universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Gender Bias in Machine Translation – The Effect of Adjectives on the Translation of Professions“

verfasst von / submitted by

Manuel Lardelli

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 070 348 331
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Translation Italienisch Deutsch
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Dagmar Gromann, BSc

Acknowledgments

A special thanks to Professor Gromann for helping me out with such patience, dedication and passion. She was the perfect supervisor with whom to conclude the amazing experience at the University of Vienna.

Another special thanks to Professor Krause without whom I wouldn't have learned so much. I wish there were more Professors as passionate as she is, and I'll terribly miss her translation courses.

To all the amazing people, be they family, friends or lovers, be they still with me or not, with whom I shared laughter, tears, unforgettable and also forgettable moments in the last two years (and not only). Thank you.

To Vienna, my second home and the city that has changed my life.

Table of Contents

1. Introduction *1*

1.1 Research Question and Hypothesis 2

1.2 Research Motivation and Objectives 2

2. Theoretical Background *5*

2.1 Language and Perception 5
2.1.1 Gender and Perception 8

2.2 Gender Bias 10
2.2.1 Gender Bias in Language 12
2.2.2 Gender Bias characterizing the Language of Different Fields 14
2.2.3 Gender Bias in NLP – The case of the Word Embeddings 15
2.2.4 Other Biases in NLP and Word Embeddings – A brief Overview 18

2.3 Machine Translation 20
2.3.1 Machine Translation – A Brief Historical Overview 20
2.3.2 MT Approaches – An Overview 22
2.3.3 Neural Networks and Neural Machine Translation 23
2.3.4 Gender Bias in Machine Translation 25
2.3.5 Available User-Friendly NMT Systems 27

3. Related Work *31*

4. Research Methodology *34*

4.1 System Selection 34

4.2 Languages Involved 35

4.3 Data for this Study 36
4.3.1 Creation of the first two corpora 37
4.3.2 Corpus Translation and Cleaning 39
4.3.3 Third Corpus Creation and Translation 41

4.4 Evaluation Process 42

4.5 Research Validity: Advantages and Disadvantages of the Selected Methodology 46

5. Results *48*

5.1 Gender Bias in Google Translate 48

5.2 Gender Bias in DeepL 61

5.3 Results Compared 73

6. Discussion *81*

6.1 Gender Bias in Google Translate and DeepL 81

6.2 The Influence of Adjectives on Translation Outputs 83

6.3 Reflection on the Research Process 87

7. Conclusion *89*

8. Bibliography *91*
10. Appendix *102*
10.1 Sentences_without_adjectives_GoogleTranslate 102
10.2 Sentences_without_adjectives_DeepL 106
10.3 Sentences_with_Adjectives_GoogleTranslate 110
10.4 Sentences_with_adjectives_DeepL 122
10.5 Sentences_with_wider_context_GoogleTranslate 141
10.6 Sentences_with_wider_context_DeepL 143
Abstract (English) *146*
Abstract (Deutsch) *146*

List of Tables

Table 1: Attributes related to agency and communality used to describe respectively males and females (Heilman 2012: 115) 11
Table 2: Representation Bias in different NLP applications based on Crawford (2017) (Sun et al. 2019) 16
Table 3: Debiasing methods dividing depending on the type (Sun et al. 2019) 17
Table 4: Randomly chosen subjects 38
Table 5: Verb sentences indicating a profession 38
Table 6: Male and female-dominated professions 38
Table 7: Adjectives related to gender (both male and female) roles 39
Table 8: Sizes of the first two corpora translated with Google Translate and DeepL after cleaning 40
Table 9: Example of how the analysis of the second corpus was conducted. Results about the translation outputs for lawyer by DeepL 44
Table 10: Overview of the subcorpora content and size 45
Table 11: Data on the cleaning of the corpus containing adjectives translated by Google Translate 50
Table 12: General results on output correctness (Google Translate) 52
Table 13: Output correctness for subject's gender (Google Translate) 52
Table 14: Bias with Liam in the corpora (Google Translate) 53
Table 15: Output correctness for typology of occupation (Google Translate) 54
Table 16: Output correctness for subject's gender and typology of occupation (Google Translate) 55
Table 17: Output correctness for female subjects, typology of profession and typology of adjective (Google Translate) 56
Table 18: Output correctness for male subjects, typology of profession and typology of adjective (Google Translate) 57
Table 19: Frequencies of correctly translated male-dominated professions in the female form (Google Translate) 58
Table 20: Results from the analysis of the third corpus by Google Translate 61
Table 21: Data on the cleaning of the second corpus translated by DeepL 62
Table 22: General results on the output correctness (DeepL) 63
Table 23: Output correctness for subject's gender (DeepL) 64
Table 24: Output correctness for typology of occupation (DeepL) 65
Table 25: Output correctness for subject's gender and typology of occupation (DeepL) 66
Table 26: Output correctness for female subjects, typology of profession and typology of adjective (DeepL) 67
Table 27: Output correctness for male subjects, typology of profession and typology of adjective (DeepL) 67
Table 28: Frequencies of correctly translated male-dominated professions in the female form (DeepL) 68
Table 29: Frequencies of mistranslations of female-dominated professions that did not generally present bias against men 71
Table 30: Results from the analysis of the third corpus by DeepL 72
Table 31: General results for output correctness - comparison between Google Translate and DeepL 73
Table 32: Output correctness for female subjects – comparison between Google Translate and DeepL 74
Table 33: Output correctness for male subjects – comparison between Google Translate and DeepL. 74

