réorganiser ses départements des placements pour mieux rationaliser et superviser la gestion d'un portefeuille qui est passé de 3 milliards de dollars en 1990 à 9 milliards de dollars à l'heure actuelle.

[7-2] Promt

IFC réorganise ses départements d'investissement pour mieux systématiser et superviser la direction d'un portefeuille qui a grandi de \$3 milliards à \$9 milliards depuis 1990.

Erreurs grammaticales :

[7-1] de 3 milliards de dollars...à 9 milliards de dollars

[7-2] de \$3 milliards à \$9 milliards

Faute de ponctuation (signe « \$ » pour « dollars »)

Mauvaise ordre des mots (du complément de nom (\$))

[7-3] Babylon

La SFI est la réorganisation des services de placement pour mieux systématiser et superviser la gestion d'un portefeuille est passée de 3 milliards de dollars à 9 milliards de dollars depuis 1990.

Erreurs grammaticales :

[7-1] entrepris de réorganiser

[7-3] est la réorganisation

Utilisation de mots appartenant à une mauvaise partie du discours (verbe et substantif).

[7-1] qui est passé

[7-3] --- est passé

Omission de mot (pronom relatif)

[7-4] Google Translate

La SFI est la réorganisation de ses services de placement afin de mieux

systematiser et de superviser la gestion d'un portefeuille qui est passé de 3 milliards de dollars 9 milliards de dollars depuis 1990.

Erreurs grammaticales :

[7-1] entrepris de réorganiser

[7-4] est la réorganisation

Mot appartenant à une mauvaise partie du discours (verbe et substantif).

[7-1] de 3 milliards de dollars...à 9 milliards de dollars

[7-4] de 3 milliards de dollars 9 milliards de dollars

Omission de mot (préposition)

Remarque : manque de cohérence avec le nom, « SFI », qui est traduit par « IFC » dans la phrase précédente.

[7-5] Bing

IFC réorganise son investissement ministères afin de mieux systematisent et supervisent la gestion d'un portefeuille qui est passé de \$3 milliards à 9 milliards \$ depuis 1990.

Erreurs grammaticales :

[7-1] rationaliser

[7-5] systematisent

Faute de conjugaison

[7-1] superviser

[7-5] supervisent

Faute de conjugaison

[7-1] départements des placements

[7-5] investissement ministères

Mauvaise ordre des mots (le modificateur, « investissement » devrait être placé après le nom)

[7-1] de 3 milliards de dollars...à 9 milliards de dollars
[7-5] de \$3 milliards à 9 milliards \$
Faute de ponctuation (signe « \$ » pour « dollars »)
Mauvaise ordre des mots (du complément de nom (\$))

Erreurs sémantiques :

[7-1] départements des placements
[7-5] investissement ministères
Faux-sens (départements ≠ ministères)

Extrait 8 :

IFC may take a maximum 35 percent stake in a venture but generally subscribes to between 5 and 15 percent of a project's equity. IFC is never the single largest shareholder in a project. (BankGloss, 1317)

[8-1] BankGloss

La SFI peut prendre une participation à hauteur de 35% dans une opération, mais se limite généralement à un pourcentage compris entre 5 et 15 % du capital du projet. Elle n'est jamais le principal actionnaire d'un projet.

[8-2] Promt

IFC peut prendre un enjeu maximum de 35 pour cent dans une entreprise, mais souscrit généralement à entre 5 et 15 pour cent de l'équité d'un projet. IFC n'est jamais l'actionnaire simple le plus grand dans un projet.

Erreurs grammaticales :

[8-1] La SFI
[8-2] IFC
Omission de mot(déterminant)

[8-1] Elle
[8-2] IFC
Omission de mot(déterminant)

[8-1] le principal actionnaire
[8-2] l'actionnaire simple le plus grand
Mauvaise ordre des mots (du nom et de l'adjectif)

Erreurs sémantiques :

[8-1] une participation
[8-2] un enjeu
Faux-sens

[8-1] opération
[8-2] entreprise
Faux-sens

[8-1] du capital
[8-2] de l'équité
Faux-sens

[8-1] le principal actionnaire
[8-2] l'actionnaire simple le plus grand
Faux-sens

[8-3] Babylon

La SFI peut prendre un maximum de 35 % dans une coentreprise mais généralement souscrit à entre 5 et 15 % des fonds d'un projet de l'équité. La SFI est jamais le plus gros actionnaire dans un projet.

