



Analyse traductologique bilingue de l'existence d'un biais de genre dans les traductions automatiques par *DeepL* et *Google Translate*

Constanza Valentina Bastías

Université de Concepción, Chili
bastiasconstanza5@gmail.com
<https://orcid.org/0000-0002-6389-7228>

Javiera Fernanda Medone

Université de Concepción, Chili
javieramedone1995@gmail.com
<https://orcid.org/0000-0001-8925-0783>

Reçu le 10-10-2023 / Évalué le 20-11- 2023 / Accepté le 01-12-2023

Résumé

Grâce aux progrès technologiques constants de ces dernières années, des outils ont été créés pour faciliter le travail de traduction et permettre ainsi un accès rapide à la communication. L'un de ces outils est le traducteur automatique qui, bien qu'il présente des aspects positifs pour mener à bien cette tâche, n'est pas exempt d'aspects négatifs, comme, par exemple, le biais de genre qui peut être observé lorsqu'il y a des ambiguïtés dans un texte. Cet article propose l'existence ou l'absence d'un biais de genre dans les traductions automatiques effectuées par les traducteurs automatiques (TA) *DeepL* et *Google Translate*. Pour ce faire, des segments spécifiques de différents types de nouvelles ont été analysés et quantifiés.

Mots-clés : traduction, traducteur automatique, biais de genre, *DeepL*, *Google Translate*

Análisis traductológico bilingüe de la existencia de sesgo de género en las traducciones automáticas de *DeepL* y *Google Translate*

Resumen

Gracias a los constantes avances tecnológicos de los últimos años, se han creado herramientas que facilitan el proceso de traducción y proporcionan un acceso rápido a la comunicación. Una de estas herramientas es el traductor automático que, aunque tiene aspectos positivos para llevar a cabo esta tarea, no está exento de aspectos negativos, como el sesgo de género que se puede observar cuando existen ambigüedades en un texto. Este artículo plantea la existencia o no de sesgo de género en las traducciones automáticas realizadas por los traductores automáticos *DeepL* y

Google Translate. Para ello, se analizaron y cuantificaron segmentos específicos de distintos tipos de noticias.

Palabras clave: traducción, traductor automático, sesgo de género, *DeepL*, *Google Translate*

Bilingual translation analysis of the existence of gender bias in the machine translations of *DeepL* and *Google Translate*

Abstract

Thanks to the constant technological advances of recent years, tools have been created to facilitate the translation process and provide rapid access to communication. One of these tools is the automatic translator which, although it has positive aspects for carrying out this task, is not without negative aspects, such as the gender bias that can be observed when there are ambiguities in a text. This article proposes the existence or absence of gender bias in machine translations carried out by the machine translators *DeepL* and *Google Translate*. To do this, specific segments of different types of news were analysed and quantified.

Keywords : translation, automatic translator, gender bias, *DeepL*, *Google Translate*

Introduction¹

Communiquer, c'est arriver à partager quelque chose de soi. C'est une qualité rationnelle et émotionnelle spécifique de l'homme qui naît du besoin d'entrer en contact avec les autres, lorsqu'il échange des idées qui acquièrent un sens ou une signification en fonction d'expériences antérieures ou communes (Fonseca, 2005 : 18). Cela signifie qu'il existe une relation entre la langue et la société et que, par conséquent, ces deux concepts ont le pouvoir de changer et de s'influencer mutuellement, ce qui entraîne le maintien d'une réalité qui ne progresse pas.

La communication est liée à la traduction, puisque le traducteur professionnel est chargé de transmettre un message compréhensible pour le destinataire d'une langue à l'autre, en tenant compte de divers aspects tels que la culture, le sens et le contexte du texte original (Hurtado, 2001).

¹ Constanza Valentina Bastías s'est chargée des résumés, de l'Introduction, de l'analyse des résultats (parties 3 et 5), du résumé des résultats (partie 2) ; Javiera Fernanda Medone des parties 1 et 2, du corpus, de l'analyse des résultats (parties 1, 2, 4), du résumé des résultats (partie 1), de la conclusion.

Par ailleurs, il convient de noter que, compte tenu des progrès technologiques rapides, des outils ont été créés pour faciliter le travail de traduction, dont les traducteurs automatiques, qui se caractérisent par une traduction littérale des textes (Hernández, 2002). Cet outil peut présenter des aspects positifs et négatifs. En effet, l'un de ces aspects négatifs est le biais de genre qui peut être remarqué lorsqu'il y a des ambiguïtés dans un texte.

En ce qui concerne le genre, il convient de préciser que dans cette étude, il s'agit d'une catégorie linguistique, c'est-à-dire le genre grammatical qui, dans la langue espagnole, se manifeste par des morphèmes de genre masculin et féminin dans les noms et les adjectifs (Minervini, 2021). Ce phénomène peut donc être observé dans les traductions de l'anglais vers l'espagnol, car l'anglais n'a pas de morphèmes de genre et de nombre. Dans ce cas, il est donc important de comprendre le contexte d'un texte pour réussir la traduction, ce qui peut être difficile pour les traducteurs en ligne. Ainsi, en raison de ce besoin dans le monde contemporain et de l'urgence de rendre visible les préjugés sexistes qui pourraient être perçus dans notre travail de traducteurs, il a été décidé d'aborder cette question dans le cadre de la présente recherche.

