



ISSN 1724-0700

ISSN en ligne 2260-8087

# La traduction automatique neuronale et les biais de genre : le cas des noms de métiers entre l'italien et le français

**Eleonora Marzi**

Université de Bologne, Italie

eleonora.marzi@unibo.it

<https://orcid.org/0000-0002-4558-9483>

Reçu le 10-02-2021 / Évalué le 15-03-2021 / Accepté le 18-04-2021

## Résumé

Au cours des dernières années, la traduction automatique (TA) a fait des progrès remarquables grâce à l'utilisation de réseaux de neurones artificiels ; toutefois, l'intérêt et le succès des logiciels de TA - rendus disponibles gratuitement pour le grand public - ont mis en évidence la présence de biais de genre dans les résultats obtenus. Ce biais pourrait dépendre d'une biodiversité insuffisante de données qui servent à l'entraînement des algorithmes d'intelligence artificielle. Prenant pour exemple les noms de métiers, cette étude entend vérifier si la présence de deux différents types d'adjectifs (d'apparence et de compétence) peut influencer la traduction automatique qui concerne le genre en allant du français vers l'italien ou vice-versa. La liste des noms de métiers analysés, qui a été rédigée à partir du système de classification des professions européennes ESCO (*European Skills, Competences, Occupation*), a permis de créer des phrases que nous avons fait traduire par trois moteurs de traduction automatique neuronale : *Google Translate*, *DeepL*, *Microsoft Translator*.

**Mots-clés :** études de genre, écart entre les sexes, apprentissage profond, ESCO (*European Skills, Competences*), traduction automatique

## La traduzione automatica neurale e il divario di genere nel linguaggio: il caso delle professioni tra italiano e francese

### Riassunto

Durante gli ultimi anni, la traduzione automatica (TA) ha fatto dei progressi notevoli grazie all'approccio delle reti neurali; l'interesse e il successo degli strumenti di TA - resi disponibili gratuitamente al grande pubblico - hanno messo in evidenza la presenza di un divario di genere nei risultati. Tale divario potrebbe dipendere da una biodiversità insufficiente dei dati che servono ad allenare l'algoritmo d'intelligenza artificiale. Prendendo in considerazione il caso dei nomi delle professioni, il presente studio vuole verificare in che modo la presenza di due diversi tipi di aggettivi (d'apparenza e di competenza) possa influenzare la traduzione automatica del genere dal francese all'italiano o viceversa. L'analisi è stata condotta utilizzando una lista di nomi di professione che abbiamo selezionato a partire dal sistema di classificazione delle professioni europee ESCO (*European Skills, Competences*,

*Occupation*) e che è servita per creare delle frasi che abbiamo fatto tradurre da tre motori di traduzione neuronale: *Google Translate*, *DeepL*, *Microsoft Translator*.

**Parole chiave:** studi di genere, *gender data gap*, *deep learning*, ESCO (*European Skills, Competences, Occupation*), traduzione automatica

### Neural machine translation and gender gap in language: the case profession names in Italian and French

#### Abstract

In recent years, machine translation (MT) has made considerable progress thanks to the neural network approach. The interest and the success of MT tools - made freely available to the general public - have highlighted the presence of a gender gap in the results. This gap could be due to an insufficient biodiversity of the data used to train the artificial intelligence algorithm. Considering as a case study the names of professions, the present article aims to verify how the presence of two different types of adjectives (appearance and competence) can influence the automatic translation of gender from French to Italian or vice versa. The analysis was carried out using a list of profession names that we selected from the ESCO Project (*European Skills, Competences, Occupation*) and which was used to create sentences that we had translated by three neural translation engines: *Google Translate*, *DeepL*, *Microsoft Translator*.

**Keywords:** gender studies, *gender data gap*, *deep learning*, ESCO (*European Skills, Competences, Occupation*), automatic translation

#### 1. La traduction automatique : fonctionnement et typologies

Dans son essai *Dire presque la même chose*, Umberto Eco donne l'exemple de textes traduits de manière automatique qui deviennent presque anecdotiques. La réflexion d'Eco exemplifie les deux sentiments opposés que la traduction automatique suscite depuis toujours. D'un côté les technophobes (Eco, 1964), convaincus d'une distance infranchissable entre la performance humaine et celle de la machine, et de l'autre, les technophiles qui rêvent d'un instrument capable d'imiter les capacités humaines.

L'histoire de la traduction automatique oscille entre ces deux positions, scandée par l'introduction de nouvelles technologies, notamment les corpus numériques et les techniques d'apprentissage liées à l'intelligence artificielle. Ce mouvement incessant entre la confiance envers une traduction automatique comparable à ce que ferait l'être humain, et le découragement qui la voit condamnée à un état d'imperfection, n'a jamais cessé et a produit diverses générations d'élaboration dans le domaine de la traduction, toutes basées chacune sur une idée différente de la traduction.

Les tout premiers systèmes de traduction automatique (TA) naissent autour des années 1950. Ils sont fondés sur une approche de la traduction entendue comme décryptage d'un code (Weaver, 1949), traduisant donc mot à mot. Les éléments qui composaient ces systèmes étaient un compilateur (*parser*), permettant d'identifier la relation entre les mots de la langue source et ceux de la langue d'arrivée, un dictionnaire monolingue, avec les informations morphosyntaxiques de la langue source, et un dictionnaire bilingue, utilisé pour permettre au système de trouver l'équivalent dans la langue cible. Cette traduction mot à mot rendait évidemment difficile la traduction des expressions métaphoriques ou ambiguës.