Erreurs grammaticales :

[8-1] une participation
[8-3] --
Omission de mot (nom)

[8-1] n'est jamais
[8-3] est jamais
Omission de mot (préposition - l'adverbe négatif)

[8-1] une opération, mais
[8-3] une coentreprise mais
Faute de ponctuation (absence du virgule avant « mais »)

[8-1] se limite généralement à
[8-3] généralement souscrit à
Mauvaise ordre des mots (du verbe et de l'adverbe)

Erreurs sémantiques :

[8-1] une opération
[8-3] une coentreprise
Faux-sens

[8-1] du capital du projet
[8-3] des fonds d'un projet de l'équité
Ajout (« de l'équité »)

[8-4] Google Translate

La SFI peut prendre une participation maximale de 35 pour cent dans une entreprise, mais souscrit généralement entre 5 et 15 pour cent des actions d'un projet. IFC n'est jamais le plus grand actionnaire unique dans un projet.

Erreurs grammaticales :

[8-1] à un pourcentage compris entre 5 et 15 %
[8-4] entre 5 et 15 pour cent
Omission de mot (préposition)

[8-1] Elle
[8-4] IFC
Omission de mot (déterminant)

Erreurs sémantiques :

[8-1] opération
[8-4] entreprise
Faux-sens

[8-5] Bing

Ceci pourrait prendre une participation maximale de 35 pour cent dans une entreprise mais souscrit généralement à entre 5 et 15 % des capitaux propres du projet. L'IFC n'est jamais le seul actionnaire dans un projet.

Erreurs grammaticales :

[8-1] peut
[8-5] pourrait
Faute de conjugaison (choix de temps)

[8-1] le principal actionnaire
[8-5] le seul actionnaire
Omission de mot(adjectif)

[8-1] opération, mais
[8-5] entreprise mais
Faute de ponctuation (absence du virgule)

Remarque : Manque de cohérence avec le signe de pourcentage.

Erreurs sémantiques :

[8-1] La SFI
[8-5] Ceci
Ambiguïté

[8-1] opération
[8-5] entreprise
Faux-sens

[8-1] le principal actionnaire
[8-5] le seul actionnaire
Faux-sens

Extrait 9 :

“We can move very quickly in response to a country’s evolution. If a country wants its private sector to develop, wants its financial sector to develop, we roll up our sleeves and try to do it. However, it takes two to tango. It is up to the government to set in place hospitable policies.” (BankGloss, 1321)

[9-1] BankGloss

« Nous pouvons intervenir très rapidement quand la situation évolue. Si un pays veut développer son secteur privé, ou son secteur financier, nous retroussons nos manches et nous nous attelons à la tâche. Mais, comme sur la piste de danse, il faut être deux pour avancer. L’État doit mettre en place un cadre réglementaire favorable. »

[9-2] Prompt

“Nous pouvons bouger très vite en réponse à l'évolution d'un pays. Si un pays veut que son secteur privé se développe, veut que son secteur financier se développe, nous enroulons nos manches et essayons de le faire. Cependant, il prend deux au tango. Il est à la hauteur du gouvernement pour mettre dans l'endroit des politiques hospitalières.”

Erreurs grammaticales :

[9-1] «... »

[9-2] “...”

Faute de ponctuation

[9-1] L'État doit mettre en place

[9-2] Il est à la hauteur du gouvernement pour mettre dans l'endroit

Utilisation de préposition erronée

Erreurs sémantiques :

[9-1] intervenir

[9-2] bouger

Faux-sens

[9-1] retroussons nos manches

[9-2] enroulons nos manches

Faux-sens

[9-1] favorable

[9-2] hospitalières

Faux-sens

Erreurs stylistiques :

[9-1] Si un pays veut développer son secteur privé, ou son secteur financier

[9-2] Si un pays veut que son secteur privé se développe, veut que son secteur financier se développe

Construction boiteuse (répétition)

[9-1] comme sur la piste de danse, il faut être deux pour avancer
[9-2] il prend deux au tango
Absence d'idiomaticité

[9-1] L'État doit mettre en place
[9-2] Il est à la hauteur du gouvernement pour mettre dans l'endroit
Absence d'idiomaticité

[9-3] Babylon

" Nous pouvons déplacer très rapidement en réponse à une évolution du pays. Si un pays veut son secteur privé à développer, veut que son secteur financier de développer, nous retrousser nos manches et de tenter de le faire. Toutefois, il faut être deux pour danser le tango. C'est au gouvernement de mettre en place politiques hospitalière."

