LINGUISTIQUE DE CORPUS, TAL ET TRADUCTION AUTOMATIQUE

Nabila BEDJAOUI
Laboratoire SEPRADIS
Université Mohamed Khider, Biskra, Algérie
n.bedjaoui@univ-biskra.dz

Résumé : La linguistique de corpus est directement liée au TAL (traitement automatique du langage). Dans le présent article, nous allons dans un premier temps, nous intéresser aux fondements de la linguistique de corpus en tant que discipline, et dans un second temps, nous nous intéresserons à la traduction automatique comme outil de TAL. Nous prendrons comme exemple Google Traduction (GT). Nous nous pencherons sur ses atouts et ses limites, aussi exploiterons-nous les données d'une enquête sur GT afin de mettre en exergue les subtilités du processus de traduction. Un questionnaire a été distribué à un échantillon de 21 enseignants universitaires algériens ayant déjà publié un article. Le corpus récolté a fait l'objet d'une double analyse: d'abord quantitative, ensuite qualitative. Un deuxième corpus constitué de phrases traduites de l'anglais au français, prises de 10 résumés d'articles publiés en ligne, a été analysé pour mettre à l'épreuve GT. Cette étude se fixe comme objectif de mettre en lumière la linguistique de corpus en tant que discipline empirique, ainsi que d'explorer le domaine du TAL pour aboutir à l'étude d'un cas, celui de la traduction automatique à travers Google Traduction.

Mots-clés : Linguistique de corpus, Corpus, TAL, Traduction, Google Traduction

CORPUS LINGUISTICS, NLP AND AUTOMATIC TRANSLATION

Abstract: Corpus linguistics is directly related to NLP (natural language processing), in this research paper, we will first focus on the foundations of corpus linguistics as a discipline, and then we will look at machine translation as a NLP tool. We will take Google Translate (GT) as an example. We will look at its strengths and limitations, so we will use data from a GT survey to highlight the intricacies of the translation process. A questionnaire was distributed to a sample of 21 Algerian university teachers who had already published a paper. The collected corpus was the subject of a double analysis: first quantitative, then qualitative. A second corpus consisting of sentences translated from English to French, taken from 10 abstracts of articles published online, was analyzed to test GT. This study therefore sets itself the objective of highlighting corpus linguistics as an empirical discipline, as well as exploring the field of NLP to end up with the study of a case, that of automatic translation through Google. Translation (GT).

Keywords: Corpus linguistics, Corpus, NLP, Translation, Google Translate

Introduction

Tel M. Jourdain nous avons fait de la linguistique de corpus sans le savoir, et sans avoir connaissance de son existence, idée empruntée à Geneviève Calame-Griaule (Calame-Griaule, 1981), qui dit avoir fait de l'ethnolinguistique d'abord sans le savoir. En effet, nous n'avons pris connaissance de l'existence de la linguistique de corpus comme discipline que quand nous avons été sollicitée pour l'enseigner en tant que

module pour les étudiants de Master 1. Et ce fut une découverte à plus d'un égard, qui a suscité chez nous une série d'interrogations autant sur le plan théorique que sur le plan pratique, d'abord sur les fondements de cette discipline, ses origines, ses pionniers, ses principes, ensuite sur sa conception des corpus et leur traitement dont le traitement automatique. De ce fait notre étude traite en amont de la linguistique du corpus et du TAL et s'articule ainsi autour des questions suivantes: peut-on ne considérer la linguistique de corpus qu'à l'aune du TAL ? Qu'en est-il du traitement manuel du langage ? Quelles sont les spécificités des corpus traités automatiquement ? Et en aval, notre étude traite de la traduction automatique étant un des outils du traitement automatique du langage (TAL), et suscite des interrogations quant aux performances qu'elle propose et aux limites qu'elle impose.

Pour rendre compte de ces interrogations, nous exploiterons un des outils de traduction les plus utilisés au monde, il s'agit de Google traduction (GT), l'objectif étant de mettre en exergue les subtilités du processus de traduction automatique. Pour ce faire nous avons construit deux corpus, le premier est constitué des réponses à un questionnaire distribué aux enseignants universitaires algériens ayant déjà publié un article. La finalité de ce questionnaire est double, dans un premier temps il s'agit de repérer les problèmes rencontrés lors de l'opération de traduction sur Google, et dans un second temps, de connaitre les démarches des enseignants pour les résoudre. Le deuxième corpus sera constitué des écarts linguistiques prélevés dans des résumés d'articles publiés sur la plateforme ASJP[1]. Dans la partie *infra* de notre article, nous proposons une étude épistémologique traitant des concepts abordés dans notre travail.