1. Outils de traduction

Il existe actuellement différents outils pour compléter les compétences en matière de traduction, tels que la traduction assistée par ordinateur (Computer-Assisted Translation en anglais) qui facilitent le processus de traduction. La traduction assistée par ordinateur (désormais TAO) consiste à utiliser diverses sources d'information (dictionnaires, dictionnaires terminologiques, textes parallèles, encyclopédies, etc.) pour résoudre les doutes qui peuvent surgir chez un traducteur au cours du processus de traduction, à l'aide d'un ordinateur (Lippmann, 1971). En outre, grâce aux mémoires de traduction, qui font partie de la TAO, des informations provenant d'anciennes traductions et d'anciens textes peuvent être récupérées et stockées, ce qui permet au traducteur humain de gagner du temps et de maintenir le parallélisme si le même mot apparaît plus d'une

fois dans le texte (Oliver, 2016). Parmi les exemples de ces plateformes, citons *Trados*, *Transit* (Star), *Catalyst* (Alchemy), *Wordfast*, entre autres.

En fait, il existe de nombreux types d'outils communément appelés outils de TAO, en plus de leurs nombreuses propositions de classification selon différents critères tels que la distinction entre la traduction automatique et la mémoire de traduction (Hutchins, Somers, 1992).

2. Traduction automatique

La traduction automatique fait également partie des outils de traduction, mais elle n'est pas considérée comme une mémoire de traduction ou une TAO, car elle traduit un texte littéralement, en ignorant « les conditions extralinguistiques, [...] elle manque de bon sens, ne distingue pas les nuances, a une flexibilité limitée et ne peut pas s'appuyer sur un niveau pragmatique ou incorporer son expérience ou sa connaissance du monde » (Hernández, 2002 : 107). En outre, le traducteur humain ne peut intervenir que dans l'édition ultérieure du texte traduit, alors qu'en TAO, c'est le traducteur humain qui réalise l'ensemble du processus de traduction, en s'appuyant sur un accès rapide aux sources d'information. Parmi les traducteurs automatiques, on peut citer *Google Translate*, *DeepL*, *Reverso*, *Systran*, etc. Cependant, dans cette recherche, l'accent sera mis sur les deux outils choisis : *DeepL* et *Google Translate*.

3. Algorithme

Les traducteurs automatiques fonctionnent grâce à un système appelé algorithme, un terme créé au IX^e siècle par le mathématicien persan Abdullah Muhammad ibn Musa Al-Khwarizmi (Nabirahni *et al.*, 2019), qui a écrit le premier livre d'algèbre connu à ce jour. Ce terme trouve ses racines dans les mathématiques et est fortement lié au monde de la technologie.

En informatique, l'algorithme consiste en « une suite d'instructions séquentielles, grâce auxquelles certains processus peuvent être exécutés et certains besoins ou décisions peuvent être satisfaits » (Concepto, n.d.),

c'est-à-dire que l'algorithme se concentre sur la résolution de problèmes à travers une série d'étapes régies par la logique.

Pour ce type de traducteurs, traduire dans une centaine de langues (*Google Translate*) ou 29 langues (*DeepL*) est un défi et ces mécanismes de traduction le font au moyen d'automatismes, c'est-à-dire au moyen d'exemples de phrases déjà traduites sur internet. Ces automatismes utilisent des logiciels de traduction automatique statistique (TAE) (Padilla, 2018), chargés de rechercher des millions de documents sur internet, de les traduire en anglais comme langue intermédiaire, puis dans la langue cible. Le TAE utilise l'anglais comme langue intermédiaire parce qu'il y a plus de documents sur des sujets spécifiques en anglais sur internet, ce qui facilite et optimise le travail du traducteur automatique, car cela permet à la traduction d'être plus précise, en agissant comme un pont entre les autres langues. En d'autres termes, plus il y a de textes en anglais sur un sujet donné sur l'internet, meilleur est le résultat de la traduction.

4. Le biais cognitif et de genre

Le biais est défini comme « une erreur systématique qui peut se produire lorsque l'échantillonnage ou le test sélectionne ou favorise certaines réponses par rapport à d'autres » (Real Academia Española, n.d., définition 7). Le terme de biais cognitif a été inventé par les psychologues israéliens Kahneman et Tversky, qui le définissent comme une mauvaise interprétation systématique des informations disponibles qui influence la façon dont les pensées sont traitées, les jugements sont portés et les décisions sont prises (Kahneman, Tversky, 1987). Le phénomène des préjugés sexistes fait partie des préjugés cognitifs, car la réalité de la société est manipulée et l'on suppose que certaines responsabilités ou professions ne peuvent être exercées que par des hommes ou des femmes. Par exemple, dans le cas d'un professionnel de la santé spécialisé (médecin) et d'un professionnel des soins infirmiers (infirmier), il est associé que la profession spécialisée sera exercée par un homme et que le professionnel des soins infirmiers sera une femme. Ainsi, les préjugés sexistes peuvent être définis comme une inclination ou une tendance à attribuer un préjugé

à une personne en raison de son sexe, et les femmes sont généralement les plus touchées par ce type de préjugé (Ruiz-Cantero, Verdú-Delgado, 2004).