Erreurs grammaticales :

[9-1] nous retroussons
[9-3] nous retrousser
Faute de conjugaison (choix de temps)

[9-3] et nous nous attelons à la tâche
[9-3] et de tenter de le faire
Faute de conjugaison (choix de temps)

[9-1] mettre en place un cadre réglementaire
[9-3] mettre en place politiques
Omission de mot (déterminant)

[9-1] réglementaire favorable
[9-3] politiques hospitalière
Faute d'accord (nom-adjectif)

[9-1] «... »
[9-3] "..."
Faute de ponctuation

[9-1] veut développer son secteur privé
[9-3] veut son secteur privé à développer
Utilisation de préposition erronée

[9-1] ou son secteur financier
[9-3] veut que son secteur financier de développer
Utilisation de préposition erronée

Erreurs sémantiques :

[9-1] intervenir
[9-3] déplacer
Faux-sens

[9-1] favorable
[9-3] hospitalière
Faux-sens

Erreurs stylistiques :

[9-1] Si un pays veut développer son secteur privé, ou son secteur financier
[9-3] Si un pays veut son secteur privé à développer, veut que son secteur
financier de développer
Construction boiteuse (répétition)

[9-4] Google Translate

"Nous pouvons nous déplacer très rapidement en réponse à l'évolution d'un pays. Si un pays veut que son secteur privé à développer, veut que son secteur financier pour développer, nous retroussons nos manches et essayer de le faire. Cependant, il faut être deux pour danser le tango. Il appartient au gouvernement de mettre en place des politiques hospitalières. "

Erreurs grammaticales :

[9-1] nous nous attelons à la tâche
[9-4] nous...essayer de le faire
Faute de conjugaison (choix de temps)

[9-1] «...»
[9-4] "..."
Faute de ponctuation

[9-1] veut développer son secteur privé
[9-4] veut que son secteur privé à développer
Utilisation de préposition erronée

[9-1] ou son secteur financier
[9-4] veut que son secteur financier pour développer
Utilisation de préposition erronée

Erreurs sémantiques :

[9-1] intervenir
[9-4] nous déplacer
Faux-sens

[9-1] favorable
[9-4] hospitalières
Faux-sens

Erreurs stylistiques :

[9-1] Si un pays veut développer son secteur privé, ou son secteur financier
[9-4] Si un pays veut que son secteur privé à développer, veut que son secteur financier pour développer
Construction boiteuse (répétition)

[9-5] Bing

"Nous pouvons déplacer très rapidement en réponse à l'évolution du pays. Si un pays veut que son secteur financier afin de développer son secteur privé à développer, veut, nous retroussons nos manches et essayez de le faire.
Toutefois, il faut être deux pour danser le tango. Il appartient au gouvernement à mettre en place des politiques hospitalier. »

Erreurs grammaticales :

[9-1] et nous nous attelons à la tâche

[9-5] et essayez de le faire

Faute de conjugaison

[9-1] un cadre réglementaire favorable

[9-5] des politiques hospitalier

Faute d'accord (nom-adjectif)

[9-1] «

[9-5] "

Faute de ponctuation (pas cohérent non plus puisque les citations sont écrites correctement à la fin)

[9-1] veut développer son secteur privé ou son secteur financier

[9-5] veut que son secteur financier afin de développer son secteur privé à développer, veut

Mauvaise ordre des mots (du nom et du verbe)

[9-1] L'État doit

[9-5] Il appartient au gouvernement à

Utilisation de préposition erronée

Erreurs sémantiques :

[9-1] intervenir

[9-5] déplacer

Faux-sens

[9-1] favorable
[9-5] hospitalier
Faux-sens

Erreurs stylistiques :

[9-1] Si un pays veut développer son secteur privé, ou son secteur financier
[9-5] Si un pays veut que son secteur financier afin de développer son secteur privé à développer, veut,
Construction boiteuse (répétition)

Extrait 10 :

The governor set his sights high: privatize the state power utility, structure and bid out an independent power project, and set up a regulatory agency. IFC accomplished all of this in a relatively quick nine months. At auction, Ceará reaped \$880 million for the utility, 27 percent above the minimum price.