1. Linguistique de corpus et corpus

Dès son apparition dans les années 60 en Angleterre, la linguistique de corpus a suscité un intérêt multidimensionnel chez les chercheurs. Cette discipline alors naissante a fait l'objet d'une série de recherches, d'ouvrages, de séminaires :

> La linguistique de corpus est un paradigme qui connaît depuis les vingt dernières années un succès croissant (Williams 2006). Cependant, sous une même appellation, de nombreuses pratiques coexistent plus ou moins pacifiquement, faisant courir aux corpus le risque d'être galvaudés.
>
> Giuliani & Hannachi, (2010 : 145)

Cette difficulté épistémologique rend difficile la tentative de cerner la discipline et de la définir d'une manière claire et objective.

> En effet, la grande diversité des pratiques et des méthodes que l'on peut regrouper derrière l'objet « corpus » interdit l'économie d'une définition précise des positionnements, méthodes, objectifs, et outils employés par le chercheur avant toute démarche de recherche.
>
> Giuliani & Hannachi, (2010 : 146)

Par conséquent la discipline devient l'apanage d'une vaste tranche de chercheurs qui aspirent à une certaine spécificité de leurs travaux de recherche. Bien que le mot corpus soit annexé à la linguistique, il ne s'agit guère de cet aspect dénué de toute dimension réelle, voire sociale. Tout au contraire, la linguistique de corpus se doit de traiter des corpus attestés, contextualisés. « La linguistique de corpus peut ainsi

[1] Plateforme d'édition électronique des revues scientifiques algériennes développée et gérée par le Centre de Recherche sur l'Information Scientifique et Technique (CERIST)

être définie comme une discipline empirique par excellence, car elle vise à tirer des conclusions sur la base de l'analyse de données externes plutôt que sur les connaissances linguistiques propres aux chercheurs. » (Zuffery, 2020 : 4). Il est donc question d'exploiter, dans le but d'une recherche linguistique, des données de différentes natures : textes, audio, vidéos… « La plupart du temps, ces échantillons sont rassemblés dans un format informatisé, ce qui permet de les étudier plus efficacement que s'ils étaient sur papier. » (Zuffery, 2020 : 4). Nous devons souligner à ce niveau que la linguistique de corpus est généralement connue et reconnue par rapport au lien qu'elle entretient avec le TAL. L'exploitation de vastes données dans un format numérisé donne souvent sa raison d'être à la linguistique de corpus. Ceci dit, l'informatique n'est pas une condition *sine qua non* pour l'analyse des corpus et la linguistique de corpus. Il est toujours possible de traiter manuellement un corpus, certes la quantité est limitée par rapport au TAL, mais la dimension ainsi que l'intuition humaines ne peuvent être remplacées. Lors d'un colloque[2] sur la linguistique de corpus, nous avons eu l'occasion de poser des questions à Philipe Blanchet, notamment sur le degré de pertinence du TAL, il a répondu que malgré les avancées technologiques, l'intervention humaine est inévitable. La machine a ses limites que seul l'homme est capable de pallier. La linguistique de corpus est souvent renvoyée à la linguistique appliquée. De ce fait elle recouvre tous les aspects linguistiques de la langue. Elle peut ainsi être appliquée en lexicographie, syntaxe, phonologie, sociolinguistique … Nous estimons d'ailleurs et en ce qui concerne la sociolinguistique qu'il faudrait évoquer une sociolinguistique de corpus, tant les corpus sont puisés dans les interactions sociales et tant il est question de langues et de variations.

> La linguistique de corpus est une méthode particulièrement efficace pour établir quels sont les contextes fréquents d'utilisation d'un mot ou d'une expression ? Dans l'étude de la syntaxe, la linguistique de corpus permet de recenser les types de structures syntaxiques usités dans différentes langues. Par exemple, par une étude de corpus, il est possible de déterminer dans quels genres textuels les phrases passives sont les plus utilisées. Enfin, grâce à l'existence de corpus de données orales, la linguistique de corpus permet également de répondre à des questions de phonologie et de sociolinguistique, par exemple d'établir la zone de diffusion géographique de certains traits de prononciation.
>
> Zuffery (2020 :5)

Le corpus est la clé de voûte de tout travail de recherche scientifique. Sa présence dissipe le doute et éclaire la lanterne du chercheur lui permettant ainsi de répondre à sa problématique et de vérifier ses hypothèses. D'où l'importance cruciale de cet élément de base, ce *pivot* autour duquel gravitent les autres éléments de la recherche et sans lequel tout s'écroule. La linguistique de corpus, elle, tend à étudier les différents corpus de manières diverses et avec des outils entre autres technologiques (numériques) dont le TAL. « Le mot *corpus* vient du latin et signifie le corps. Ainsi, un *corpus de textes* constitue littéralement un corps de textes, c'est-à-dire un rassemblement d'un certain nombre de textes dans le but de les étudier. » (Zuffery (2020 :3) Il a été défini par Saint clair (1996) comme étant « une collection de données