L'attention portée à la grammaire a permis de développer une deuxième génération de systèmes informatiques à base de règles (*rule-based*). Ces systèmes partent d'une conception théorique de la traduction pour laquelle, grâce à l'analyse de la représentation du signifié dans la langue source, il est possible de créer une représentation équivalente dans la langue cible<sup>1</sup> ; au niveau technologique, elles sont construites à travers une série de règles morphologiques, syntaxiques et sémantiques. Si ce type de système a l'avantage de traduire correctement les structures syntaxiquement non parallèles entre deux langues, il présente encore des défauts. Tout d'abord, l'impossibilité d'incorporer dans une liste toutes les règles grammaticales d'une langue. Ensuite, pour une interprétation correcte des ambiguïtés liées aux termes homographes et polysémiques, il est nécessaire de prendre en considération le cotexte et le contexte. Les travaux successifs ont emprunté cette direction et, aux alentours des années 1980, l'apparition de corpus de grande taille en format numérique a permis de créer des systèmes basés sur les corpus (*corpus-based*). Ce modèle s'appuie sur une grande quantité d'exemples corrects et alignés, qui sont identifiés grâce à des calculs statistiques. Il s'agit, en quelque sorte, d'une tentative de résoudre le problème lié à la connaissance du contexte : par exemple, l'expression idiomatique italienne « *in bocca al lupo* » est traduite en français par « bonne chance » en raison de la récurrence de ces séquences, et de leur co-occurrence régulière dans des segments alignés correspondants. Toutefois, cette solution présente une limite : se basant sur des statistiques tirées d'exemples du corpus, le système n'est pas en mesure de traduire des expressions linguistiques inhabituelles ou poétiques et plus généralement les expressions qui présentent un certain niveau de créativité ou qui sont peu représentées dans le corpus.

Au début des années 2000, un tournant important est marqué par l'introduction des techniques issues de l'intelligence artificielle. Le système de TA devient alors capable d'apprendre et d'améliorer ses performances grâce à des algorithmes entraînés sur des quantités énormes de données non structurées. En 2014, les réseaux de neurones artificiels apparaissent : cette architecture entend imiter

les réseaux biologiques des neurones du cerveau humain grâce à une organisation par niveaux qui interagissent entre eux à travers des nœuds contenant des centres de connaissance. L'élaboration des données grâce aux réseaux de neurones artificiels permet de créer des modèles de connaissance qui s'améliorent sans cesse. Cette technologie, qui s'appuie sur des systèmes définis à base de connaissance (*knowledge-based*), est utilisée pour la traduction automatique neuronale (*Neural Machine Translation - NMT*). Bien qu'elle exploite divers outils - tels que dictionnaires, règles grammaticales et sémantiques, glossaires spécialisés et mémoires de traductions - elle représente en quelque sorte une évolution de la traduction basée sur les corpus (Wilks, 2009).

## 2. Le biais de genre et la traduction automatique

En 2016, *Google* a lancé son traducteur automatique neuronal ; en 2018, *Microsoft*, *Systran* et *Linguee* en ont fait de même. La mise à disposition gratuite des outils de TA a modifié l'interface des logiciels suggérant par exemple à l'utilisateur une traduction correcte ou lui permettant d'évaluer la traduction faite par le moteur. Cette pratique ne découle pas d'une sensibilité démocratique envers la connaissance, mais elle vise à tirer profit de la compétence linguistique des utilisateurs qui, de cette manière, deviennent des experts linguistiques bénévoles.

La visibilité majeure acquise récemment par la traduction automatique ainsi que l'usage massif de ce type de traduction ont conduit à une prise de conscience à propos des stéréotypes de genre dans les traductions<sup>2</sup>. Cette tendance s'est accrue grâce à un intérêt grandissant et à une sensibilité renouvelée pour la question du genre dans le langage, dont témoigne, entre autres, le débat récent sur le langage épïcène<sup>3</sup>.

De nombreuses études, que nous préciserons dans cette section, supposent que l'existence du biais de genre dans les traductions automatiques serait due à une biodiversité insuffisante des données d'entrées. En effet, les architectures neuronales constituent l'évolution des architectures à base de corpus, qui sont alors unies aux approches d'apprentissage profond (*deep learning*) : cela veut dire que les données de sortie sont grandement affectées par la qualité des données d'entrées qui entraînent les algorithmes d'intelligence artificielle. Bien que l'on obtienne de meilleurs résultats lorsque le moteur de traduction automatique est entraîné sur un corpus représentant le même type de discours spécifique que celui du texte à traduire, il faut aussi considérer que la langue (et par conséquent le corpus) est le reflet de la société : longtemps, certains noms de métiers n'ont pas disposé d'équivalents féminins, puisque ces métiers n'étaient jamais exercés par

des femmes. Par exemple, si le mot « infirmière » est normalement bien attesté, c'est aussi parce que cette profession, qui a subi une forte croissance depuis la première et la seconde guerre mondiale, était exercée habituellement par des femmes ; il n'en va pas de même pour le mot « ingénieure », qui est apparu bien plus tard parce que ce métier a été longtemps réservé aux hommes.

Bien que les stéréotypes de genre présents dans le langage varient au fil des années dans le sens d'un rééquilibrage (Jones *et al.*, 2020), l'écart est encore présent. Caroline Criado Perez (2019) parle de décalage dans les données sexospécifiques (*gender data gap*), en faisant référence à l'absence d'une représentation adéquate des femmes dans les bases de données statistiques ; de manière similaire, Ricardo Baeza-Yates (2018) parle de biais algorithmique (*algorithmic bias*) en signalant que le bas pourcentage de femmes dans le secteur de la programmation par rapport à celui des hommes aurait des conséquences sur la conception des algorithmes.

À la fin de l'année 2018, Google a officiellement admis la présence des stéréotypes de genre au sein de ses applications automatiques<sup>4</sup> (traduction, suggestion de rédaction dans *Gmail*, etc.) et a également publié un article sur les mesures adoptées pour résoudre cet écart<sup>5</sup>.