(BankGloss, 1327)

[10-1] BankGloss

Le gouverneur s'est fixé d'ambitieux objectifs : il a en effet décidé de privatiser la compagnie d'électricité de l'État, de structurer un projet de production d'énergie indépendante, de lancer un appel d'offres en vue de sa réalisation, et de mettre en place un organisme réglementaire. La SFI s'est acquittée de cette tâche en seulement neuf mois. La vente aux enchères de la compagnie d'électricité a rapporté à l'État 880 millions de dollars, soit 27 % de plus que le prix minimum demandé.

[10-2] Prompt

Le gouverneur avait des visées haut : privatisez l'utilité de pouvoir publique, la structure et offrez un projet de pouvoir indépendant et remettez une agence de contrôle sur pied. IFC a accompli tout cela dans des neuf mois relativement rapides. À la vente aux enchères, Ceará a moissonné \$880 millions pour l'utilité, à 27 pour cent au-dessus du prix minimal.

Erreurs grammaticales :

[10-1] Le gouverneur s'est fixé d'ambitieux objectifs

[10-2] Le gouverneur avait des visées haut

Faute d'accord (nom-adjectif)

[10-1] la compagnie d'électricité de l'État

[10-2] l'utilité de pouvoir publique

Faute d'accord (nom-adjectif)

[10-1] il a en effet décidé de privatiser...de structurer...de lancer...de mettre

[10-2] privatisez...la structure et offrez...et remettez

Faute de conjugaison (choix de temps)

[10-1] de structurer

[10-2] la structure

Mots appartenant à une mauvaise partie du discours (substantif et verbe)

[10-1] un projet de production d'énergie indépendante

[10-2] un projet de pouvoir indépendant

Omission de mot (préposition et nom : « de production »)

[10-1] La SFI

[10-2] IFC

Omission de mot (déterminant)

[10-1] en...neuf mois

[10-2] dans des neuf mois

Utilisation de préposition erronée (« dans »),

Utilisation d'article erroné (« des »)

[10-1] 880 millions de dollars
[10-2] \$880 millions pour l'utilité
Faute de ponctuation (signe « \$ » pour « dollars »),
Mauvaise ordre des mots (du complément de nom (\$))

[10-1] soit 27 % de plus que
[10-2] à 27 pour cent au-dessus du
Utilisation de préposition erronée (« à »)

Erreurs lexicales :

[10-1] la compagnie d'électricité de l'État
[10-2] l'utilité de pouvoir publique
Anglicisme (« utilité » pour « *utility* »)
Anglicisme (« pouvoir » pour « *power* »)

[10-1] un projet de production d'énergie indépendante
[10-2] un projet de pouvoir indépendant
Anglicisme (« pouvoir » pour « *power* »)

[10-1] La vente aux enchères de la compagnie d'électricité a rapporté à l'État 880 millions de dollars
[10-2] À la vente aux enchères, Ceará a moissonné \$880 millions pour l'utilité
Anglicisme (« utilité » pour « *utility* »)

Erreurs sémantiques :

[10-1] de mettre en place un organisme réglementaire
[10-2] remettez une agence de contrôle sur pied
Faux-sens (« remettre sur pied » au lieu de « mettre en place »)

[10-1] La vente aux enchères a rapporté à l'État
[10-2] À la vente aux enchères, Ceará a moissonné
Faux-sens (choix malheureux du verbe « moissonner »)

[10-3] Babylon

Le gouverneur définir ses sites haute: privatiser le pouvoir de l'état utilitaire, la structure et les soumissions à un pouvoir indépendant projet, et de mettre en place un organisme de réglementation. La SFI accompli tout cela dans un relativement rapide neuf mois. Dans le cadre d'une vente aux enchères, Ceará récolté 880 millions de dollars pour l'utilitaire, 27 pour cent au-dessus du prix minimal.