Table 34: Output correctness for male-dominated professions - comparison between Google Translate and DeepL 75
Table 35: Output correctness for female-dominated professions - comparison between Google Translate and DeepL 75
Table 36: Output correctness for female subjects and male-dominated professions - comparison between Google Translate and DeepL 76
Table 37: Output correctness for female subjects and female-dominated - comparison between Google Translate and DeepL 76
Table 38: Output correctness for male subjects and male-dominated - comparison between Google Translate and DeepL 77
Table 39: Output correctness for male subjects and female-dominated - comparison between Google Translate and DeepL 77
Table 40: Output correctness for female subjects, male-dominated professions and typology of adjectives - comparison between Google Translate and DeepL 78
Table 41: Output correctness for female subjects, female-dominated professions and typology of adjective form - comparison between Google Translate and DeepL 78
Table 42: Output correctness for male subjects, male-dominated professions and typology of adjectives - comparison between Google Translate and DeepL 79
Table 43: Output correctness for male subjects, female-dominated professions and typology of adjectives - comparison between Google Translate and DeepL 79
Table 44: Adjective co-occurrence in the first subcorpus - comparison between Google Translate and DeepL 80
Table 45: Adjective co-occurrence in the second subcorpus - comparison between Google Translate and DeepL 80

List of Figures

Figure 1: The visible spectrum (Wikimedia Commons 2006) 7
Figure 2: Biases found in different vectors for word representation (Speer 2017) 19
Figure 3: A neural network architecture (Mueller & Massaron 2018: 159) 24
Figure 4: Example of one hot encodings of the words Rome, Paris, Italy and France (Bonanzini 2017) 25
Figure 5: Example of gender bias against women in Google Translate (Google Translate) 26
Figure 6: Example of gender bias against women in DeepL (DeepL a) 26
Figure 7: Example of gender bias against men in Google Translate (Google Translate) 26
Figure 8: DeepL outputs quality in comparison to other translation systems (DeepL n.d.b) 28
Figure 9: Translating with SYSTRAN Translate (SYSTRAN Translate) 29
Figure 10: Example of gender bias in the language pair Hungarian – English (Prates et al. 2018: 4) .. 31
Figure 11: Example of gender bias in the language pair English – Spanish (Stanovsky et al. 2019) 33
Figure 12: Comparisons among translators regarding output quality (DeepL n.d.b) 35
Figure 13: Example from the first corpus 38
Figure 14: Second corpus translated with Google Translate. The adjective conceited is sometimes wrongly rendered as presunta, namely supposed. 40
Figure 15: Example of the use of Sketch engine during the study (Sketch Engine) 43

Figure 16: Frequencies (number) of adjectives occurring when male-dominated professions were translated in the female form (Google Translate) 59
Figure 17: Frequencies (%) of adjectives occurring when male-dominated professions were translated in the female form (Google Translate) 59
Figure 18: Frequencies of adjectives occurring when professions presenting bias against men were correctly translated in the male form (Google Translate) 60
Figure 19: Frequencies of adjective (numbers) when male-dominated professions were correctly translated in the female form (DeepL) 68
Figure 20: Frequencies of adjective (%) when male-dominated professions where correctly translated in the female form (DeepL) 69
Figure 21: Frequencies of adjective (numbers) when professions presenting bias towards women where correctly translated in the male form (DeepL) 70
Figure 22: Frequencies of adjective (%) when professions presenting bias towards women where correctly translated in the male form (DeepL) 70
Figure 23: Results from the analysis of the third DeepL subcorpus 72

1. Introduction

The topic of the current master's thesis is gender bias in machine translation (MT) and it refers to the issue of how MT's outputs can be discriminatory against different genders. This research field relates to two broader subject areas, namely gender studies and Artificial Intelligence (AI). Gender studies is an academic discipline that questions gender as social constructs and studies relationships, inequalities and differences occurring within them (Pilcher & Whelehan 2004: xif). There is instead no agreement on a general definition of Artificial Intelligence (AI) (Nilson 2009: 13). However, Nilson (2009: 13) defines it as "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment." In order to identify better the research issue, some terminology related to the subject area must be explained.

First of all, gender bias relates to unfair differences against people based on their gender (Macmillan Dictionary n.d.). The term Machine Translation (MT) refers to a computer software used in order to translate a text from one source language (SL) into a target language (TL) (Hutchins 1994: 51). MT is powered by Natural Language Processing (NLP), namely a research area that has the purpose of making computers understand language texts or speech so as to carry out different tasks (Chowdhury 2003: 51). In particular, some NLP applications use neural networks and are trained on big collections of texts, thus reflecting different biases, for example gender and occupational bias, present in them (Lu et al. 2018). Such notions, which are examined in detail in the second chapter of the current work, lead to the crux of the matter.