Le sociologue et écrivain allemand Eichler (1991) a été l'un des précurseurs de la recherche sur les préjugés sexistes et a rendu ce phénomène visible en classant les préjugés en trois types :

a) L'androcentrisme, qui consiste en l'identification du masculin à l'humain en général, rendant ainsi les femmes invisibles, ce qui se traduit dans le langage par l'utilisation générique du genre grammatical masculin dans un contexte inapproprié.

b) L'insensibilité au genre, qui ne considère pas le sexe et le genre comme des variables significatives dans les contextes, car elle ne remet pas en question les effets différenciés sur les femmes et les hommes, et favorise donc les inégalités. Ce type de biais peut être associé à l'échange des rôles attribués aux hommes et aux femmes dans les professions.

c) Les doubles standards, dus à l'utilisation de critères différents pour traiter et évaluer des situations ou des problèmes similaires ou identiques pour les deux sexes. Ce dernier est l'inverse de l'androcentrisme (Eichler, 1991).

Il faut donc noter que si ce phénomène modifie la perception d'une réalité, il peut également modifier le langage et donc la traduction des textes.

5. Recherches antérieures sur les préjugés sexistes dans la traduction automatique

En ce qui concerne les préjugés sexistes et la traduction elle-même, une étude de Stanovsky *et al.* (2019) a mené une évaluation à grande échelle des préjugés sexistes dans quatre traducteurs automatiques (*Google Translate*, *Microsoft Translator*, *Amazon Translate* et *Systran*). L'étude a mesuré la présence de biais de genre dans des phrases anglaises simples qui placent les participants dans des rôles de genre non stéréotypés. Par exemple, la phrase démonstrative utilisée était la suivante : « The doctor asked the nurse to help *her* in the operation » (« Le médecin a demandé à l'infirmière de l'aider dans l'opération »). Il est précisé que le médecin est une femme et que le pronom « *her* » est utilisé. Plus précisément, la phrase

était formulée de manière à ce qu'un traducteur humain puisse identifier le sexe du sujet de la phrase. Les phrases simples ont été traduites dans huit langues différentes, à savoir l'espagnol, le français, l'italien, le russe, l'ukrainien, l'hébreu, l'arabe et l'allemand, sur la base d'une analyse plus morphologique des résultats. Enfin, l'étude a conclu que les quatre traducteurs automatiques et « les modèles académiques de TA sont significativement enclins à traduire sur la base de stéréotypes de genre plutôt que sur le contexte » (Stanovsky *et al.*, 2019, p. 1683).

De même, la recherche de Medina (2021) a utilisé la traduction de mots et de phrases simples de l'anglais vers l'espagnol et a étudié l'influence de l'intelligence artificielle sur les traductions automatiques et l'existence d'une mécanisation des traditions, par exemple dans les rôles attribués aux professions, ainsi que les préjugés culturels et de genre, voire la discrimination dans les formes mêmes du langage, qu'il soit formel ou informel. Cette recherche montre l'existence de biais de genre dans les traducteurs automatiques *Google Translate*, *Bing Microsoft Translator* et *DeepL* et, par ailleurs, démontre l'intérêt d'améliorer et d'éradiquer ce phénomène dans ces outils de traduction. Plus précisément, l'étude montre que dans certains traducteurs automatiques, le pronom féminin de la troisième personne du singulier en anglais (*she*) est constamment traduit par le pronom masculin de la troisième personne du singulier (*he*). Toutefois, cette traduction n'est pas pertinente dans les langues sans genre grammatical, comme le turc, le finnois et le hongrois. Il n'en va pas de même pour le français et l'espagnol, qui sont considérés comme des langues à genre grammatical (Medina, 2021).

Une autre étude est celle de Minervini (2021) qui a examiné les traductions automatiques par *DeepL* et *Google Translate* de discours politiques et journalistiques dans la combinaison linguistique espagnol-italien. Le résultat est que les stéréotypes et les préjugés sexistes sont effectivement présents dans ce type de textes et que, dans la plupart des cas, les noms de professions ont une préférence pour le genre grammatical masculin, de sorte que, dans ce sens, il est important que les programmeurs qui choisissent les données à partir desquelles l'intelligence artificielle est alimentée aient une vision du monde informée et équilibrée, et même qu'ils soient des femmes, car les données reflètent la vision du monde du

programmeur. Il est essentiel de poursuivre les recherches sur ce phénomène pour éliminer les préjugés sexistes dans la traduction automatique (Minervini, 2021).

De même, Farkas et Németh (2021) ont mené une étude systématique sur la présence de préjugés sexistes uniquement dans *Google Translate*. Dans ce cas, des phrases contenant des noms de professions ont été traduites du hongrois, une langue neutre sur le plan de la prononciation, vers l'anglais. Des préjugés contre les deux sexes (féminin et masculin) ont été trouvés, mais les résultats biaisés contre les femmes sont beaucoup plus fréquents. On pourrait donc en conclure que les traductions sont plus proches de notre perception des professions et que les professions ont un effet plus important sur la traduction que les adjectifs (Farkas, Németh, 2021).