Erreurs grammaticales :

[10-1] Le gouverneur s'est fixé

[10-3] Le gouverneur définir

Faute de conjugaison (choix de temps)

[10-1] d'ambitieux objectifs

[10-3] ses sites haute

Faute d'accord (nom-adjectif)

[10-1] de privatiser...de structurer...de lancer...de mettre

[10-3] privatiser...la structure et les soumissions...et de mettre

Utilisation de mots appartenant à une mauvaise partie du discours (substantif et verbe)

[10-1] un projet de production d'énergie indépendante

[10-3] un pouvoir indépendant projet

Omission de mot(préposition et nom : « de production »)

[10-1] La SFI s'est acquittée de

[10-3] La SFI accompli

Faute de conjugaison (choix de temps)

[10-1] a rapporté à l'État

[10-3] Ceará récolté...pour l'utilitaire

Omission de mot(verbe : auxiliaire)

[10-1] soit 27 % de plus que

[10-3] 27 pour cent au-dessus du

Omission de mot(verbe)

[10-1] la compagnie d'électricité de l'État
[10-3] le pouvoir de l'état utilitaire
Faute de ponctuation (majuscule à « État »)

[10-1] un projet de production d'énergie indépendante
[10-3] un pouvoir indépendant projet
Mauvaise ordre des mots (du nom et de l'adjectif)

[10-1] en seulement neuf mois
[10-3] dans un relativement rapide neuf mois
Mauvaise ordre des mots (du nom et de l'adjectif),
Utilisation de préposition erronée

[10-1] 880 millions de dollars
[10-3] \$880 millions
Faute de ponctuation (signe « \$ » pour « dollars »),
Mauvaise ordre des mots (du complément de nom (\$))

Erreurs lexicales :

[10-1] la compagnie d'électricité
[10-3] le pouvoir de l'état utilitaire
Anglicisme (« pouvoir » pour « *power* »)
Anglicisme (« utilitaire » pour « *utility* »)

[10-1] un projet de production d'énergie indépendante
[10-3] un pouvoir indépendant projet
Anglicisme (« pouvoir » pour « *power* »)

[10-1] La vente aux enchères de la compagnie d'électricité a rapporté à l'État 880 millions de dollars
[10-3] Dans le cadre d'une vente aux enchères, Ceará récolté 880 millions de dollars pour l'utilitaire
Anglicisme (« utilitaire » pour « *utility* »)

Erreurs sémantiques :

[10-1] d'ambitieux objectifs
[10-3] ses sites haute
Faux-sens (« sites » pour « *sights* »)

[10-1] la compagnie d'électricité
[10-3] le pouvoir de l'état utilitaire
Faux-sens (« état utilitaire »)

[10-4] Google Translate

Le gouverneur jeté son dévolu élevé: la privatisation de la puissance publique de l'état, de la structure et enchérir sur un projet énergétique indépendant, et mettre en place un organisme de réglementation. IFC accompli tout cela dans neuf relativement rapides mois. Aux enchères, Ceará a récolté 880 millions de dollars pour l'utilité, 27 pour cent au-dessus du prix minimum.

Erreurs grammaticales :

[10-1] Le gouverneur s'est fixé
[10-4] Le gouverneur jeté
Faute de conjugaison (choix de temps)
Omission de mot (verbe : auxiliaire)

[10-1] de structurer un projet de production d'énergie indépendante, de lancer un appel d'offres en vue de sa réalisation
[10-4] de la structure et enchérir sur un projet énergétique indépendant
Utilisation de mots appartenant à une mauvaise partie du discours (substantif et verbe)

[10-1] un projet de production d'énergie indépendante
[10-4] un projet énergétique indépendant
Omission de mot (préposition et nom : « de production »)

[10-1] La SFI s'est acquittée de
[10-4] IFC accompli
Faute de conjugaison (choix de temps)
Omission de mot (déterminant)

[10-1] soit 27 % de plus que
[10-4] 27 pour cent au-dessus du
Omission de mot (verbe)

[10-1] :
[10-4]:
Faute de ponctuation

[10-1] l'État
[10-4] l'état
Faute de ponctuation

[10-1] en...neuf mois
[10-4] dans neuf...mois
Utilisation de préposition erronée

[10-1] en seulement neuf mois
[10-4] dans neuf relativement rapides mois
Mauvaise ordre des mots (de l'adverbe et de l'adjectif)

Erreurs lexicales :

[10-1] La vente aux enchères de la compagnie d'électricité a rapporté à l'État 880 millions de dollars
[10-4] Ceará a récolté 880 millions de dollars pour l'utilité
Anglicisme (« utilité » pour « *utility* »)