If MT learns how to translate on the basis of biased corpora, then its outputs are discriminatory too. As a matter of fact, in the last few years, research (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019) proved that automated translation can be discriminatory against women as well as other genders, e.g. non-binary – but the latter are not taken into consideration in the present master's thesis. Indeed, MT tools seem to generally render male-dominated jobs in the male form even though the subject of the sentences to be translated is a female (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019).

Having described the topic of this thesis as well as the wider context to which it relates and the research problem, an overview of the structure of the present work is provided. The following subchapters introduce the research question and the hypothesis as well as the motivation and objectives for writing this thesis. The second chapter represents the theoretical background, i.e. the basis on which the study was carried out. As a consequence, the main concepts presented in this introduction are thoroughly explained. The third chapter is dedicated

to the current work conducted in this very field, whilst the fourth deals with the selected methodology. Results from the study are presented in the fifth and discussed in the sixth. The seventh chapter serves as the conclusion.

1.1 Research Question and Hypothesis

The current work leans on previous research conducted by Prates et al. (2018), Cho et al. (2019) and Stanovsky et al. (2019). The focus is on the translation from English into Italian of simple and short sentences containing gender-dominated occupations, i.e. professions carried out mostly by (1) men or (2) women only and hence associated with that specific gender, as well as adjectives related to gender roles and/or stereotypes. Such terms are explained in the second chapter of this thesis. The purpose of the study is to find whether the above-mentioned adjectives can affect the MT's outputs of professions. As a consequence, the research question can be phrased as follows:

RQ1: How do adjectives collocated with nouns affect the machine translation of professions?

Hence, the hypothesis to be evaluated and analysed is:

H1: In general, MT correctly renders gender-dominated professions in the female or male form if the gender stereotype related to the adjectives preceding such job titles is consistent with the gender of the subject.

1.2 Research Motivation and Objectives

There are several reasons for writing this master's thesis. In fact, at least three different motivations can be provided to support the validity of such a project. Indeed, this research is conducted (1) to show the great extent to which MT is biased, (2) demonstrating that such a phenomenon is surely due to the bias present in human-generated corpora but it is also amplified by MT and finally (3) it is carried out to underline the importance of using a gender fair language. These motivations are further illustrated in detail.

First of all, it is important for researchers to be aware that MT is still extremely biased. In 2018 Google claimed to have reduced bias in its own translation system (Kuczmarski 2018). According to the company, Google Translate algorithms were improved so as to render both

the female and the male version whether of gender-neutral sentences or of single words (Kuczmarski 2018). Therefore, the purpose of the proposed thesis is to show that the quality of MT must be enhanced because the outputs produced are still extremely sexist.

In addition, I would personally like to demonstrate that MT not only reproduces bias present in the training corpora, but it also takes it to extremes. Hence, results of this study should highlight such a phenomenon so as to confirm once again the importance of humans in the translation process.

Since gender bias is present in our language, this topic should ultimately raise the possibility of reflection on how gender stereotypes affect society and consequently lead to consider the use of a gender fair language as a matter of crucial importance. Bem and Bem (1973), for instance, demonstrated that women are more likely to apply for a stereotypically male job, if the advert for this is written in a gender inclusive language. Besides, research on how gender affects the way we think is provided in the course of the present work. For the moment, it might be meaningful to quote Santacreu-Vasaut et al. (2013). The scholars collected evidence of how distinctions in terms of gender could actually have an impact on our society. Indeed, they showed a correlation between countries in which distinction in terms of female/male are stronger and the adoption of a gender quota with a consequently higher participation of women.

For the objectives of the current master's thesis, the study of gender bias in MT was conducted through the translation of three different datasets of sentences from English into Italian. More specifically, the first dataset contained sentences including professions but without adjectives related to gender stereotypes whereas the second and the third included such attributes. By collecting statistics on the correctness of the automated translation of gender-dominated professions, it was possible to analyse to what extent the aforementioned type of adjectives affected the machine translation of professions. This means that data collected during the study were quantified, evaluated and compared in order to simply describe the used sample. As a consequence, tendencies regarding the base assumptions of the current work are presented in detail in the sixth chapter by discussing the results of the study.

The results demonstrate the urge for research in this very field by demonstrating that the collected data is extremely biased. Therefore, having verified that MT can produce sexist outputs when translating words belonging to a certain field (e.g. professions), scholars could undertake research so as to find whether other domains are biased and consequently develop and/or improve debiasing techniques. Hence, results could be interesting for people working or doing research in the area of machine translation, such as scholars or programmers.

Furthermore, the results could also be important for translators and/or translation companies, as they must take into account that translation outputs can present gender bias when MT is employed in the translation process. I think that such a study could also be fascinating for people fond of themes like gender, gender fair language and sexism as these were the reasons leading me to propose this topic for my master's thesis.