Les études examinées montrent que la plupart des recherches menées dans ce domaine datent de 2019 et que l'éventail n'est pas très large (4 seulement), car l'étude de ce phénomène est relativement récente. D'autre part, les traducteurs automatiques qui ont été étudiés dans ces études sont les mêmes que ceux qui seront repris dans la recherche suivante, à savoir *DeepL* et *Google Translate*. De plus, selon Johnson (2020), ingénieur logiciel et chercheur appartenant à l'équipe de *Google*, *Google Translate* a apporté des modifications pour éviter ce phénomène dans les phrases simples. Par conséquent, cette thèse ne se contentera pas de vérifier l'existence ou la non-existence de biais de genre, comme l'ont fait les recherches précédentes, mais vérifiera également si les deux, et pas seulement *Google Translate*, ont apporté des changements ou des améliorations pour éviter les biais de genre dans leurs traductions.

6. Corpus

Tout d'abord, le corpus est composé de 20 nouvelles ou articles :

- Source : The Guardian (journal anglais)
- Langue : Anglais.
- Longueur : entre 500 à 1000 mots.

- Section du journal : Politique (7), Événements internationaux (2), Sports (6), Culture (3), Écologie (2).

De même, les nouvelles se caractérisent par la présence de personnalités féminines en tant que protagonistes afin de vérifier la présence ou l'absence de préjugés sexistes.

Enfin, les nouvelles étant en anglais, elles seront considérées comme un corpus de traduction, car il analyse la traduction faite par le traducteur automatique *DeepL* et *Google Translate* du corpus collecté.

Une fois le corpus de base constitué, le corpus de traduction a été analysé. Pour cela, la traduction de chaque actualité a été réalisée à l'aide de *DeepL* puis le même exercice a été réalisé avec *Google Translate*. Les deux traducteurs automatiques disposent d'une version gratuite sur internet qui ne diffère pas des versions payantes, à l'exception du numéro de mots qu'ils traduisent. Cependant, cette fonctionnalité ne nuit pas au travail d'analyse.

Les textes ou articles ont été divisés en segments, qui correspondent aux paragraphes, pour faciliter l'analyse, et lorsque les traductions automatiques des deux traducteurs automatiques sont obtenues, l'existence ou la non-existence de préjugés sexistes dans le corpus est identifiée en développant un tableau composé de 6 colonnes : numéro de segment, segment en langue originale, segment traduit par *DeepL*, biais de genre (BG), segment traduit par *Google Translate* et biais de genre (BG). Ensuite, il est indiqué si un segment traduit du corpus compilé manifeste des préjugés sexistes, en marquant l'option « oui » ou « non » dans la case BG, comme illustré ci-dessous dans le tableau 1 :

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non

Tableau 1 : Tableau démontrant le développement de l'analyse.

De même, il est mentionné quel type d'aspect linguistique le préjugé sexiste présente dans les traductions générées automatiquement, c'est-à-dire si le traducteur automatique montre le préjugé sexiste à travers des

marques morphologiques de la langue, telles que l'androcentrisme ou l'insensibilité au genre (Eichler, 1991), liés aux métiers dans les thèmes abordés.

Enfin, le nombre de fois que ce phénomène se répète dans les traductions de chaque article généré par les TA est quantifié, dans le but de déterminer lequel des deux TA est le plus fiable lorsque le traducteur s'appuie sur cet outil de traduction. Les données quantifiées sont affichées sous forme de deux graphiques à barres ; le premier graphique montre l'apparition et la présence de préjugés sexistes dans les segments traduits par chaque TA, et le deuxième graphique montre la fréquence du type de préjugés sexistes dans les sujets abordés.

7. Analyse des résultats

7.1 Section de politique

Cette section présente deux exemples de politiques mettant en vedette des personnalités féminines comme protagonistes. La principale caractéristique de cette nouvelle est que la personnalité est introduite avec un prénom et un nom, puis reprise uniquement avec son nom de famille, ce qui donne une ambiguïté au contexte. Cela rend le texte difficile à comprendre. Ceci est illustré dans l'exemple suivant :

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non
4	Much of her support often stems from not being <u>a typical politician</u> [...].	Gran parte de su apoyo se debe a que no es <u>un político típico</u> [...].	X		Gran parte de su apoyo a menudo se debe a que no es <u>una política típica</u> [...].		X

Tableau 2 : Segment présentant de préjugé sexiste contre les femmes.

Dans ce segment, qui correspond au quatrième de la première nouvelle de la section politique, on observe que le poste accordé à Francia Márquez a été traduit par *DeepL* avec le genre grammatical masculin (« *un político típico* » en espagnol, « un politicien typique » en français), ce qui

correspondrait au biais de type insensibilité au genre, selon Eichler (1991), puisqu'il échange le rôle assigné aux femmes et le traduit comme si le poste était occupé par un homme.

En effet, du point de vue du TA, les postes politiques ou ceux qui nécessitent de plus grandes responsabilités sont occupés uniquement par des hommes et non par des femmes, c'est-à-dire que ces postes sont stéréotypés comme étant masculins.

De même, le même phénomène se produit pour les postes de *minister*, *frontrunner*, *prime minister*, *ambassador*, dans lesquels les préjugés sexistes sont plus présents dans TA *Google Translate* que dans *DeepL*.