Erreurs sémantiques :

[10-1] d'ambitieux objectifs
[10-4] son dévolu élevé
Faux-sens

[10-1] il a en effet décidé de privatiser la compagnie d'électricité
[10-4] la privatisation de la puissance publique
Faux-sens

[10-1] de structurer un projet de production d'énergie indépendante, de lancer un appel d'offres en vue de sa réalisation
[10-4] de la structure et enchérir sur un projet énergétique indépendant
Faux-sens (« enchérir » pour « lancer un appel d'offres »)

[10-5] Bing

Le gouverneur a jeté son dévolu haute : privatiser l'électricité d'État, la structure et la soumission d'un projet d'électricité indépendants et de mettre en place un organisme de réglementation. IFC accompli tout cela dans un neuf mois relativement rapide. Aux enchères, Ceará engrangé 880 millions \$ pour l'utilitaire, 27 pour cent supérieur au prix minimum.

Erreurs grammaticales :

[10-1] d'ambitieux objectifs

[10-5] son dévolu haute

Faute d'accord (nom-adjectif)

[10-1] privatiser la compagnie d'électricité de l'État

[10-5] privatiser l'électricité d'État

Omission de mot (nom)

Omission de mot (déterminant)

[10-1] de privatiser...de structurer...de lancer...de mettre

[10-5] privatiser...la structure et la soumission...de mettre

Utilisation de mots appartenant à une mauvaise partie du discours (substantif et verbe)

[10-1] un projet de production d'énergie indépendante

[10-5] un projet d'électricité indépendants

Faute d'accord (nom-adjectif)

Omission de mot (nom)

[10-1] La SFI s'est acquittée de

[10-5] IFC accompli

Faute de conjugaison

Omission de mot (déterminant)

[10-1] seulement neuf mois

[10-5] neuf mois relativement rapide

Faute d'accord (nom-adjectif)

[10-1] La vente...a rapporté à l'État
[10-5] Ceará engrangé
Omission de mot (verbe : auxiliaire)

[10-1] 880 millions de dollars
[10-5] 880 millions \$
Omission de mot (préposition « de »)

[10-1] soit 27 % de plus que le prix minimum demandé
[10-5] 27 pour cent supérieur au prix minimum
Omission de mot (verbe)

[10-1] en...neuf mois
[10-5] dans un neuf mois
Utilisation de préposition erronée,
Utilisation d'article erroné

[10-1] 880 millions de dollars
[10-5] 880 millions \$
Faute de ponctuation (signe « \$ » pour « dollars »),
Mauvaise ordre des mots (du complément de nom (\$))

Erreurs lexicales :

[10-1] La vente...a rapporté à l'État
[10-5] Ceará engrangé...pour l'utilitaire
Anglicisme : « utilitaire » pour « *utility* »

Erreurs sémantiques :

[10-1] d'ambitieux objectifs
[10-5] son dévolu haute
Faux-sens

[10-1] privatiser la compagnie d'électricité de l'État
[10-5] privatiser l'électricité d'État
Non-sens (« l'électricité de l'État » n'a rien à voir avec « la compagnie d'électricité de l'état »)

APPENDICE B Erreurs commises par type

		Prompt	Babylon	Google	Bing
1	Nombre d'erreurs	13	15	5	8
	Erreurs grammaticales	8	9	4	5
	Erreurs lexicales	1	1	0	1
	Erreurs sémantiques	4	5	1	2
	Divergences	12	14	11	12
2	Nombre d'erreurs	5	4	5	4
	Erreurs grammaticales	2	2	3	2
	Erreurs lexicales	1	1	1	1
	Erreurs sémantiques	2	1	1	1
	Divergences	12	11	9	10
3	Nombre d'erreurs	4	2	1	3
	Erreurs grammaticales	2	1	1	2
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	2	1	0	1
	Divergences	13	11	12	13
4	Nombre d'erreurs	3	2	3	2
	Erreurs grammaticales	1	1	1	0
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	2	1	2	2
	Divergences	8	8	8	8
5	Nombre d'erreurs	3	4	8	6
	Erreurs grammaticales	0	2	3	3
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	3	2	5	3
	Divergences	12	17	14	13
6	Nombre d'erreurs	3	1	1	2
	Erreurs grammaticales	1	1	0	1
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	2	0	1	1
	Divergences	12	10	11	12
7	Nombre d'erreurs	2	2	2	6
	Erreurs grammaticales	2	2	2	5