2. Theoretical Background

This chapter provides the theoretical background needed in order to understand the research question and the resulting hypothesis as well as the main ideas on which the current work is based. To sum up, the intimate relationship between language and perception is addressed, describing how the idiom one speaks can actually shape how s/he perceives reality and vice versa. This is very important, because it leads to a central point. Being gender expressed in language, idioms are a vehicle for gender bias which is consequently to be found in Natural Language Processing (NLP) applications, such as Machine Translation (MT). As a consequence, gender bias is explained by referring also to gender stereotypes and gender roles, which lie at its origin. Furthermore, the phenomenon of gender bias, and bias in general, in NLP systems is analysed. Afterwards, a brief historical overview of the creation and development of MT as well as the description of the current state-of-the-art tools are provided. Finally, gender bias in MT and some debiasing methods are discussed.

2.1 Language and Perception

In this chapter the close relationship between language and reality is revealed. The idea that the use of language can affect the way we think and consequently our reality as well as vice versa is a central concept for this thesis. Indeed, we reproduce gender bias and sexism present in our society while speaking (Menegatti & Rubini 2017: 1, Weatherall 2002: 5). As a result, Natural Language Processing (NLP) applications that are trained with biased corpora end up being biased too (Lu et al. 2018). Before jumping to conclusions, studies on the mutual influence between language and thought are presented.

Thousands of languages are spoken around the world. As Boroditsky (2011: 64) underlines, this leads people to think differently in order to express the same event. For instance, some languages use different pronouns for *we*, one meaning *you and I*, another meaning *you, someone else and I* and finally the last one meaning *someone else and I but not you* (Deutscher 2010: 15). Or again, some idioms have tenses specifying whether a past event took place in an immediate, recent or a distant past, whereas others do not (Boroditsky 2011: 64). Examples of linguistic differences were collected throughout the years and this made scholars ask themselves whether concepts of time, space and matter are derived from experience and

therefore the same for everyone or whether linguistic structures can partially exert some sort of influence on it (Whorf 1939: 138).

This so-called Whorf-effect, namely that the mother tongue language affects the way a person thinks and perceives reality, has been debated for a long time (Zhu 2015: 1). Scholars seemed to agree that such an effect was not to be observed, advancing the theory that all languages share the same universal underlying concepts (Boroditsky 2011: 63, Deutscher 2010: 6). However, this view has changed in recent years since different studies demonstrated how certain areas of language, such as time, space, colours, gender and others, actually affect the way we think (Boroditsky 2001, 2003, 2011, Casasanto 2008, Deutscher 2010, Haviland & Levinson 1994, Philipps & Boroditsky 2003).

Indeed, Deutscher (2010: 9f) explains that language is the result of cultural conventions. Concrete concepts depend on nature and are the same for everyone whilst labels are used to name them and vary from one language to another (Deutscher 2010: 13). Thus, both abstract and some concrete concepts depend on cultural conventions (Deutscher 2010: 13f) (see the aforementioned examples relating to differences in pronouns between idioms or see the area of colours). In order to better understand how language can affect thought, the results of the studies mentioned must be reported.

Boroditsky's (2001) starting point for research was that languages do often refer to time by using spatial metaphors. However, these can vary across idioms. Indeed, English speakers usually refer to time by using horizontal metaphors, such as *move a meeting forward* or *push a deadline back* (Boroditsky 2001: 4). On the other hand, in Mandarin vertical metaphors are preferred, i.e. an early event is referred to as *up* whilst a later one as *down* (Boroditsky 2001: 5). The researcher claims that the use of spatial metaphors when describing time can lead speakers to think of it as actually being horizontal (Boroditsky 2001: 7). This was confirmed since in the study she found that (1) Mandarin speakers usually thought about time vertically, even when using English; (2) Mandarin-English bilinguals generally think about time vertically as well, but the extent varies according to the age at which they learned English; (3) if English speakers are taught to refer to time by using vertical metaphors, then they also start to think about it in such terms. Boroditsky's conclusion is that language exerts a profound influence in the shaping of thought (Boroditsky 2001: 20). In addition, Casasanto (2008: 70f) underlines that English and Greeks, for example, describe quantities of times differently. The former use expressions referring to distances (e.g. *a long time*) whereas the latter to amounts (e.g. *much time*). Casasanto found that people actually do represent time mentally in the same way they

talk about it in their mother tongue, concluding that language affect the non-linguistic, cognitive representation of time (Casasanto 2008: 75).

As regards space, Boroditsky (2011: 64) notes that Aboriginal languages do not refer to object positions with words like *left* or *right*. Instead, they use cardinal directions, thus improving their orientation ability. Indeed, it turns out that populations speaking such languages can orient themselves much better than scientists thought would be possible. These findings are not surprising as other scholars like Haviland and Levinson (1994) studied the astonishing ability of Mayan speakers to understand their position just by using absolute directions.

Colour terms also change across languages. The problem is that colours are actually a continuum, one gradually blurring into the others, as can be seen in Figure 1. However, language sets boundaries and they are not the same in every idiom (Deutscher 2010: 17). For instance, in the Philippines there is no distinction between yellow and green, whereas in Russian there are two different terms for *light blue* and *dark blue* (Deutscher 2010: 63f). Witthoft et al. (2003) discovered that the ability to discriminate colours is strictly related to the distinction made among them within a language. For example, in their study they found that Russians were quicker than English speakers at distinguishing different shades of blue.