D'autre part, il convient de mentionner que, bien qu'il soit plus courant de constater des préjugés sexistes à l'égard des femmes que des hommes, comme le soulignent les recherches de Farkas et Németh (2021), dans la présente étude, des cas ont été présentés dans lesquels les préjugés le genre était orienté vers le genre masculin, comme dans la nouvelle 1 de la section politique, dans laquelle seulement deux cas ont été observés, présentés ci-dessous :

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non
1	<u>The 23-year-old fashion designer</u> dressing Colombia's	<u>El diseñador de moda</u> de 23 años viste a la primera vicepresidenta		X	<u>La diseñadora de moda</u> de 23 años que viste a la primera vicepresidenta	X	
6	<u>The designer</u> said Márquez's outfits [...].	<u>La diseñadora</u> dijo que los trajes de Márquez [...].	X		<u>El diseñador</u> dijo que los atuendos de Márquez [...].		X

Tableau 3 : Segments présentant des préjugés sexistes contre les hommes.

Cette situation se produit également parce que le monde de la mode a toujours été lié au genre féminin et même si de nombreux hommes sont actuellement impliqués dans ce sujet, il existe encore un certain préjugé. De plus, une incohérence peut être observée dans les traductions des deux TA, puisque dans un premier temps *DeepL*, par exemple, reconnaît que le métier lié à la mode est exercé par un homme, mais reprend ensuite le métier avec le genre grammatical féminin.

7.2. Section d'événements internationaux

Puisque le corpus a été extrait d'un journal anglais, les événements survenus spécifiquement au Chili sont qualifiés d'internationaux. Comme dans les exemples illustrés dans le tableau 2, le type de préjugé sexiste d'insensibilité au genre est observé dans cette section, puisque les traductions de *leader*, *president* et *geographer* font référence à une personnalité féminine, Camila Vallejo. Le tableau suivant montre que les préjugés sexistes sont présents dans les traductions réalisées par *Google Translate* :

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non
1	The then leader of Chile's [...], Vallejo – a would-be geographer [...].	Líder entonces del sindicato [...] de Chile, Vallejo - una aspirante a geógrafa [...].		X	El entonces líder del sindicato [...] de Chile, Vallejo, un aspirante a geógrafo , [...].	X	
6	Because Vallejo is no longer the Fech president .	Porque Vallejo ya no es la presidenta de la Fech.		X	Porque Vallejo ya no es presidente de la Fech.	X	

Tableau 4 : Segments présentant des préjugés sexistes dans la section d'événements internationaux.

7.3. Section des sports

Ce type de préjugé sexiste (insensibilité au genre) se produit non seulement dans le domaine de la politique et des événements internationaux, mais également dans le domaine du sport, notamment dans le football, qui est considéré comme un sport masculin. Dans cette section, même s'il est fréquent de voir des femmes diriger dans de nombreux domaines sportifs, les TA ont encore du mal à assimiler leur participation. Pour cette raison, on observe que des positions dans le football telles que « *entrenador* » (« entraîneur »), « *goleador* » (« buteur »), « *capitán* » (« capitaine »), entre autres, ont été traduites par *Google Translate* avec le genre grammatical masculin, comme l'illustre le tableau suivant :

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non
1	[...] White was the tournament's joint top scorer with six goals.	[...] White fue la máxima goleadora del torneo con seis tantos.		X	[...] White fue el máximo goleador conjunto del torneo con seis goles.	X	
2	[...] presence of the Lionesses' « third captain ».	[...] la presencia de la " tercera capitana " de las Leonas.		X	[...] la presencia del « tercer capitán » de las Lionesses.	X	
3	[...] <u>England's manager</u> identified [...]	[...] la seleccionadora de <u>Inglaterra</u> consideró [...].		X	[...] el entrenador de <u>Inglaterra</u> identificó [...].	X	

Tableau 5 : Segments présentant des préjugés sexistes dans la section des sports.

Dans d'autres sports comme le cyclisme, l'athlétisme et le tennis, non seulement ce type de préjugé sexiste existe, mais aussi celui associé à l'utilisation du genre grammatical masculin comme générique dans un contexte inapproprié, c'est-à-dire l'androcentrisme (Eichler, 1991). L'exemple suivant montre que la personnalité féminine est mentionnée plus tôt (Raducanu), mais les deux TA ne reconnaissent pas le genre grammatical féminin. Par conséquent, dans le tableau 6 ci-dessous, les deux traducteurs automatiques présentent un préjugé sexiste de type androcentrisme (Eichler, 1991).

N° segment	Segment en langue original	Segment traduit par <i>DeepL</i>	BG		Segment traduit par <i>Google Translate</i>	BG	
			Oui	Non		Oui	Non
3	Raducanu proved to be her own worst enemy , [...]	Raducanu demostró ser su peor enemigo , [...].	X		Raducanu demostró ser su peor enemigo , [...].	X	

Tableaux 6 : Segment présentant de préjugé sexiste dans le domaine du sport.

7.4. Section de culture

Dans le domaine de la culture, il est démontré que *Google Translate* est le traducteur qui présente le plus de préjugés sexistes d'insensibilité au

genre, car, comme dans l'exemple précédent, les personnalités féminines des trois actualités analysées dans cette section sont utilisées uniquement avec leur nom de famille et les professions qu'elles exercent sont traduites avec le genre grammatical masculin. Les traducteurs automatiques ont été confrontés à l'ambiguïté que présente le texte lorsqu'il fait référence à des personnalités féminines uniquement avec leur nom de famille, ce qui a empêché la compréhension du contexte de ces outils de traduction, comme le montre l'exemple suivant :

N° segment	Segment en langue original	Segment traduit par DeepL	BG		Segment traduit par Google Translate	BG	
			Oui	Non		Oui	Non
2	[...] Leonskaja was like <u>a climber</u> [...].	[...] Leonskaja fue como <u>una escaladora</u> [...].		X	[...] Leonskaja era como <u>un escalador</u> [...].	X	

Tableau 7 : Segment présentant de préjugé sexiste dans le domaine de la culture.