[2] « Choix de corpus et de méthodes : Contextualiser sa recherche en lettres et langues étrangères » tenu le 1er et le 2 juin 2022 à l'université de Batna2.http://fr.univbatna2.dz/sites/default/files/fra/files/programme_colloque_corpus_01.pdf

langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon de langage » (Sinclair, 1996, p. 4) les auteurs du *Dictionnaire de praxématique,* quant à eux nous décrivent le contexte de construction du corpus ainsi que son sens,

> Afin d'étudier un phénomène, quel qu'il soit, il convient de recueillir un ensemble d'énoncés manifestant ledit phénomène, pour pouvoir l'analyser dans son objectivité. Le corpus peut être naturel (ou attesté) : les occurrences sont prélevées dans des interactions sociales ; il peut être plus ou moins artefactuel : sa production est provoquée par le chercheur qui fait réaliser les occurrences à une population cible…ou les produit lui-même en faisant appel et confiance à son intuition de sujet parlant. En analyse du discours, le corpus est forcément attesté. L'utilisation de l'informatique permet actuellement le traitement de vastes unités, qui fera découvrir, on peut le penser, de nouveaux objets.
>
> Détrie, Siblot, & Vérine (2001 :68)

Les corpus représentent des échantillons linguistiques de nature très variée et cette variété permet de répondre à des questions de recherches très diverses dans tous les domaines de la (socio) linguistique. D'où l'existence de différents types de corpus. « Corpus échantillons et corpus évolutifs, corpus de langue générale et corpus de spécialité, corpus brut et corpus annoté, corpus monolingue et corpus multilingue (…) » (Zuffery, 2020, p. 18) En plus des corpus synchroniques et des corpus diachroniques. (*Ibid.*)

2. TAL

Le traitement automatique du langage ou TAL est désormais un domaine en plein essor qui prend en charge statistiquement des corpus volumineux pouvant atteindre des milliers, voire des millions de données. À la question qu'est-ce que le TAL ? Kurdi (2017 : 10) répond : « Il s'agit d'une discipline qui se trouve à l'intersection de plusieurs autres branches de la science comme l'informatique, l'intelligence artificielle, la linguistique et la psychologie cognitive. ». L'auteur donne des exemples d'outils de TAL (logiciels), il cite les étiqueteurs de discours, les analyseurs morphologiques, et les analyseurs syntaxiques. Nous pouvons de même citer les concordanciers qui,

> […] permettent notamment de rechercher toutes les occurrences d'un mot avec leur contexte d'utilisation et de faire apparaitre les résultats ligne par ligne en une seule requête. Ces outils permettent également d'établir la liste des mots contenus dans le corpus avec leur fréquence et de générer une liste de mots-clés correspondant au contenu d'un corpus. Dans le cas des corpus contenant des textes et leur traduction, certains outils appelés des aligneurs permettent d'aligner phrase par phrase le contenu du corpus.
>
> Zuffery (2020 :11)

Un autre exemple de TAL, l'ACM, (Analyse des Correspondances multiples) qui est une méthode statistique, descriptive, employée pour explorer un grand nombre de données dans un format numérisé. Ces méthodes ont commencé à se développer dès les années 1950. Elle est souvent intégrée à des méthodes en relation avec l'informatique et l'intelligence artificielle, tel est le cas du *Data mining*, « fouilles de données » en français. Quand les données se présentent sous forme de texte, comme

c'est le cas pour les questions ouvertes, il s'agit du *Text Mining* (Lebart, 1999). Cette méthode est utilisée pour décrire la corrélation entre des individus et des variables, elle permet de dresser des représentations graphiques qui mettent en relief cette corrélation. « Dans une ACM, la distance entre les individus est calculée pour mettre en évidence les relations entre eux. Plus cette distance est grande, plus les individus sont différents et ne présentent pas les mêmes modalités. L'ACM est un outil qui permet de vérifier certains aspects des hypothèses formulées, qui ne peuvent pas être abordés par l'observation directe sur terrain. » (Bedjaoui, 2016 : 164).