Si l'existence de stéréotypes de genre est donc prouvée, les méthodes utilisées pour les détecter peuvent être variées. Dans une étude sur le moteur de TA de Google, Marcelo Prates (Prates *et al.*, 2019) a sélectionné plus de 1 000 professions, une liste des adjectifs les plus fréquents selon le *Corpus of Contemporary American English* (COCA) et une structure syntaxique à double possibilité, sans ou avec l'adjectif. Les résultats montrent que le système traduit préférentiellement les pronoms personnels au masculin plutôt qu'au féminin et que la présence de certains adjectifs comme « *Arrogant, Cruel and Guilty* » (« arrogant, cruel et coupable », *ibidem* : 24) porte l'algorithme à privilégier la traduction au masculin, alors que d'autres adjectifs comme « *Shy, Attractive, Happy, Kind and Ashamed* » (« timide, séduisant, heureux, gentil, honteux/qui a honte », *idem*) facilitent la traduction au féminin, suggérant une représentation stéréotypée des genres. Stanovsky (Stanovsky, G. *et al.* 2019) parvient à la même conclusion en constatant que la traduction devient plus précise en présence d'une sorte d'amalgame entre certaines professions et le genre, comme dans « *female nurse* » (« infirmière »). De même, la structure syntaxique a fait l'objet d'études qui ont essayé d'identifier des éléments critiques ou de faciliter une traduction correcte. Costa Jussà (2019), qui propose d'intégrer des mots pour mitiger l'impact des stéréotypes de genre (*words embeddings*), constate que les noms propres contribuent à réduire l'ambiguïté. Enfin, Monti (2017) identifie des constructions comme les anaphores,

les appositions et l'accord sujet/objet avec le complément prédicatif qui seraient problématiques. Toutes ces études nous ont servi de point de départ théorique pour sélectionner une liste de noms de métiers que nous allons présenter dans la section suivante.

### 3. Méthodologie d'analyse : les noms de métiers et la structure syntaxique

En partant de la prémisse que les systèmes de TA génèrent des traductions erronées en termes de genre, dues entre autres à un manque de biodiversité dans les données d'entrée, nous allons nous intéresser au type d'erreurs de TA, dans la conviction que cette analyse pourrait éclairer la nature des stéréotypes présents. Notre objectif n'est pas d'évaluer la performance des logiciels de TA, mais de savoir si les erreurs sont simplement aléatoires, ou bien si elles sont révélatrices de stéréotypes provenant d'un imaginaire caché.

Tout en considérant les travaux précédemment mentionnés, nous avons tout d'abord sélectionné une liste de noms de métiers, extraite de celle du projet ESCO<sup>6</sup> (*European Skills Competences and Occupations*), cette dernière comptant 2 942 noms de professions en 27 langues européennes. La liste ESCO s'appuie à son tour sur un code ISCO 08<sup>7</sup> qui divise les professions en 10 groupes : Professions militaires ; Directeurs, cadres de direction et gérants ; Professions intellectuelles et scientifiques ; Professions intermédiaires ; Employés de type administratif ; Personnel des services directs aux particuliers, commerçants et vendeurs ; Agriculteurs et ouvriers qualifiés de l'agriculture, de la sylviculture et de la pêche ; Métiers qualifiés de l'industrie et de l'artisanat ; Conducteurs d'installations et de machines, et ouvriers de l'assemblage ; Professions élémentaires.

Nous avons sélectionné 73 noms de métiers équivalents en français et en italien (tableau 1), en suivant un critère de représentativité générale en termes de typologie ; nous avons également préféré retenir les seuls hyperonymes plutôt que de multiplier les hyponymes. Par exemple, en ce qui concerne le secteur médical, nous avons préféré le mot « docteur » aux hyponymes « dermatologue », « neurologue », « pharmacologue », « pneumologue », « psychiatre », « pédiatre », etc.

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
directeur	directrice	direttore	direttrice
editeur	editrice	editore	editrice
physicien	physicienne	fisico	fisica
docteur	docteure	dottore	dottoressa

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
infirmier	infirmière	infermiere	infermiera
professeur	professeure	professore	professoressa
programmeur d'applications	programmeuse d'applications	sviluppatore di applicazioni	sviluppatrice di applicazioni
avocat	avocate	avvocato	avvocata
écrivain	écrivaine	scrittore	scrittrice
traducteur	traductrice	traduttore	traduttrice
compositeur	compositrice	compositore	compositrice
danseur	danseuse	ballerino	ballerina
acteur	actrice	attore	attrice
technicien	technicienne	tecnico	tecnica
opticien	opticienne	ottico	ottica
décorateur	décoratrice	decoratore	decoratrice
cuisinier	cuisinière	cuoco	cuoca
employé	employée	impiegato	impiegata
serveur	serveuse	cameriere	cameriera
coiffeur	coiffeuse	parruchiere	parrucchiera
commerçant	commerçante	commerciante	commerciante
vendeur	vendeuse	commesso	commessa
pompier	pompière	vigile del fuoco	vigilessa del fuoco
agent de police	agente de police	agente di polizia	agente di polizia
agriculteur	agricultrice	agricoltore	agricoltrice
ouvrier	ouvrière	operaio	operaia
éleveur	éleveuse	allevatore	allevatrice
pêcheur	pêcheuse	pescatore	pescatrice
chasseur	chasseuse	cacciatore	cacciatrice
potier	potière	vasaio	vasaia
artisan	artisane	artigiano	artigiana
boulangier	boulangère	panettiere	panettiera
pâtissier	pâtissière	pasticcere	pasticcera
barman	barmaid	barman	barmaid
ingénieur	ingénieure	ingegnere	ingegnere
pharmacien	pharmacienne	farmacista	farmacista
esthéticien	esthéticienne	estetista	estetista
magistrat	magistrat	magistrato	magistrato

Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
musicien	musicienne	musicista	musicista
chanteur	chanteuse	cantante	cantante
metteur en scène	metteuse en scène	regista	regista
chauffeur de taxi	chauffeuse de taxi	autista di taxi	autista di taxi
astronome	atronome	astronomo	astronoma
météorologue	météorologue	meteorologo	meteorologa
chimiste	chimiste	chimico	chimica
géologue	géologue	geologo	geologa
biologiste	biologiste	biologo	biologa
botaniste	botaniste	botanico	botanica
zoologiste	zoologiste	zoologo	zoologa
architecte	architecte	architetto	architetta
graphiste	graphiste	grafico	grafica
vétérinaire	vétérinaire	veterinario	veterinaria
bibliothécaire	bibliothécaire	bibliotecario	bibliotecaria
sociologue	sociologue	sociologo	sociologa
anthropologue	anthropologue	antropologo	antropologa
peintre	peintre	pittore	pittrice
psychologue	psychologue	psicologo	psicologa
chorégraphe	chorégraphe	coreografo	coreografa
photographe	photographe	fotografo	fotografa
secrétaire	Secrétaire	segretario	segretaria
concierge	concierge	custode d'immobile	custode d'immobile
orfèvre	orfèvre	orafo	orafa
cadre	cadre	dirigente	dirigente
urbaniste	urbaniste	urbanista	urbanista
médecin	médecin	medico	medico
physiothérapeute	physiothérapeute	fisioterapista	fisioterapista
comptable	comptable	contabile	contabile
économiste	économiste	economista	economista
journaliste	journaliste	giornalista	giornalista
interprète	interprète	interprete	interprete
pilote d'avion	Pilote d'avion	pilota di aeromobili	pilota di aeromobili
athlète	athlète	atleta	atleta



Titre FR_masc	Titre FR_fem	IT_masc	IT_fem
conducteur d'autobus et de tramways	conductrice d'autobus et de tramways	conducente di autobus e tram	conducente di autobus e tram

**Tableau 1** : Liste des noms de métiers

En nous appuyant sur les études de TA mentionnées précédemment, nous avons inséré ces noms dans des phrases dont la structure syntaxique est la suivante : article indéfini, nom du métier, adverbe + adjectif, passé composé (auxiliaire + participe passé), complément de lieu. Cette structure fournit de nombreux indices d'accord de genre même dans le cas d'une profession ayant un nom épïcène (comme « secrétaire » en français ou « *commerciant*e » en italien). Pour ce qui concerne les adjectifs, nous avons distingué l'adjectif d'apparence, tels « beau/belle » en français ou les équivalents italiens « *bello/bella* », de l'adjectif de compétence, comme « intelligent/intelligente » en français ou les équivalents italiens « *intelligente/intelligente* ». Nous avons sélectionné les 500 adjectifs les plus utilisés dans les corpus frWac et itWac (M. Baroni *et al.*, 2008) et avons construit une base de données d'un total de 584 phrases (en tableau 2).

		TOT
apparence m/f (fr)	<i>Un &lt;PROFESSION&gt; très &lt;ADJ. APPARENCE&gt; est entré dans la salle</i> <i>Une &lt;PROFESSION&gt; très &lt;ADJ. APPARENCE&gt; est entrée dans la salle</i>	146 phrases (m/f)
apparence m/f (it)	<i>Un &lt;PROFESSION&gt; molto &lt;ADJ. APPARENCE&gt; è entrato nella stanza</i> <i>Una &lt;PROFESSION&gt; molto &lt;ADJ. APPARENCE&gt; è entrata nella stanza</i>	146 phrases (m/f)
compé- tence m/f (fr)	<i>Un &lt;PROFESSION&gt; très &lt;ADJ. COMPETENCE&gt; est entré dans la salle</i> <i>Une &lt;PROFESSION&gt; très &lt;ADJ. COMPETENCE&gt; est entrée dans la salle</i>	146 phrases (m/f)
compé- tence m/f (it)	<i>Un &lt;PROFESSION&gt; molto &lt;ADJ. COMPETENCE&gt; è entrato nella stanza</i> <i>Una &lt;PROFESSION&gt; molto &lt;ADJ. COMPETENCE&gt; è entrata nella stanza</i>	146 phrases (m/f)
		584 phrases

**Tableau 2** : Composition des bases de données.

Ensuite, nous avons fait traduire ces phrases du français vers l'italien ou vice-versa par trois moteurs de TA neuronales : *Google Translate* (GT), *Microsoft Translator* (MT), *DeepL* (DT). Les résultats obtenus sont analysés dans la section suivante.

#### 4. Analyses et résultats

Après avoir fait élaborer les phrases par les moteurs, nous avons analysé manuellement les résultats obtenus, en les annotant selon leur exactitude. Elles ont été jugées « fausses » en ne considérant que le nom de métier, c'est-à-dire que, même si d'autres éléments grammaticaux (accord du participe, article, etc.) ont été correctement traduits au féminin, le nom de métier ne l'était pas. Dans ce cas, la phrase a été annotée comme « incorrecte » puisque le concept a été bien traduit mais que le nom n'a pas été accordé au genre de manière correcte. Dans le cas des noms de métiers épiciènes, l'erreur n'est considérée que lorsque la totalité de la phrase est traduite avec l'accord de l'autre genre par rapport à l'original ou lorsque le nom de métier est traduit par un autre mot.

Comme on peut l'observer dans les tableaux 3 et 4, les taux de traductions correctes des phrases au masculin de la langue source (qu'elle soit en français ou en italien) ont un pourcentage qui va de 93 % à 97% et les taux des traductions correctes pour le féminin oscillent entre 8% et 86%. Plus précisément le féminin ne présente que deux sous-groupes qui ont un taux de réponses correctes assez élevé tandis que tous les autres ont un niveau très bas. Le taux de réponses correctes plus élevé au masculin par rapport au féminin confirme la présence d'un déséquilibre en faveur du masculin, comme l'indiquaient aussi les études mentionnées.