7.5. Section d'écologie

Il est démontré que dans le troisième segment de cette section, les deux TA présentent des préjugés sexistes, puisqu'ils ne reconnaissent pas que le nom « *hero* » (« héros ») fait référence à une personnalité féminine. L'exemple suivant montre le type de préjugé sexiste d'androcentrisme, puisque les deux TA n'ont pas reconnu le sexe de la protagoniste de l'actualité, Jane Fonda.

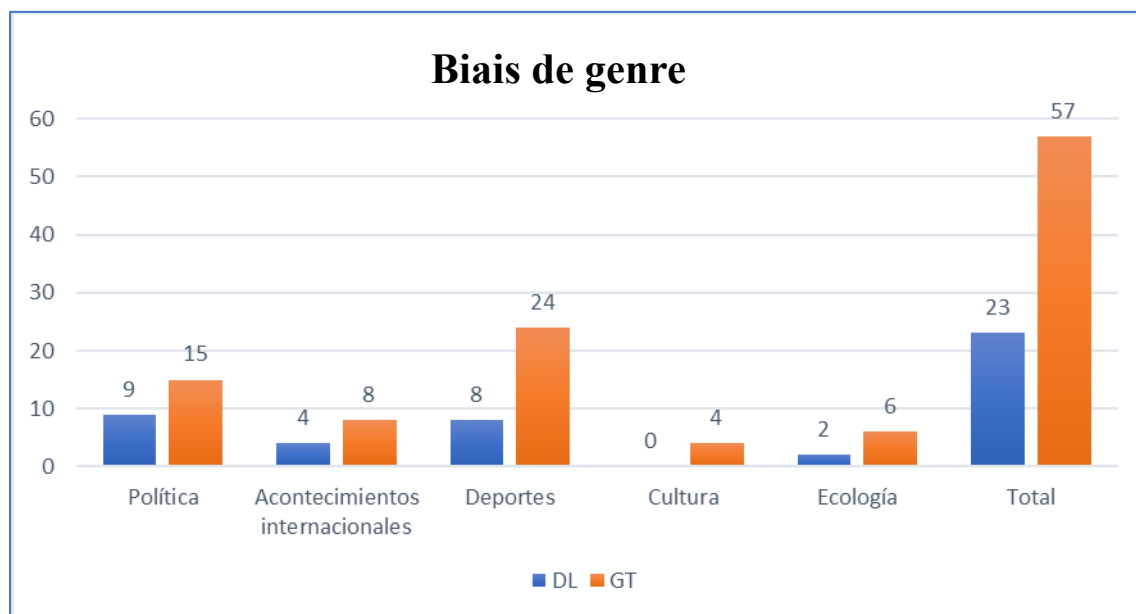
N° segment	Segment en langue original	Segment traduit par DeepL	BG		Segment traduit par Google Translate	BG	
			Oui	Non		Oui	Non
3	I don't mean to cast myself as <u>a hero</u> : [...]	No pretendo presentarme como <u>un héroe</u> : [...].	X		No pretendo presentarme como <u>un héroe</u> : [...].	X	

Tableau 8 : Segment présentant de préjugé sexiste dans le domaine de l'écologie.

8. Résumé des résultats

Bien que des recherches antérieures confirment la présence de préjugés sexistes dans les TA choisies, aucune ne présente de données quantifiables

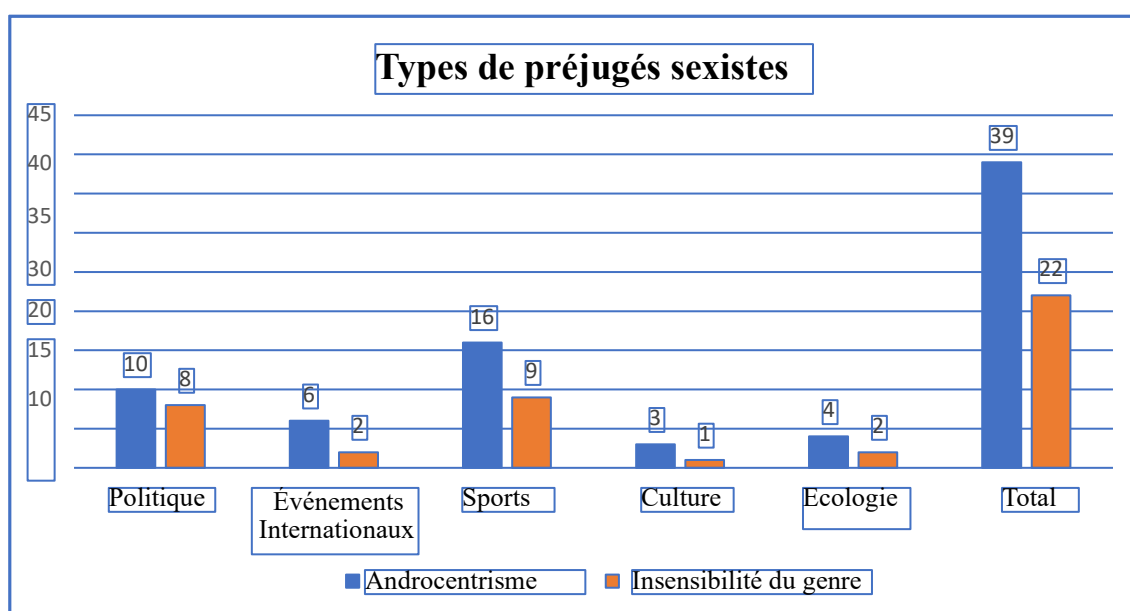
sur les résultats. Dans ce cas, les résultats par sections étudiées sur l'apparition et la présence de préjugés sexistes dans les traducteurs *DeepL* et *Google Translate*, ainsi que la fréquence du type de préjugé présenté, seront présentés en préparant des graphiques à barres, comme illustré ci-dessous :