3.Traduction automatique

D'après Larousse, *traduire* vient du latin *traducere,* faire passer d'un point à un autre, et c'est « Transposer un discours, un texte, de l'exprimer dans une langue différente. », transposer dans le sens de permuter, de replacer. La traduction a côtoyé la diversité linguistique tout au long de l'histoire. Traduit-on les mots ou les idées ? Rappelons, à ce niveau, l'injonction de Cicéron selon laquelle il ne faut pas traduire mot à mot, « verbum pro verbo » (Oustinoff, 2015). Dans le même contexte nous pouvons citer l'Ut Pictura d'Horace où il condamne la peinture et la littérature à n'être qu'une interprétation pure et simple l'une de de l'autre. La peinture et la littérature sont ainsi en situation de conformité ou d'équivalence. Rensselaer, cité par (Ambrasian, 2001, p. 304) estime que « La poésie est une peinture parlante et la peinture une poésie muette. ». le peintre traduit le poème à traves son pinceau et le poète traduit la peinture à travers sa plume. Que traduit-on ? Comment traduit-on ? Et pourquoi traduit-on ? Tout dépend du contexte, des besoins ainsi que des objectifs de l'opération de traduction. Pour comprendre la subtilité du processus de traduction, nous pouvons lire les textes sous leur forme originelle (langue source) et sous leur forme seconde (langue cible), nous pensons à nos expériences personnelles, lire Ahlem Mostghanemi en arabe, Shkespeare en anglais, ou De Musset en français diffère de les lire en d'autres langues. Il y a ce petit *quelque chose* qu'on ne peut traduire, *l'âme de l'œuvre*.

Venons-en maintenant à la traduction automatique, qui diffère de la traduction *classique* par son recours à l'outil informatique. Les premières machines à traduire automatiquement les langues naturelles remontent à la seconde guerre mondiale où des ordinateurs ont été utilisés par des savants américains pour décoder des messages de l'armée japonaise. Technique également exploitée lors de la guerre froide entre les USA et l'URSS. (Raemdonck & Siouffi, 2012). D'après les mêmes auteurs la technique de traduction la plus rudimentaire consiste à introduire dans la mémoire de l'ordinateur deux répertoires de mots le premier appartenant à la langue « source » et le deuxième à la langue « cible », par la suite des systèmes de correspondances sont installés entre eux. » (*Ibid.*).

« La TA consiste à saisir un texte, à le soumettre à un traitement automatique et à récupérer en sortie « une traduction brute ». Aucune intervention humaine n'est possible pendant le traitement » (Loffler-Lorian, 1996 : 7). Néanmoins des problèmes persistent et entravent le processus de la traduction à plus d'un niveau.

> According to its original creator Franz-Josef Orch (2005), Google Translate does not adopt a grammar rule analysis in their translation method, but a statistical machine translation method. This means instead of teaching the machine basic grammatical rules such as word order and sentence structure of the target language, the translation results are based on the frequency of word collocations that is sequences of words that appear together in the sentence."
>
> Lee (2016: 106)

Chose que nous essayerons d'expliciter à travers notre enquête qui sera consacrée à l'étude d'un outil de la TA, il s'agit de GT (Google Traduction). Qui est donc derrière GT ? Franz-Josef Och actuellement responsable de GT, qui a fait plusieurs recherches et a publié plusieurs articles sur la traduction automatique. Tout a commencé en 2006 avec *statistical machine translations* de l'arabe, du chinois et du russe de et vers l'anglais. Franz-Josef Och, explique comment: « Our approach was to use statistical translation models learned from parallel text, that is, sets of documents and their translations. The system learns a model automatically from the parallel data. » (Freif, 2010 : 95).

4. Enquête

Notre enquête présente deux pans, le premier est consacré à un questionnaire sur la TA, en général et sur GT en particulier. Il est constitué de 5 questions, la première est semi-directive, les autres questions sont ouvertes. Le questionnaire est distribué à un échantillon de 21 enseignants appartenant à différentes universités algériennes, (Biskra, Constantine, Jijel, Sétif, Annaba, Tizi Ouzou et Mostaganem). Nous soumettrons notre corpus à une double analyse, d'abord quantitative, ensuite qualitative. Le deuxième pan de notre enquête concerne une étude effectuée sur des résumés d'articles publiés sur la ASJP. Le corpus sera constitué de 10 résumés en anglais que nous allons traduire en français via Google Traduction, pour vérifier si : l'auteur a utilisé cet outil ; l'outil en question a bien effectué l'opération de traduction ; l'auteur est intervenu pour rectifier la traduction.

5. Résultats

5.1 Le questionnaire

Dans un premier temps nous exposerons et analyserons les résultats obtenus à travers le questionnaire, et dans un second temps nous traiterons les résumés.

Question 1

Pour traduire le résumé de votre article en anglais, (cochez la réponse choisie)
a- vous le traduisez, vous-même?
b- vous utilisez Google Traduction?
c- vous utilisez un autre logiciel, si oui, lequel?
d- vous demandez à une personne spécialisée en anglais de le traduire pour vous?
L'objectif à travers cette question est de connaitre, même indirectement, la capacité des enseignants universitaires à traduire du français à l'anglais. Préciser leur relation à l'anglais, à la traduction.
La majorité des enseignants, soit 9 d'entre eux ont choisi la réponse *d.* ils affirment donc que pour traduire le résumé de leur article en anglais, ils demandent à une personne spécialisée en anglais de le traduire pour eux. Ces enseignants ne font pas recours à Google traduction Qu'en est-il des traducteurs sollicités ?
7 enseignants utilisent GT. 3 enseignantes traduisent elles-mêmes leurs articles, elles précisent néanmoins qu'elles demandent à un anglophone de vérifier leurs résumés. 1 enseignant utilise un autre logiciel, « J'utilise le traducteur de Word, il est bien meilleur que Google traduction ». Une enseignante n'a pas répondu à la question.