Moteur de TA	masculin			féminin		
	GT	DT	MT	GT	DT	MT
fr>it	96%	96%	97%	33%	86%	11%
it>fr	96%	95%	96%	21%	75%	16%

**Tableau 3 :** Pourcentages totaux des traductions correctes avec l'adjectif d'apparence, différenciés par moteurs de traduction

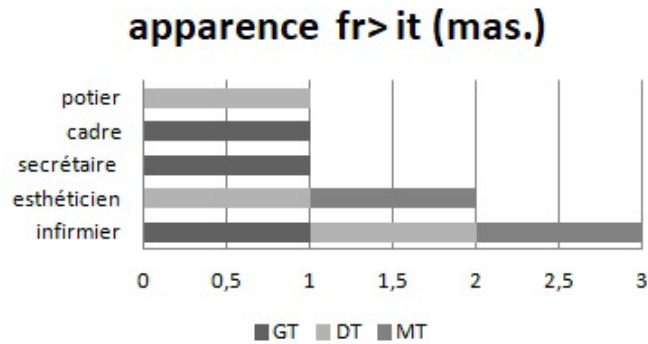
Moteur de TA	masculin			féminin		
	GT	DT	MT	GT	DT	MT
fr>it	96%	96%	93%	11%	9%	7%
it>fr	96%	95%	97%	18%	8%	11%

**Tableau 4 :** Pourcentages totaux des traductions correctes avec l'adjectif de compétence, différenciés par moteurs de traduction

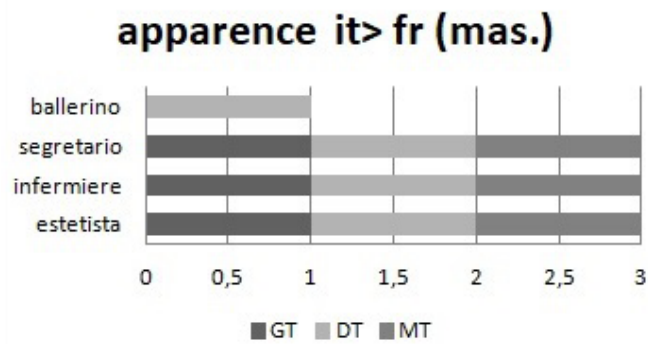
Nous observons encore que le résultat lié au sous-groupe des noms de métiers au féminin montre que la féminisation est moins fréquente dans des phrases avec l'adjectif de compétence (de 7% à 18%) que dans les phrases avec l'adjectif de

l'apparence (11%-86%), ce qui confirme l'hypothèse de départ selon laquelle deux adjectifs appartenant à deux champs sémantiques différents influencent l'exactitude des traductions.

En revanche, la présence d'adjectifs d'apparence et de compétence n'affecte nullement les performances de traduction dans le cas du masculin et les noms des métiers sur lesquels se concentrent les erreurs sont toujours les mêmes : « *infermiere*/infirmier » est traduit à 100% au féminin, suivi de « *estetista*/esthéticien » avec un taux d'erreur de 83 %, de « *segretario*/secrétaire » avec un taux d'erreur de 75%. Certains cas qui semblent isolés comme « potier », « orfèvre », « *ballerino* », « *coreografo* » présentent un taux d'erreur de 8%.



**Tableau 5 :** Pourcentage des traductions erronées du français vers l'italien pour les noms masculins avec l'adjectif d'apparence



**Tableau 6 :** Pourcentage des traductions erronées de l'italien vers le français pour les noms masculins avec l'adjectif d'apparence

### competence fr>it (mas.)

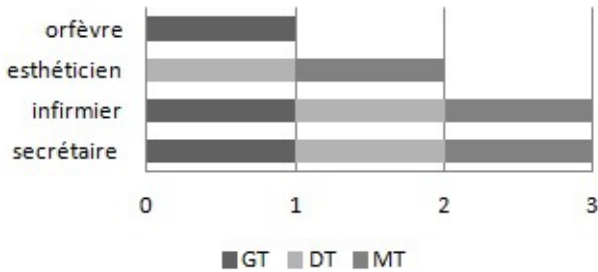
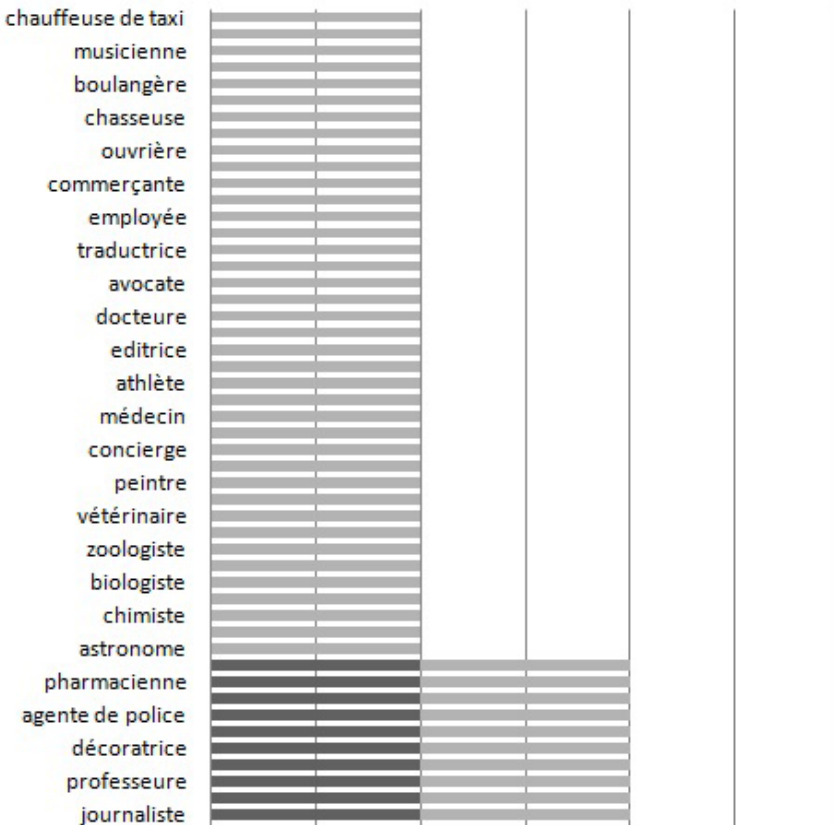
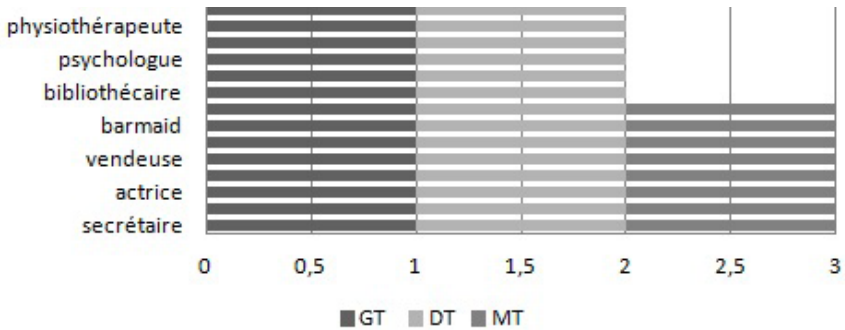