Graphique 1 : Apparition et présence de préjugés sexistes dans toutes les sections.

Le graphique 1 nous permet de conclure qu'il existe un total de 61 segments analysés dans lesquels l'existence et l'apparition de préjugés sexistes sont observées dans le corpus compilé à partir de 20 articles journalistiques appartenant au journal The Guardian. De même, on constate que le TA *DeepL*, représenté en bleu, a présenté ce phénomène en 23 segments, soit 38 % du total, tandis que le TA *Google Translate*, surligné en orange, a exposé ce phénomène en 57 segments, ce qui correspond à 93 % du total. Ainsi, comme Medina (2021), on peut conclure que *DeepL* interprète la langue espagnole avec une plus grande précision. En outre, Minervini (2021) indique qu'il est plus évident de trouver des valeurs masculines, c'est-à-dire des préjugés sexistes, dans *DeepL* que dans *Google Translate*, dans la traduction de discours politiques et journalistiques, surtout lorsqu'ils concernent des domaines culturellement sensibles, attribués aux hommes (Minervini, 2021). Cependant, les résultats de cette

recherche différent de ceux de Minervini (2021), puisqu'il a été constaté que *Google Translate* présentait plus de préjugés sexistes que *DeepL* dans les segments analysés. De plus, même si les préjugés sexistes avaient déjà été mis en évidence, analysés et confirmés dans *Google Translate*, cet outil ne semble pas avoir apporté de modifications pour les éviter dans ses traductions, comme l'avait souligné Johnson (2020). Par ailleurs, compte tenu du point de vue sociologique soulevé par Eichler (1991), on conclut que les types de préjugés sexistes présentés dans les deux TA étaient l'androcentrisme et l'insensibilité au genre, comme l'illustre le graphique suivant :



Graphique 2 : Fréquence du type de préjugé sexiste.

Les données présentées dans le graphique 2 permettent de percevoir que le type de biais le plus fréquent dans les 61 segments analysés est l'androcentrisme. De plus, le thème du sport est celui qui présente le plus de marques de ce type de biais, liés à l'utilisation du genre grammatical masculin comme générique dans un contexte inapproprié. Cela signifie que les TA traduisent les adjectifs et les noms qui font référence à des personnalités féminines de ce genre grammatical dans les sports exposés dans l'actualité sur ce sujet (football, cyclisme, entre autres), et supposent que ceux-ci sont pratiqués par des hommes. En revanche, ces données indiquent que le résultat obtenu dans cette étude est en partie contraire à

celui qui a été obtenu par Minervini (2021), puisqu'elle conclut que, dans la plupart des cas, les noms de professions ont une préférence pour le genre masculin grammatical, ce qui peut s'expliquer par une insensibilité au genre (Eichler, 1991). Cependant, dans cette étude, l'insensibilité au genre est moins clairement mise en évidence.

Conclusion

En conclusion, il a été prouvé que les préjugés sexistes sont bel et bien un phénomène retrouvé dans les traductions de l'anglais vers l'espagnol générées par les traducteurs automatiques *DeepL* et *Google Translate*. Il convient de noter que l'analyse des segments nous a permis d'identifier certains contextes d'apparition de ce phénomène dans les deux TA, comme l'androcentrisme et l'insensibilité au genre, selon Eichler (1991), c'est-à-dire l'utilisation du genre grammatical masculin générique dans un contexte inapproprié, et l'échange des rôles attribués aux hommes et aux femmes. En outre, nous pensons, à travers les preuves existantes et les résultats analysés, qu'il existe davantage de préjugés sexistes à l'égard des femmes, car nous sommes toujours confrontés à un monde caractérisé par une pensée sexiste et de nombreuses disparités entre les sexes. De cette manière, la communication joue un rôle extrêmement important, car elle est le reflet de la société dans laquelle nous vivons. Il est donc évident que nous devons encore faire face à de nombreux obstacles pour renverser les stéréotypes et les préjugés imposés par la société.

En revanche, la raison pour laquelle un corpus de traduction a été choisi, composé de textes avec des personnalités féminines spécifiques comme protagonistes, est ce qu'expliquent Farkas et Németh (2021). Comme mentionné ci-dessus, ces chercheurs indiquent qu'il est plus courant de constater des préjugés sexistes à l'encontre des femmes que des hommes. Cet aspect a donc été considéré pour analyser à la fois l'aspect morphologique et l'aspect lié à l'échange des rôles attribués aux hommes et aux femmes.