Question 2

Quand vous utilisez Google Traduction, est-ce que vous êtes satisfait à 100%.
Oui Non ☐ ☐
Pourquoi? (Si c'est non)

Les enseignants sont presque unanimes par rapport aux réponses à cette question, ils ont tous répondu *non.* Ils soulignent ainsi leur insatisfaction de l'outil GT. Leurs raisons sont les suivantes (E=Enseignant) : E4 : « Les termes pertinents ne sont pas traduits mot à mot dans l'autre langue. », E5 : « Google Traduction fournit des traductions littérales. Cela ne suffit pas pour des traductions professionnelles et aussi pour les traductions qui requièrent un haut niveau de précision. », E6 : « La traduction est de très mauvaise qualité, c'est une traduction automatique qui ne tient pas compte de beaucoup de paramètres. », E7 : « Il y a des nuances au niveau du sens de certains mots et expressions. De même, il ne constitue pas logiquement et grammaticalement la structure phrastique d'une langue à une autre. Il reste un logiciel automatique Si j'ose écrire, c'est une machine automatique "*bête*" mais "*disciplinée*". », E8 : « Il faut toujours relire le texte traduit pour voir sa conformité avec le texte original surtout lorsqu'il s'agit de concepts de spécialité. », E9 : « Je ne suis pas satisfaite dans la mesure où la traduction n'est pas faite de manière parfaite et exacte. », E10 : « Parce que la traduction est littérale. », E12 : « La traduction est souvent erronée car il ne prend en compte que le premier sens des mots qui n'est toujours pas le bon. En plus, les règles morphosyntaxiques ne sont pas toujours respectées. » Les réponses soulignent le manque de pertinence, le manque de précision et la traduction littérale.
2 enseignants insatisfaits n'ont pas donné de raisons, ils disent ne plus l'utiliser. 1 enseignante a répondu *oui.* 1 enseignante n'a pas répondu à la question.

Question 3

Quelles sont selon vous, les erreurs commises par l'outil Google Traduction?
*lexicales *morphosyntaxiques * sémantiques * pragmatiques autres*
Donnez des exemples.

L'ensemble de nos informateurs ont mentionné tous les types d'erreurs proposées dans le questionnaire. E1 a ajouté le manque de cohérence et de cohésion. E2 précise que GT « ne donne jamais le sens souhaité. », pour E4 : « les équivalents donnés dans la langue cible sont moins précis que ceux utilisés dans la langue source, (sens et registre non adéquat). », pour E6 : « Google traduction traduit systématiquement, ne peut avoir la subtilité de choisir les mots en fonction du contexte de leur insertion, ce sont des erreurs d'ordre sémantique, lexical et pragmatique. », E7 souligne le problème de nuances dans le sens. E13 : « Je n'ai jamais essayé de chercher les erreurs faites par Google Traduction, je prends la phrase traduite telle qu'elle est écrite. »

Question 4

Traduire est-il un problème pour vous?

Pour 16 enseignants, la traduction pose problème. 5 enseignants pensent le contraire.

Question 5

Que proposez-vous comme solution?

13 enseignants proposent d'apprendre la langue étrangère (la langue cible) l'anglais dans le cas des enseignants de français qui doivent traduire les résumés de leurs articles en anglais. Nous devons avouer que c'est en effet la meilleure solution, étant nous- même diplômée d'une licence en langue anglaise de 4 ans, (système classique).

Le fait de maitriser la langue cible, permet pendant le processus de traduction de vite repérer les anomalies quelles que soit leurs natures et de les corriger.
8 enseignants privilégient la formation de traducteurs spécialisés dans le domaine des articles scientifiques et qui puissent être disponibles de préférence en ligne.
Sachant que cette catégorie de traducteurs existe bel et bien, ils proposent leurs services payants en ligne, dans les groupes de doctorants et dans ceux des enseignants universitaires.
Mais est-ce la solution ?
Peut-être faudrait-il essayer de mieux perfectionner GT ainsi que les autres outils de traduction, tout en suivant une formation d'anglais.