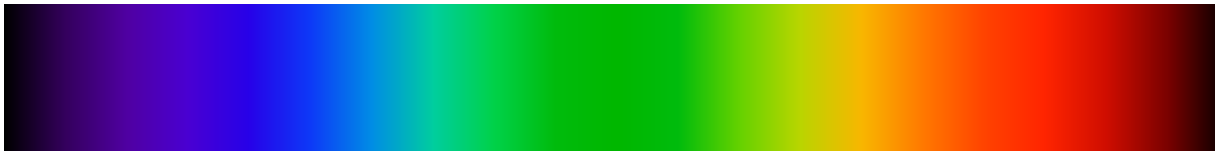


Figure 1: The visible spectrum (Wikimedia Commons 2006)

The aforementioned studies demonstrate how one's mother tongue deeply affects the way s/he sees reality. Indeed, even though the world is just one, this can be thought of and described in many different ways, may it be because we conceptualise time and space in a whole different way or just because we use different words to describe the very same concept. Furthermore, languages also refer differently to gender (Stahlberg et al. 2007: 164f) and several studies demonstrated how this language area has an impact on the way people think as well (Boroditsky et al. 2003, Philipps & Boroditsky 2003, Prewitt-Freilino et al. 2012, Santacreu-Vasaut et al. 2013, Stahlberg et al. 2007). This topic will be dealt with in a specific subchapter because gender – and more specifically gender bias – is, together with MT, the focus of the current master's thesis and therefore requires particular attention.

2.1.1 Gender and Perception

The starting point for understanding how gender affects thought is the distinction among (1) grammatical gender, (2) natural gender and (3) genderless languages introduced by Stahlberg et al. (2007: 164f). This distinction is based on how languages relate to gender. In (1) every noun has a gender, i.e. also inanimate objects. As a consequence, nouns and pronouns as well as articles and adjectives are gendered. An example for this category is the Italian language, which was used in the current study to answer the research question. In (2) most nouns do not have a gender, such as in the English language. Therefore, only pronouns have different forms referring to male or female and/or eventually neutral. In (3), finally, neither nouns nor pronouns express gender. However, this can be achieved by adding an additional expression marking the noun as male or female. An idiom belonging to this group is Turkish. Having made this distinction, an overview of some studies on how gender affects thought can be provided.

Philipps and Boroditsky (2003) focused on German and Spanish, i.e. languages with grammatical gender. Their starting point was that some grammatical distinctions such as plural forms are the reflection of tangible differences. As a consequence, when children learn a language presenting a grammatical gender system, they could actually believe that this grammatical property reflects significant differences between objects (Philipps & Boroditsky 2003: 929) To find this, the scholars carried out different experiments. For instance, they tested the memory of a group of Spanish and English speakers to see if gender exerts some sort of influence. Objects were given a masculine or a feminine proper name. The gender name was actually consistent with the grammatical gender only half of the time. They found that the participants tended to remember the object's name when their grammatical gender and the gender of their proper name corresponded (e.g. Germans were more likely to remember Patrick as a name for an apple, since apple is masculine in German) (Philipps & Boroditsky 2003: 930, Boroditsky et al. 2003: 68). Another experiment consisted in describing objects that have opposite genders in the two languages. It resulted that participants used different adjectives according to the gender of the object. The word *key* is masculine in German and feminine in Spanish. German speakers provided adjectives such as *hard* or *heavy*, whereas Spanish speakers preferred *beautiful* or *elegant*. This tendency was observed throughout the whole experiment, confirming that if objects are provided with gender, then people tend to mentally represent that object by underlying its masculine or feminine qualities (Boroditsky et al. 2003: 65).

Prewitt-Freilino et al. (2012) compared gender equality across 134 countries. Their purpose was to understand whether there is a correlation between gender equality and the typology of language spoken, namely with grammatical gender, with natural gender or genderless. Based on the results provided by Boroditsky et al. (2003), the scholars wondered if the influence of gender was to be extended to the relationships between women and men as well (Prewitt-Freilino et al. 2012: 269). So, the scholars compiled a list of 134 countries and compared the levels of gender equality by using the gender system of their official languages as metrics. To sum up, they found that (1) higher levels of gender equality are to be found in countries where a natural gender language is spoken; (2) countries where a gendered language is spoken show instead lower levels of gender equality and (3) countries with genderless languages are to be found in the middle of this ranking. If language can shape thought and reality, such results might be explained because natural gender languages are more likely to promote a gender-inclusive language. Indeed, they can easily increase the visibility of women as their pronouns and nouns can have gender-symmetrical forms. Genderless languages, instead, are not so suitable as the use of gender neutral forms do not automatically cause bias to disappear. Furthermore, as opposed to idioms with grammatical gender, natural gender ones do not present gendered structures which could limit a symmetrical revision of language (Prewitt-Freilino et al. 2012: 279). Results also show that where a more gender-inclusive language is spoken, higher levels of gender equality are to be found. Other similar studies were also carried out.