Tableau 7 : Pourcentage des traductions erronées du français vers l’italien pour les noms masculins avec l’adjectif de compétence

### adjectif apparence fr > it (fém.)



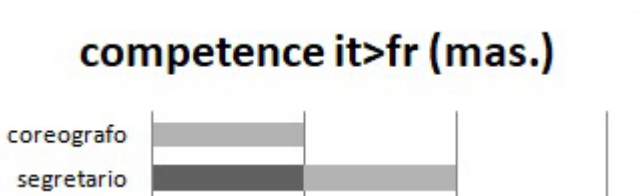


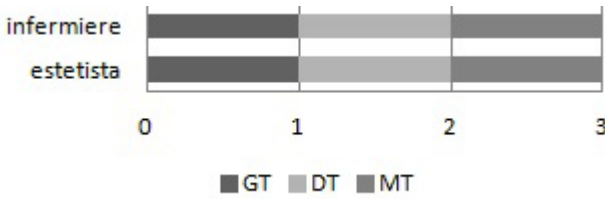
**Tableau 8 :** Pourcentage des traductions erronées de l’italien vers le français pour les noms masculins avec l’adjectif de compétence

En analysant la typologie d’erreurs, nous observons qu’il s’agit là de noms de métiers traduits au féminin. Nous trouvons donc, en correspondance de « Un infirmier très beau est entré dans la salle », la traduction « *Un’infermiera molto bella è entrata nella stanza* ». Dans ces cas, le nom du métier l’emporte sur la structure syntaxique, qui pourtant possède plusieurs éléments comme les accords ou les adjectifs pouvant conduire à une interprétation correcte de la part du moteur de traduction automatique. La traduction est donc forcée au féminin, et, quelle que soit la langue source, les trois moteurs de traductions montrent toujours les mêmes noms de métiers ayant le plus haut taux d’erreurs.

Dans le cas du masculin, nous avons vu que les noms de métiers traduits de manière erronée par les trois moteurs de traduction sont toujours les mêmes, que ce soit de l’italien vers le français ou du français vers l’italien. En comparant les mêmes noms de métiers avec les résultats des traductions des phrases au féminin, nous observons qu’ils donnent lieu aux rares phrases qui sont traduites correctement.

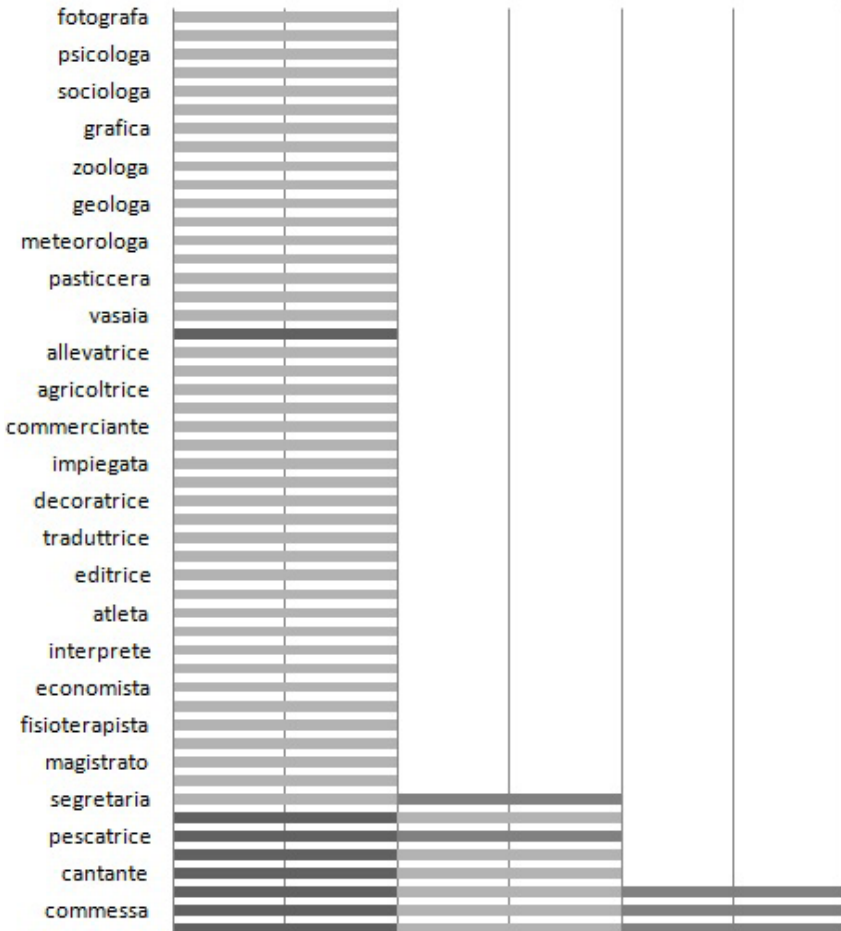
Comme on peut le voir dans les tableaux 9, 10, 11, 12, « *estetista/esthéticienne* », « *segretario/secrétaire* », « *infermiere/infermiera* » ont des taux de réponses correctes très élevés.