5.2 Les résumés

Comme nous l'avons mentionné supra, le deuxième pan de notre enquête est constitué de l'analyse de 10 résumés pris d'articles en français, publiés sur la plateforme ASJP. Pour souligner les erreurs de traduction, nous allons soumettre les résumés en français à une traduction automatique vers l'anglais via GT, nous allons ensuite comparer le résumé obtenu avec le résumé publié. Nous avons opté pour des articles de français dans les trois spécialités, sciences du langage, littérature et didactique. Dans un premier temps nous allons procéder au repérage des écarts et dans un second temps, nous essayerons de définir la nature de l'écart selon les critères mentionnés dans le questionnaire proposé aux enseignants universitaires. L'écart en question peut donc être de nature : lexicale, morphosyntaxique, sémantiques, pragmatiques ou autre (à préciser), tout en proposant l'exemple, objet de l'écart.

Résumé N°1

« (…) obéit à des règles qui lui sont propres et à des ***canons*** spécifiques pour atteindre un objectif précis : faire disparaître la malédiction. »

« (…) obeys its own rules and specific ***guns*** to achieve a specific objective: to remove the curse. »

Commentaire

Dans cet exemple, le mot *canons* a été littéralement traduit (selon son premier sens) *armes,* or le sens exprimé par l'auteur est *norme.* GT, ne fait pas la part des choses quand il s'agit du contexte.
L'auteur est intervenu en modifiant le mot *guns* par *canons*.

« (…) obeys its own rules and specific ***canons*** to achieve a specific objective: to remove the curse. »

Résumé N°2

« (…) et nous comparons les résultats à ceux des élèves d'une classe ***au fonctionnement traditionnel*** »

« (…) and we compare the results to those of students in a ***class to traditional operation***? »

Commentaire

Nous pouvons remarquer que GT n'a pas trouvé la bonne traduction et que l'auteur a dû rectifier l'erreur. Cependant, l'auteur n'a pas remarquer que le mot ***élèves*** a été mal traduit ***students*** *qui veut dire étudiants,* même dans la version publiée.
L'auteur a ensuite rectifié de la sorte :

« (…) and compare the results with those of students in a ***traditionally operated classroom»***

Résumé N°3

« (…) l'existence d'un ***effet*** de médiation de l'impact pro-social perçu du travail. »
« (…) existence of a mediation of the perceived pro-social impact of work. »

Commentaire

GT n'a tout simplement pas trouver l'équivalent pour le mot *effet*.
Erreur que l'auteur a rectifiée : « (…) the existence of a mediating ***effect*** the perceived pro-social impact of work. »

Résumé N°4

« (…) ***et ce qu'endurent les*** parents d'enfants autistes. »
« (…) ***and this endured by*** parents of autistic children. »

GT a littéralement traduit les mots sans tenir compte de leur aspect grammatical. En effet GT a confondu ***ce que*** (pronom relatif COD), avec ***ce*** (adjectif démonstratif). L'auteur a rectifié : « (…) ***and what*** the parents of autistic kids ***endure***.»

Résumé N°5

« A ce sujet, il est question ***de mettre en lumière***. »
« In this regard, it is a question ***of highlight***. »

Commentaire

Of , est une préposition suivie d'un adjective se terminant par ***ing.*** GT n'a pas réussi à surpasser cette ambiguïté grammaticale corrigée par l'auteur dans la version publiée : « In this regard, it is a question ***of highlighting***. »

Résumé N°6

« ***L***'analyse des résultats obtenus indiquent ***une faible orientation concurrentielle*** dans le marché. »
« Analysis of the results obtained indicate ***low orientation competitive*** in the market. »

Commentaire

Dans un premier temps, GT n'a pas mis l'article défini ***L(a), the***, en anglais devant le mot *analyse*, la même erreur est reproduite avec l'article indéfini ***une*** qui n'a pas été traduit devant le mot faible. Dans un second temps, GT n'a pas bien placé l'adjectif ***competitive***, qui se place avant le nom en anglais, et qui a été placé après par GT. L'auteur a certes corrigé la traduction, néanmoins une erreur lui a échappé dans l'exemple cité : ***The*** avant *analysis.* Aussi a-t-il changé le mot *low* qui a le sens de *bas,* et l'a remplacé par *weak* qui a le sens de *faible* :
«Analysis of the results obtained indicate **a weak competitive orientation** in the market. »

Résumé N°7

« Dans quelle mesure ***la formation hybride*** peut-elle favoriser l'apprentissage à l'université ? Et comment elle permet de **développer chez nos étudiants l'autonomie individuelle** ? »
« To what extent does the ***training hybrid*** can promote learning at University? And how it allows develop ***autonomy in our students individual***? »

Commentaire

Même remarque que pour l'exemple précèdent, l'adjectif ***hybrid*** est placé après le nom. Dans le deuxième emplacement, GT a mis les mots en désordre affectant ainsi le sens de la phrase, ***autonomy in our students individual***, entrainant ainsi un triple écart tant au niveau lexical en affectant le sens de l'ensemble de la phrase, qu'au niveau grammatical en désorganisant la structure de la phrase et au niveau pragmatique où le contexte n'est nullement pris en charge. L'auteur a rectifié ainsi ses phrases :