Santacreu-Vasaut et al. (2013) created an index in order to measure how strong and common female/male distinctions in languages using grammatical gender are. On this basis, the scholars compiled a list of 84 countries in order to assess the effect of female/male distinctions on the political participation of women in such countries. Their findings were that (1) the more a language emphasises the aforementioned distinctions, the more the country in which this idiom is spoken is likely to adopt a gender quota in politics; (2) this very same countries show also a larger increase of women participating in the political agenda after the adoption of the quota. The explanation for (1) is that distinction in gender could actually trigger gender stereotypes, thus representing barriers for women who would like to actively participate in politics.

2.2 Gender Bias

As anticipated in the introduction, gender bias describes discriminations against people because of their gender (Macmillan Dictionary n.d.). Hence, it is important to make a distinction between sex and gender. Sex is actually biological since it relates those biological and physical features like sexual and reproductive organs which distinguish females from males (Halilovic et al. 2017: 18). Instead, gender is a social construct and therefore relates to a set of features, roles, and behaviours which define males and females in a particular socio-cultural context. As society and culture change during time, so does gender (Halilovic et al. 2017: 18). Moreover, sex categories are three, namely male, female and intersex (Halilovic et al. 2017: 18) whereas gender categories are at least 64 (Mere Abrams 2019), meaning that sex and gender do not always correspond. With regard to the proposed topic, however, two aspects are worth examining: gender roles and gender stereotypes.

Gender roles are those social norms which establish the considered appropriate behaviour of individuals on the basis of their actual or perceived sex (Cultural Anthropology n.d.). Such roles can deeply affect women's and men's lives. An example connected to the proposed thesis is in the field of work. Women are generally related to family care whilst men to production, i.e. earning for their family (Halilovic et al. 2017: 19). Nevertheless, gender roles change across time and societies (Halilovic et al. 2017: 19).

As already mentioned, gender refers to attributes as well. This aspect in particular is important in order to understand the research question. Indeed, the purpose of the current thesis is to assess to what extent adjectives can affect the MT's outputs of professions. However, specific attributes are normally used in order to distinguish women and men, hence resulting in behavioural stereotypes (Halilovic et al. 2017: 19). Indeed, gender stereotypes are "generalizations about the attributes of men and women" (Heilman 2012: 114). These include their main features as well as their inclination for some specific roles or responsibilities (Halilovic et al. 2017: 22). Thus, gender stereotypes create expectations about what men and women are like (descriptive) as well as what they should be like (prescriptive) (Heilmann 2012: 114, Menegatti & Rubini 2017: 3). As Table 1 shows, males are described with expressions or words referring to the field of agency whereas women are related to the field of communality (Heilman 2012: 114).

Table 1: Attributes related to agency and communality used to describe respectively males and females (Heilman 2012: 115)

AGENCY	COMMUNALITY
Achievement-orientation (competent, ambitious, task-focused)	Concern for others (kind, caring, considerate)
Inclination to take charge (assertive, dominant, forceful)	Affiliative tendencies (warm, friendly, collaborative)
Autonomy (independent, self-reliant, decisive)	Deference (obedient, respectful, self-effacing)
Rationality (analytical, logical, objective)	Emotional sensitivity (perceptive, intuitive, understanding)

Besides, in her paper Heilman (2012) points out that gender stereotypes can affect one's career since the decision making in the hiring process can be biased because of them.

Now that the wider framework is clear, gender bias can be discussed in detail. First of all, Schafran (1993: 397f) identifies three components of gender bias. (1) "Stereotyped thinking about the nature and roles of women and men"; (2) "how society values women and men, including what is perceived as women's work" and (3) "myths and misconceptions about the social and economic realities of women's and men's lives". Consequences can be different. For instance, women are not considered suitable to hold male-dominated positions as they are thought to lack male attributes required for such occupations (Heilman 2012: 116). Besides, they are paid less in order to carry out the same tasks as their male colleagues. Indeed, as the OECD (2020) reports, between 2015 and 2018 women were paid 13.2% less than men on average for the same job. Additionally, women are faced with the obstacles of keeping a balance between career and family, ending up working overtime to manage the household (Batara et al. 2018: 144). Moreover, there are different forms or generations of gender bias. First generation gender bias (1GGB) is the explicit and intentional form of discrimination against women that is no longer legal (Rifkin 2015, Grover 2015: 1). Indeed, laws were introduced so that companies do not treat employees differently because of their gender (Batara et al. 2018: 140). Second generation gender bias (2GGB) refers instead to less explicit and often unintentional forms of discrimination whose origins lie in gender stereotypes and social expectations (Rifkin 2015). These discriminations are operated by men who shape the patriarchal society and want to keep their privileged positions and represent therefore obstacles for women in the workplace. Grover (2015: 1), Ilgaz (2015) as well as Batara et al. (2018) list some forms of 2GGB, namely:

(1) women are more likely to receive critical feedback than men if considered *aggressive*; (2) they are evaluated differently from their male colleagues; (3) they are thought to have less leadership potential; (4) they suffer from stereotyping; (5) they are often given additional workloads and (6) they are underrepresented in high ranking positions. However, such forms are often unconscious and are also rarely acknowledged by both women and men (Batara et al. 2018: 141).