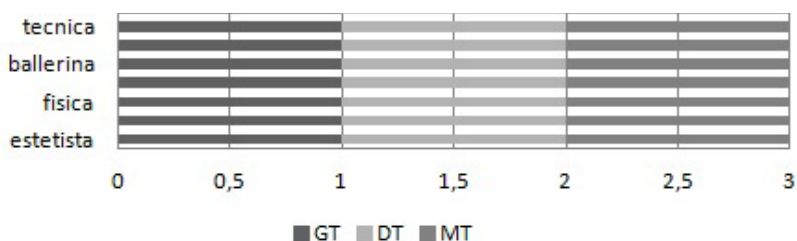




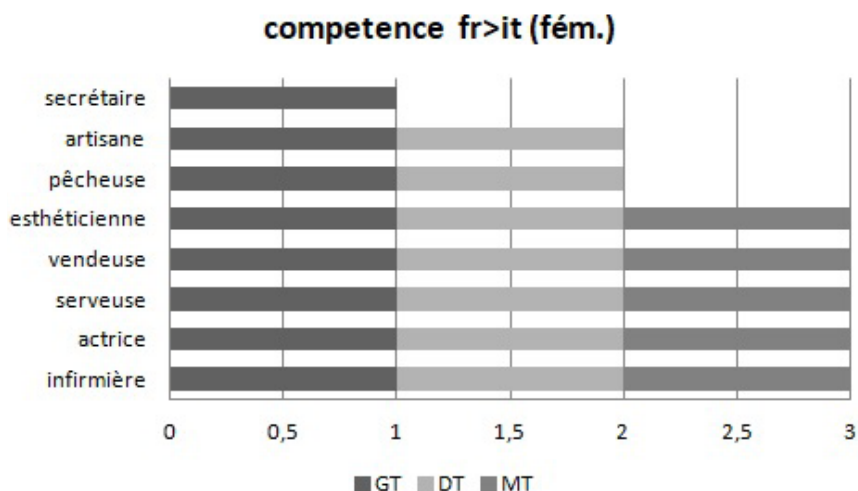
**Tableau 9 :** Pourcentage des traductions correctes du français vers l’italien pour les noms féminins avec l’adjectif d’apparence

**adjectif apparence it > fr (fém.)**

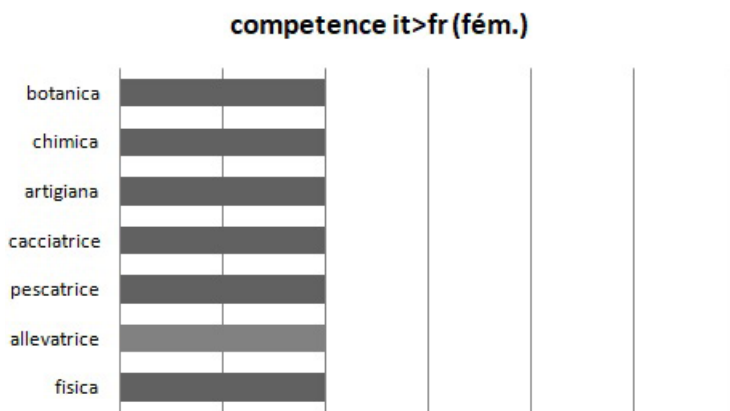


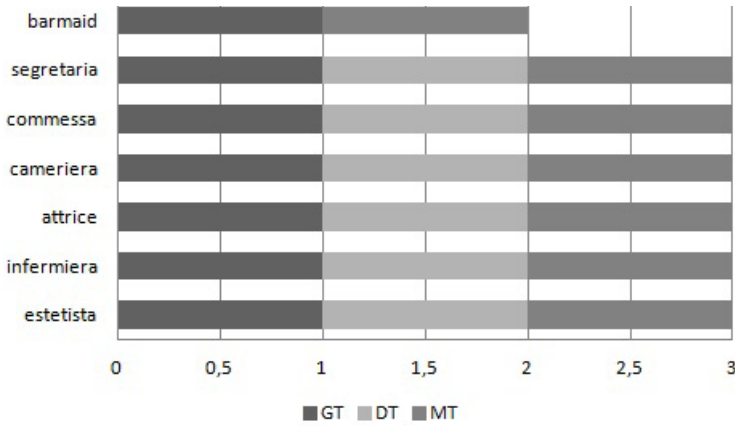


**Tableau 10** : Pourcentage des traductions correctes de l'italien vers le français pour les noms féminins avec l'adjectif d'apparence



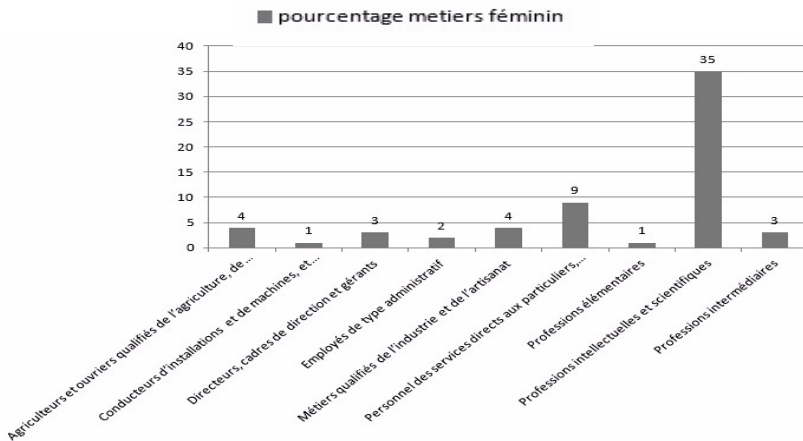
**Tableau 11** : Pourcentage des traductions correctes du français vers l'italien pour les noms féminins avec l'adjectif de compétence





**Tableau 12 :** Pourcentage des traductions correctes de l’italien vers le français pour les noms féminins avec l’adjectif de compétence

Le choix des mots n’est pas aléatoire : en effet, en comparant les noms de métiers traduits de manière incorrecte au masculin et de manière correcte au féminin, quelle que ce soit la langue source, les mots qui ont un taux très élevé d’erreurs ou de réponses correctes sont toujours les mêmes. De plus, on peut regrouper les noms de métiers qui sont traduits correctement au féminin en utilisant les 10 catégories de la taxonomie du projet ESCO qui, comme nous l’avons dit auparavant, s’inspire de la classification ISCO-08 : nous avons pu ainsi identifier que le secteur ayant une couverture majeure est celui des « Professions intellectuelles et scientifiques » (tableau 13).