	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	0	0	0	1
	Divergences	12	8	9	13
8	Nombre d'erreurs	7	6	3	6
	Erreurs grammaticales	3	4	2	3
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	4	2	1	3
	Divergences	16	13	12	11
9	Nombre d'erreurs	8	10	7	8
	Erreurs grammaticales	2	7	4	5
	Erreurs lexicales	0	0	0	0
	Erreurs sémantiques	3	2	2	2
	Erreurs stylistiques	3	1	1	1
	Divergences	20	18	17	19
10	Nombre d'erreurs	17	19	14	19
	Erreurs grammaticales	11	13	10	16
	Erreurs lexicales	4	4	1	1
	Erreurs sémantiques	2	2	3	2
	Divergences	33	28	28	25

TOTAL :

	Promt	Babylon	Google	Bing	TOTAL
Nombre d'erreurs (total)	65	65	49	64	243
Erreurs grammaticales	32	42	30	42	146
Erreurs lexicales	6	6	2	3	17
Erreurs sémantiques	24	16	16	18	74
Erreurs stylistiques	3	1	1	1	6
Erreurs d'orthographe	0	0	0	0	0
Divergences	150	138	131	136	555
% Divergences (sur 405 mots)	37	34	32	34	

APPENDICE C Erreurs commises par sous-type

	1	2	3	4	5	6	7	8	9	10	TOTAL
PROMT											
Erreurs grammaticales	8	2	2	1	0	1	2	3	2	11	32
Faute d'accord	1	0	0	0	0	0	0	0	0	2	3
Mots appartenant à une mauvaise partie du discours	1	0	1	0	0	0	0	0	0	1	3
Faute de conjugaison	0	0	0	0	0	0	0	0	0	1	1
Omission de mots	1	1	1	0	0	1	0	2	0	2	8
Faute de ponctuation	2	1	0	1	0	0	1	0	1	1	7
Mauvaise ordre des mots	1	0	0	0	0	0	1	1	0	1	4
Utilisation de prépositions/articles erronés	2	0	0	0	0	0	0	0	1	3	6
Erreurs lexicales	1	1	0	0	0	0	0	0	0	4	6
Choix de termes inappropriés	1	0	0	0	0	0	0	0	0	0	1
Anglicisme	0	1	0	0	0	0	0	0	0	4	5
Erreurs sémantiques	4	2	2	2	3	2	0	4	3	2	24
Faux-sens	3	2	2	2	2	2	0	4	3	2	22
Imprécision	1	0	0	0	0	0	0	0	0	0	1
Ambiguïté	0	0	0	0	0	0	0	0	0	0	0
Ajout	0	0	0	0	1	0	0	0	0	0	1
Erreurs stylistiques	0	0	0	0	0	0	0	0	3	0	3
Construction boiteuse	0	0	0	0	0	0	0	0	1	0	1
Absence d'idiomaticité	0	0	0	0	0	0	0	0	2	0	2

BABYLON

Erreurs grammaticales	9	2	1	1	2	1	2	4	7	13	42
Faute d'accord	1	0	0	0	0	1	0	0	1	1	4
Mots appartenant à une mauvaise partie du discours	1	0	0	0	1	0	1	0	0	1	4
Faute de conjugaison	1	0	0	0	0	0	0	0	2	2	5
Omission de mots	1	1	0	0	1	0	1	2	1	3	10
Faute de ponctuation	1	1	0	1	0	0	0	1	1	2	7
Mauvaise ordre des mots	1	0	0	0	0	0	0	1	0	3	5
Utilisation de préposition/article erroné	3	0	1	0	0	0	0	0	2	1	7
Erreurs lexicales	1	1	0	0	0	0	0	0	0	4	6
Choix de termes inappropriés	1	0	0	0	0	0	0	0	0	0	1
Anglicisme	0	1	0	0	0	0	0	0	0	4	5
Erreurs sémantiques	5	1	1	1	2	0	0	2	2	2	16
Faux-sens	3	1	1	1	2	0	0	1	2	2	13
Non-sens	1	0	0	0	0	0	0	0	0	0	1
Imprécision	1	0	0	0	0	0	0	0	0	0	1
Ambiguïté	0	0	0	0	0	0	0	0	0	0	0
Ajout	0	0	0	0	0	0	0	1	0	0	1
Erreurs stylistiques	0	0	0	0	0	0	0	0	1	0	1
Construction boiteuse	0	0	0	0	0	0	0	0	1	0	1
Absence d'idiomaticité	0	0	0	0	0	0	0	0	0	0	0