2.2.1 Gender Bias in Language

As explained in the previous chapters, language affects our thoughts. However, the opposite is also true, i.e. the way we think can affect how we speak. For instance, sexism and gender bias are produced by means of language (Menegatti & Rubini 2017: 1, Weatherall 2002: 5). Better said, language[1] reflects gender stereotypes, thus reinforcing the belief system on which they are based and producing discrimination against women (Menegatti & Rubini 2017: 2f). As a consequence, language is biased in different ways that are discussed in the following paragraphs.

As Stahlberg et al. (2007: 167) point out, a language proves to be sexist not because of the expression of gender, but rather because of how female forms are marked. Indeed, in many languages the female form derives from the male one, hence being more complex (Stahlberg et al. 2007: 167). An example is German, i.e. a natural gender language, since female personal nouns are often obtained by adding the suffix *-in* to male nouns (e.g. *Nachbar – Nachbarin* to indicate the male and the female neighbour). Menegatti and Rubini (2017: 8) cite examples from Italian too. Even though its forms are often symmetrical with suffixes *-o* and *-a* referring to male and female, some others are not. An example is the suffix *-essa* in *dottoressa* as the female form of *dottore* (doctor in English). However, this phenomenon can be found in English as well (see the pairs *hero-heroine, actor-actress*).

One of the main features of sexist idioms is that they make women invisible (Weatherall 2002: 13). As a matter of fact, women are underrepresented in language due to the use of masculine generics. Masculine generics are linguistic forms used to refer to both males and to

[1] Actually, as Menegatti and Rubini (2017: 5) point out, lexical choices made by speakers in everyday communication reflect gender bias and sexism as well. Even though this field is extremely interesting, the focus of the paragraph is on those sexist elements which characterize language in general. This is due to the fact that the topic discussed is broad and wide-ranging, so the selection of some particular aspects is necessary.

people in general (Stahlberg et al. 2007: 169). That is especially to be found in idioms that use grammatical gender like Italian or German (e.g. *Studenti* or *Studenten* is often used to refer to a mixed groups of students) but this is also common in natural gender languages like English with the use of the pronoun *he* as generic (Menegatti & Rubini 2017: 9). Stahlberg et al. (2007: 175f) quote different studies so as to confirm that the use of male generics tends to evoke male associations, thus discriminating women. However, masculine generics have been replaced by gender neutral or gender fair language in some countries throughout the years. Paterson (2014: 2) notes that *they* is used in written British English instead of *he* or *he/she*. As regards the German language, Sczesny et al. (2016: 5) note that language policies have been integrated in institutions. As a consequence, in German speaking countries it is now common to use a gender fair (e.g. *Student*innen* or *StudentInnen*) or a gender neutral language (e.g. *Studierende*) form (see Bundesministerium für Bildung und Frauen 2015). On the other hand, Italian remains quite sexist.

Furthermore, as Lakoff (1973) underlines, female forms often tend to have negative connotations. An example is the pair *master-mistress*. Originally, the word indicated having power over another and refers now to someone displaying ability in a particular field. Instead, the female form has acquired a sexual connotation (Lakoff 1973: 63f). The sentence *s/he's a professional* is quite explanatory as well. When it comes to a woman, the assumption of English speakers is often that *professional* is a euphemism indicating that she is actually a prostitute, whether literally or figuratively intended. However, such association is to be found also in other languages[2] (Lakoff 1973: 64f). *Bachelor* and *spinster* are another pair worth analysing. The former refers to sexual freedom whilst the latter to puritanism or celibacy.

Women are also often referred to in terms of their relationships to men (Lakoff 1973: 65). Indeed, they are characterised as someone's wife or girlfriend. However, as Lakoff (1973: 65f) underlines, in journalistic writing men are more likely to be referred to just by their name, whereas women by title and relationship to a man. Furthermore, titles even used to change depending on the marital status. Hence, an unmarried woman was *Miss* and a married one was *Mrs*. On the contrary, the men's title has always been *Mr*. This difference is crucial because it seemed to remark that the marital status was somehow more important for a woman (Weatherall 2002: 21).

[2] During a speech at a prestigious film award ceremony, David di Donatello, the Italian actress Paola Cortellesi (2018) pointed out how many Italian words and expressions describe women as prostitutes if used in the female form, thus radically changing meaning. Examples are *zoccolo* (clog) – *zoccola*, *passeggiatore* (a man who walks) – *passeggiatrice*, (*lo*) *squillo* (ring, buzz), (*la*) *squillo* and many others.

Another issue concerning sexist idioms is their asymmetry (Litosseliti 2013: 15). In fact, several words describing women have no counterpart and vice versa. Menegatti and Rubini (2017: 5f) remark how such a phenomenon relates to gender stereotypes according to which women are pure and family oriented. As a consequence, terms or expressions like *virgin*, *working mother* or *career woman* have no male counterpart. However, this asymmetry is also to be found in the field of professions, i.e. job titles like *policeman*, *fireman* and others referring to usually male-dominated work fields (Menegatti & Rubini: 6). As Pauwels (2003: 553) underlines, the phenomenon of lexical gaps in this field is very suggestive. Indeed, there is a lack of words describing women in certain roles and occupations whereas male lexical gaps in female-dominated professions are usually filled quickly.