**Tableau 13 :** Couverture des secteurs professionnels des noms de métiers au féminin selon les catégories ESCO qui sont traduits correctement



## Conclusions

L'hypothèse d'une influence des adjectifs est partiellement confirmée : les taux de résultats corrects sont clairement déséquilibrés en faveur des noms au masculin pour lesquelles la présence des adjectifs de compétence et d'apparence ne semblent jouer aucun rôle lors de la traduction. En revanche, pour les noms au féminin le taux d'erreur est bien plus élevé et le rôle des adjectifs d'apparence et de compétence semble jouer un rôle déterminant. Cela impliquerait que dans le cas d'une performance déjà basse - due aussi à une biodiversité insuffisante des données dans le corpus d'entraînement de l'algorithme d'intelligence artificielle - la présence d'adjectifs liés à l'apparence et à la compétence rend la représentation encore moins stable en raison de l'existence d'un stéréotype.

Des études futures en ce domaine pourraient contribuer à approfondir cette question, en élargissant la liste des adjectifs pour couvrir un spectre sémantique plus ample.

## Bibliographie

- Baeza-Yates, R. 2018. « Bias on the web ». *Communications of the ACM*, n° 61 (6), p. 54-61.
- Baroni M. et al. 2008. *The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora*. Pays-Bas : Kluwer Academic Publishers.
- Berger, A.-E. 2019, « Le genre de la traduction : introduction ». *Revue De Genre*, n° 5, p. I-XII.
- Cardon, D. et al. 2018. « La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle ». *Réseaux*, n° 211 (5), p. 173-220.
- Costa-Jussà, M. R. 2019. « An analysis of gender bias studies in natural language processing ». *Nature Machine Intelligence*, n° 1(11), p. 495-496.
- Criado-Perez, C. 2019. *Invisible Women: Exposing Data Bias in a World Designed for Men*. Londres : Chatto & Windus.
- Eco, U. 1964, *Apocalittici e integrati*. Milan : Bompiani.
- Eco, U. 2003, *Dire presque la même chose*. Traduction de M Bouzaher. Paris : Grasset [or. *Dire quasi la stessa cosa*, Milan : Bompiani].
- Federici, E. (dir.) 2011. *Translating Gender*. Berne : Peter Lang.
- Greco L. 2014. « Les recherches linguistiques sur le genre : un état de l'art ». *Langage et société*, n° 148 (2), p.11-29.
- Jones, J. J. et al. 2020. « Stereotypical Gender Associations in Language Have Decreased Over Time ». *Sociological Science*, n° 7, p. 1-35.
- Loffler-Laurian, A-M. 1996, *La traduction automatique*, Presses universitaires du Septentrion : Villeneuve d'Ascq.
- Monti, J. 2017. « Questioni di genere in traduzione automatica ». In : De Meo, A., Di Pace, L., Manco, A. *Al femminile, scritti linguistici in onore di Cristina Vallini*. Florence : Franco Cesati editore, p. 411-431.
- Morel, H. 2020 « Au-delà de la langue, les femmes invisibles ». *Cahiers d'histoire. Revue d'histoire critique*, p. 147-148.
- Nida, E., Taber, C.R. 1969. *The Theory and Practice of Translation*. Leiden : E.J. Brill.

- Prates, M. *et al.* 2019. « Assessing Gender Bias in Machine Translation - A Case Study with Google Translate ». *Neural Computing and Application*. Londres : Springer. [En ligne] : <https://arxiv.org/pdf/1809.02208.pdf> [consulté le 07 février 2021].
- Nirenburg, S., *et al.* 2003. *Readings in Machine Translation*. Cambridge, Massachussets : The MIT Press.
- Paresh, D. 2018. « Fearful of bias, Google blocks gender-based pronouns from new AI tool ». [En ligne] : <https://news.trust.org/item/20181127055628-022s5> [consulté le 07 février 2021].
- Stanovsky, G. *et al.* 2019. « Evaluating Gender Bias in Machine Translation ». *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Firenze.
- Viennot, E. 2014. *Non, le masculin ne l'emporte pas sur le féminin*. Paris : Editions iXe.
- Von Flotow, L. 2011. *Gender in Translation*. Amsterdam-Philadelphia : John Benjamins Publishing Company.
- Weaver, W. 1949. *Memorandum*, New York : The Rockefeller Foundation.
- Wilks, Y. 2009. *Machine Translation, Its Scope and Limits*. Londres : Springer.

### Sitographie<sup>8</sup>

DeepL

<https://www.deepl.com/translator>

ESCO - European Skills Competences and Occupations

<https://ec.europa.eu/esco/portal>

Google Translate (GT)

<https://translate.google.com>

Microsoft Translator (MT) : <https://www.microsoft.com/it-it/translator>

### Notes

1. Un des principaux supporteurs de cette approche et aussi fondateur de la moderne « Science de la traduction » fut le linguiste américain Eugene Nida (1969).
2. Voir, entre autres, Luise Von Flotow, L. (2011) et Anne-Emmanuelle Berger (2019).
3. Le langage épïcène ou rédaction épïcène vise à éviter le sexisme du langage, grâce à l'usage de techniques de reformulation lexicale, syntaxique, l'attention à l'accord morphologique et à la typographie. Le langage épïcène permet d'éviter la discrimination de genre et rejette l'idée d'un androcentrisme marqué par l'utilisation du masculin comme universel générique. À ce sujet, voir, entre autres, Éliane Viennot (2014) et Luca Greco (2014).
4. <https://blog.google/products/translate/reducing-gender-bias-google-translate>
5. <https://ai.googleblog.com/2018/12/providing-gender-specific-translations.html> Voir aussi Dave Paresh (2018).
6. ESCO est un projet de la Commission Européenne qui organise une ontologie multilingue en trois piliers : professions, compétences et certifications.
7. À chaque métier équivaut un code ISCO 08 qui est superposable à la catégorie ESCO. Voir : <https://ec.europa.eu/esco/portal/occupation?resetLanguage=true&newLanguage=fr>
8. Date de dernière consultation des sites : 7 février 2021.