«To what extent can ***hybrid training*** promote learning at university? And how does it help ***our students develop individual autonomy?*** »

Résumé N°8

« (...) sont des ***outils linguistiques*** de liaison des idées, (...) »

« (...) are ***tools linguistics*** of linking ideas, (...) »

Commentaire

Dans ce résumé, la version traduite par GT contient la même erreur dans toutes les phrases similaires. Erreur signalée dans le résumé N°7, et corrigé par l'auteur comme suit :

« (...) are ***linguistic tools*** for linking ideas, (...) »

Résumé N°9

« Les résultats de cette recherche, ***montre*** (...) »

«The results of this research, **shows** (...) »

Cette traduction est intéressante dans la mesure où elle souligne l'incapacité de GT à déceler les éventuelles erreurs notamment celles de conjugaison, en effet l'auteur a mal conjugué le verbe montrer *montre* au lieu de *montrent,* GT a traduit le verbe au singulier *shows* au lieu de *show,* erreur que l'auteur a corrigé dans la version anglaise :

«The results of this research ***show*** (...) »

Résumé N°10

« L'objectif de cet article est de voir comment la subjectivité de l'énonciateur ***transparaît*** dans son ***discours de presse écrite algérienne*** à travers les déictiques? »

«The objective of this article is to see how the subjectivity of the enunciator ***reflected*** in his ***written press speech Algerian*** through deictics? »

Commentaire

Écart grammatical, erreur de conjugaison, GT a conjugué *transparaît* (présent de l'indicatif), au (*simple past*) *reflected,* erreur que l'auteur a « cru » corriger en ajoutant *is* au participe passé transformant ainsi le verbe à la forme passive, forme inexistante dans la version en français. Autre erreur commise, cette fois par l'auteur qui a validé la traduction de GT du verbe *paraitre*, (*to reflect)*, qui a plutôt le sens de *refléter* et ne pas celui de *paraitre* qui a le sens de *appear.*

Dans la même phrase nous remarquons une autre erreur déjà rencontrée dans les résumés précédents, GT a éventuellement un problème quant à la place de l'adjectif par rapport au nom en anglais, erreur corrigé correctement par l'auteur :

The objective of this article is to see how the subjectivity of the enunciator ***is reflected*** in his speech in the ***Algerian written press*** through the deictics ? »

Une troisième erreur s'offre à nous dans le même exemple, celle du point d'interrogation qui clôture la phrase qui n'est nullement interrogative car aucune question n'y est posée. L'interrogation est en fait indirecte et ne nécessite donc pas le point d'interrogation. Forme que GT n'a su prendre en charge, vu la subtilité de la tournure pour une machine, qu'en est-il de l'auteur ?

6. Discussion

Il n'était nullement évident pour nous de rechercher des écarts de traduction automatique via GT. En effet, nous avons parcouru un bon nombre de résumés avec zéro écart. La question qui s'est imposée de suite était : est-ce qu'il s'agit d'une traduction automatique ou manuelle ? Nous avons fait l'expérience de retraduire un ou deux résumés et le résultat obtenu était fidèle à la version publiée, mais dans les deux cas, il s'agissait de petits résumés dont la langue se rapprochait plus vers un lexique simple et ordinaire, de même les phrases utilisées par les auteurs étaient plutôt

de structure simple. GT est certes l'outil de traduction le plus utilisé au monde, mais force est de constater que des lacunes persistent notamment au niveau syntaxique et sémantique. « L'exercice de ce métier, comme celui de tous les métiers, se heurte à des difficultés concrètes et pratiques (...) et qui tiennent essentiellement à ce qu'on pourrait appeler les différences des langues et des styles. » (Testard, 1985, p. 12).