GOOGLE TRANSLATE

Erreurs grammaticales	4	3	1	1	3	0	2	2	4	11	31
Faute d'accord	0	1	0	0	0	0	0	0	0	0	1
Mots appartenant à une mauvaise partie du discours	0	0	0	0	0	0	1	0	0	1	2
Faute de conjugaison	1	0	0	0	3	0	0	0	1	2	7
Omission de mots	1	0	1	0	0	0	1	2	0	4	9
Faute de ponctuation	0	1	0	1	0	0	0	0	1	2	5
Mauvaise ordre des mots	1	1	0	0	0	0	0	0	0	1	3
Utilisation de prépositions ou d'articles erronés	1	0	0	0	0	0	0	0	2	1	4
Erreurs lexicales	0	1	0	0	0	0	0	0	0	1	2
Choix de termes inappropriés	0	0	0	0	0	0	0	0	0	0	0
Anglicisme	0	1	0	0	0	0	0	0	0	1	2
Erreurs sémantiques	1	1	0	2	5	1	0	1	2	3	16
Faux-sens	1	1	0	2	2	1	0	1	2	3	13
Imprécision	0	0	0	0	0	0	0	0	0	0	0
Ambiguïté	0	0	0	0	0	0	0	0	0	0	0
Ajout	0	0	0	0	3	0	0	0	0	0	3
Erreurs stylistiques	0	0	0	0	0	0	0	0	1	0	1
Construction boiteuse	0	0	0	0	0	0	0	0	1	0	1
Absence d'idiomaticité	0	0	0	0	0	0	0	0	0	0	0

BING

Erreurs grammaticales	5	2	2	0	3	1	5	3	5	16	42
Faute d'accord	0	0	0	0	0	0	0	0	1	3	4
Mots appartenant à une mauvaise partie du discours	1	0	1	0	0	0	0	0	0	1	3
Faute de conjugaison	1	0	0	0	2	0	2	1	1	1	8
Omission de mots	0	2	1	0	0	1	0	1	0	7	12
Faute de ponctuation	0	0	0	0	0	0	1	1	1	1	4
Mauvaise ordre des mots	2	0	0	0	1	0	2	0	1	1	7
Utilisation de prépositions ou d'articles erronés	1	0	0	0	0	0	0	0	1	2	4
Erreurs lexicales	1	1	0	0	0	0	0	0	0	1	3
Choix de termes inappropriés	0	0	0	0	0	0	0	0	0	0	0
Anglicisme	1	1	0	0	0	0	0	0	0	1	3
Erreurs sémantiques	2	1	1	2	3	1	1	3	2	2	18
Faux-sens	2	1	1	2	1	1	1	2	2	1	14
Non-sens	0	0	0	0	0	0	0	0	0	1	1
Imprécision	0	0	0	0	0	0	0	0	0	0	0
Ambiguïté	0	0	0	0	0	0	0	1	0	0	1
Ajout	0	0	0	0	2	0	0	0	0	0	2
Erreurs stylistiques	0	0	0	0	0	0	0	0	1	0	1
Construction boiteuse	0	0	0	0	0	0	0	0	1	0	1
Absence d'idiomaticité	0	0	0	0	0	0	0	0	0	0	0

ERREURS COMMISES GLOBALEMENT

	14
Erreurs grammaticales	7
Faute d'accord	12
Mots appartenant à une mauvaise partie du discours	12
Faute de conjugaison	21
Omission de mots	39
Faute de ponctuation	23
Mauvaise ordre des mots	19
Mots appartenant à une mauvaise partie du discours	12
Utilisation de prépositions ou d'articles erronés	21
Erreurs sémantiques	74
Faux-sens	62
Non-sens	2
Imprécision	2
Ambiguïté	1
Ajout	7
Erreurs lexicales	17
Choix de termes inappropriés	2
Anglicisme	15
Erreurs stylistiques	6
Construction boiteuse	4
Absence d'idiomaticité	2