En ce qui concerne les limites au cours de la réalisation de cette étude, il convient de mentionner, tout d'abord, la disponibilité limitée de recherches sur les préjugés sexistes autour de la traduction automatique,

car, comme mentionné tout au long de ce travail, il s'agit d'un phénomène relativement récent. De plus, il peut également y avoir une limitation culturelle car ces documents peuvent avoir été rédigés dans une langue inconnue. Deuxièmement, le corpus constitué est petit et, étant limité à un seul domaine, celui des textes journalistiques, il ne permet pas de généralisations. Enfin, troisièmement, les traducteurs automatiques analysés n'étaient que deux (*DeepL* et *Google Translate*), qui, bien qu'ils soient deux des TA les plus reconnus et utilisés par l'utilisateur commun, sont également les deux outils les plus analysés dans les recherches précédentes. De même, on pourrait approfondir le nombre de langues étudiées, comme dans les recherches menées par Stanovsky *et al.* (2019).

À partir des preuves recueillies dans cette étude, il est possible de souligner que les préjugés sexistes sont un phénomène présent dans la traduction et qu'il est donc nécessaire d'approfondir encore plus ce domaine. Par exemple, il est suggéré d'étudier d'autres types de genres textuels, de considérer d'autres langues et d'autres TA et d'approfondir les contextes dans lesquels les préjugés sexistes se produisent, car, en les approfondissant, ou plutôt en étudiant le contexte dans lequel ce phénomène est présent, il signifierait une évolution des algorithmes et des logiciels, ce qui permettrait d'améliorer les TA, et ainsi d'être plus utiles et efficaces lors de la traduction.

Bibliographie

Concepto. (s.f.). Algoritmo en Informática. En la Enciclopedia Concepto.

[En ligne] : <https://concepto.de/algoritmoen-informatica/>

[consulté le 24 septembre 2022].

Eichler, M. 1991. *Nonsexist Research Methods : A Practical guide*. Routledge.

Farkas, A., Németh, R. 2021. « How to measure gender bias in machine translation : Real-world oriented machine translators, multiple reference points ». *Social Sciences & Humanities Open*, 5, 1-11.

Fonseca, M. 2005. *Comunicación oral: Fundamentos y práctica estratégica* (2^a. ed). Prentice Hall.

Hernández, P. 2002. «En torno a la traducción automática». *Cervantes*, 2, p. 101-117.

Hurtado, A. 2001. *Traducción y Traductología. Introducción a la traductología*. Cátedra.

- Hutchins, W.J., Somers, H. L. 1992. *An introduction to Machine Translation*. Cambridge University Press.
- Johnson, M. (22 de abril de 2020). A Scalable Approach to Reducing Gender Bias in Google Translate. Google AI Blog. <https://ai.googleblog.com/2020/04/ascaleable-approach-to-reducing-gender.html> [consulté le 24 septembre 2023].
- Kahneman, D., Tversky, A. 1987. «Teoría prospectiva: un análisis de la decisión bajo riesgo». *Infancia y Aprendizaje*, 30, p. 95-124.
- Lippmann, E. 1971. An Approach to Computer-Aided Translation. *IEEE Transactions on Engineering Writing and Speech*, 14 (1), p. 10-33.
- Medina, C. 2021. Influencias en el idioma español de las traducciones del inglés basadas en Inteligencia Artificial. IA EñTM. [En ligne] : <https://doi.org/10.21428/39829d0b.6db1fabb> [consulté le 24 septembre 2023].
- Minervini, R. 2021. «La traducción automática del género (español-italiano): análisis de ejemplos traducidos con *DeepL* y *Google Traductor*». *International Journal of Translation*, n° 23, p. 105-127.
- Nabirahni, D. Evans, B., Persaud, A. 2019. Al-Khwarizmi (Algorithm) and the Development of Algebra. *Mathematics Teaching Research Journal*, 11 (1-2), p. 13-17.
- Oliver, A. 2016. *Herramientas tecnológicas para traductores*. Barcelona: Editorial UOC.
- Padilla, O. (15 de diciembre de 2018). ¿Cómo funciona el traductor de Google? Te lo explicamos en detalle. El Español. [En ligne] : https://www.elespanol.com/elandroidelibre/tutoriales/20181215/funcionatraductor-google-explicamos-detalle/360964596_0.html [consulté le 24 septembre 2023].
- Real Academia Española. (s.f.). Sesgo. En el Diccionario de la lengua española. [En ligne] : <https://dle.rae.es/sesgo> [consulté le 27 juin 2023].
- Ruiz-Cantero, M., Verdú-Delgado, M. 2004. «Sesgo de género en el esfuerzo terapéutico». *Gaceta Sanitaria*, 18 (4), p. 17-21.
- Stanovsky, G., Smith, N., Zettlemoyer, L. 2019. Evaluating Gender Bias in Machine Translation. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. [En ligne] : <https://aclanthology.org/P19-1164.pdf> [consulté le 27 juin 2023].



© *Synergies Chili*, n° 19, Année 2023.

Revue du GERFLINT

ARK : <https://catalogue.bnf.fr/ark:/12148/cb426944659>

Bibliothèque nationale de France.

- Décembre 2023 -

Éléments sous droits d'auteur. Modalités de lecture et de citation, politique d'archivage et mentions légales consultables sur le site de l'éditeur et de la revue :

<https://gerflint.fr/>

<https://gerflint.fr/synergies-chili>