Conclusion

Nous aspirions à travers notre étude à *informer* à propos de la linguistique de corpus, beaucoup de chercheur la pratique à leur insu, sans savoir qu'elle existe en tant que telle. La connaitre permet d'avancer sur le terrain en connaissance de cause, cerner son (ses) objet (s), maitriser ses méthodes, construire et analyser ses corpus et enfin la reconnaitre en tant que discipline à part entière. Le TAL serait-il devenu la condition *sine qua non* aux pratiques de la linguistique de corpus ? Nonobstant son utilisation presque systématique en linguistique de corpus, il n'en demeure pas moins un outil conditionné par le degré de développement technologique notamment par rapport à l'exploitation de l'intelligence artificielle par les tenants du domaine. Google Traduction en est un exemple, malgré le nombre d'utilisateurs qui s'élève en millions, et en dépit des améliorations apportées à cet outil, des lacunes persistent à tous les niveaux si l'on se réfère aux résultats obtenus dans notre recherche. Ladite recherche a permis de connaitre les avis des enseignants universitaires utilisateurs de GT dans leurs articles, précisément dans leurs résumés, quand il s'agit de les traduire en anglais. Les résultats obtenus à travers le questionnaire nous ont éclairée sur différents points soulevés dans les questions posées, notamment sur l'utilisation ou non de GT et sur le taux de satisfaction de cet outil. D'après les résultats obtenus, moins de la moitié de notre échantillon utilise GT, les autres enseignants demande à des spécialistes en anglais de le faire pour eux. Seulement rien ne dit que ces spécialistes n'utilisent pas GT d'autant plus que dans la totalité des résumés choisis dans la deuxième partie de notre enquête, nous avons trouvé une similitude quasi complète entre les résumés publiés et ceux traduits, hormis bien sur quelques différences que nous avons décelées. Quant à la satisfaction de GT, elle est quasi absente, ou presque chez les utilisateurs, les raisons ont été données, elles concernent essentiellement l'aspect sémantique et l'aspect syntaxique de la langue.

Des écarts que nous avons décelés à travers notre analyse des 10 résumés d'articles publiés sur ASJP dans des revues classées B et C. A ce propos il y avait plus d'écarts dans les résumés des articles publiés dans les revues classées C que dans ceux des articles publiés dans des revues classées B. Les articles publiés sur ASJP sont soumis à des experts (reviewers), mais ces derniers n'expertisent que la matière rédigée dans leur langue de spécialité, négligeant ainsi le résumé rédigé, si l'article est en français, en anglais ou en arabe. Notre degré de maitrise de l'anglais nous permet de corriger les résumés en anglais, qu'en est-il des experts qui eux ne le maitrisent pas forcément ? En fait ils essayent autant que faire se peut d'améliorer la version obtenue, mais rares sont ceux qui obtiennent un résultat satisfaisant. Notre étude se veut un préambule pour d'autres travaux qui se fixent, nous l'espérons, comme objectif de *vulgariser* la linguistique de corpus en tant que discipline et le TAL en tant qu'outil d'analyse permettant le traitement de corpus volumineux. Aussi faudra-t-il prendre en charge les utilisateurs du TAL, notamment les enseignants universitaires en les formant au maniement de ces outils, ces derniers devraient être perfectionnés pour une meilleure exploitation et pour des résultats plus fiables. Traduire est une compétence qui nécessite un savoir-faire et un savoir-être, c'est presque une seconde

nature, « Le bon traducteur, en effet, ne traduit pas de l'extérieur, mais en se situant délibérément à l'intérieur du texte à traduire dont il a retrouvé toute l'inspiration et toutes les harmoniques » (Testard, 1985 :3)

Références bibliographiques

Ambrasian, N. (2001). Le parallèle art et littérature. *Revue de littérature comparée*, pp. 304-309.

Bedjaoui, N. (2016). *La Perception du Français Chez Les Apprenants Algériens des Ecoles Privées de Langues Etrangères.* Constantine: Université des Frères Mentouri.

Calame-Griaule, G. (1981). Délimitation d'un terrain ethnolinguistique : la parole chez les dogons. *Langage et société*, pp. 9-11.

Détrie, C., Siblot, P., & Vérine, B. (2001). *Termes et Concepts pour l'Analyse du Discours. Une Approche Praxématique.* Paris: éd Honore Champion.

Freif, N. (2010). *Enabling Globalization: A Guide to Using Localization to Penetrate International Markets.* Montana: Global Vision International.

Giuliani, D., & Hannachi, R. (2010). Linguistique de corpus et didactique du F.L.E. Une exploitation du corpus IntUne. *Cahiers de Praxématique*, pp. 145-160.

Kurdi, M.-Z. (2017). *Traitement Automatique des Langues et Linguistique Informatique.* Londres: ISTE Group.

Lee, C. (2016). *Multilinguisme Online.* Abingdon: Taylor &Francis.

Loffler-Lorian, A.-M. (1996). *La Traduction Automatique.* Villeneuve-d'Ascq: Presses Univ. Septentrion.

Oustinoff, M. (2015). Histoire de la traduction. *La traduction*, pp. 25-46.

Raemdonck, D.-V., & Siouffi, G. (2012). *100 Fiches pour Comprendre le Linguistique.* Paris: Bréal.

Sinclair, J. (1996). *Preliminary recommendations on Corpus Typology.* Piza: EAGLE.

Testard, M. (1985). Les problèmes de la traduction. *Bulletin de l'Association Guillaume Budé*, pp. 2-29.

Zuffery, S. (2020). *Introduction à la Linguistique de Corpus.* Londres: ISTE Group.