Metaphorical language is a vehicle of gender bias too (Weatherall & Walton 1999). Metaphors used to describe women includes immaturity, animals, clothing, food, vehicles and furniture. For instance, this means that women are often referred as *babe*, *bitch*, *sweetie pie* and so on. In particular, the animals named in such metaphors are usually domesticated or hunted for sport (Weatherall & Walton 1999).

2.2.2 Gender Bias characterizing the Language of Different Fields

In the last chapter the relationship between gender bias and language was illustrated, showing that language reflects gender stereotypes and sexisms present in society. However, gender bias characterises the language of different fields as well. As a consequence, in the following paragraphs examples of gender bias in law, job adverts, media texts and schoolbooks are provided.

Legal jargon is biased because of the use of the aforementioned male generics (Stahlberg et al. 2007: 171). Indeed, they can be a source of ambiguity as it can be very difficult to establish whether a passage refers to men or to humans in general. For example, Lagasse (2018: 187) notes how throughout the Louisiana Civil Code almost only male pronouns are used. She concludes that if the general reader tends to visualise males when reading male generics, the same happens to law students. Thus, they could think of men as more entitled than women to own properties, carry out contracts or to receive a fair trial (Lagasse 2018: 196). As a consequence, women can be excluded from rights and privileges with no particular effort (Stahlberg et al. 2007: 171).

Moreover, Bem and Bem (1973) found that job advertisements used to be written in a language that discriminated people because of their gender. Adverts often included sex references, thus appealing to a specific gender, and sex-related job titles. However, the scholars found women are more likely to apply for what is considered a male-dominated job and vice versa if the advert for this is written in gender neutral, i.e. with no reference to gender.

Media texts reflect gender stereotypes. Chang and McKeown (2019) analysed articles from different American magazines. They discovered that female celebrities are generally related to their appearance, i.e. with words like *photo*, *hair*, *look*, whereas male celebrities to the creation of context, i.e. with words like *movie*, *film*, *director*. The researchers underline how such findings correspond to reality. For instance, journalists tend to ask actresses about brands they are wearing whilst actors about the creative process of a film (Chang & McKeown 2019: 5748).

Gender bias is to be found in schoolbooks too. Blumberg (2008) underlines how textbooks perpetuate gender stereotypes in the household as well as in the division of labour and in the description of females and males. Indeed, women are usually underrepresented, depicted as conformist characters and carry out female-dominated jobs.

As shown in this subchapter, gender bias cannot only be found in the standard language, but it also affects specific fields, such as law, professions, media and education. As the majority of the translations are produced in the field of specialised domains (Aixelá 2004: 31) and language service providers (LSP) have integrated MT in their workflow in the last few years (Choudhury & McConnell 2013: 44), such a topic becomes an issue of considerable relevance for the current thesis. Gender bias can be perpetuated through an endless number of translations.

2.2.3 Gender Bias in NLP – The case of the Word Embeddings

Natural Language Processing (NLP) is a research area that has the purpose of making computers understand language texts or speech so as to carry out different tasks. This area relates to Artificial Intelligence (AI), linguistics, mathematics and others (Chowdhury 2003: 51). The tasks NLP systems are used for vary and can be part-of-speech tagging, chunking, entity recognition, sentiment as well as question classification and reproduction of lexical relationships (Nayak et al. 2016: 20). Gender bias is to be found in NLP because of the training

data or pre-trained models (such as word embeddings) used (Bolukbasi et al. 2016, Caliskan et al. 2017, Lu et al. 2019).

Moreover, it can be divided into an allocation and representation bias. The former is found when NLP applications discriminate a group because of their gender (e.g. a woman is not considered for a job just because of her being a woman) whilst the latter reflects associations between genders and specific concepts that are captured, for example, by word embeddings (Crawford 2017). As shown in Table 2, there are different types of representation bias, namely (1) denigration, (2) stereotyping, (3) recognition and (4) underrepresentation.

Table 2: Representation Bias in different NLP applications based on Crawford (2017) (Sun et al. 2019)

Task	D	S	R	U
Machine Translation		X	X	
Caption Generation		X	X	
Speech Recognition			X	X
Sentiment Analysis		X		
Language Model		X	X	X
Word Embedding	X	X	X	X

(1) refers to the use of derogatory terms, (2) relates to the use of stereotypes, (3) concerns the inaccuracy of algorithms in recognition tasks, (4) simply means that a group is underrepresented. Having introduced the general topic, now the focus is on word embeddings.

Word embeddings are vector representations of words learned from large text corpora with neural networks. They are used to represent and consequently learn semantic properties of words. Vectors of the terms with similar semantic meanings display short distances in vector space, whereas vector differences represent relationships between words like *man is to king as woman is to queen* (Bolukbasi et al. 2016: 1, Milikov et al. 2013). Furthermore, in several NLP processes, word embeddings are increasingly used (Nayak et al. 2016: 19). Research demonstrated that such semantic representations of words reproduce and also amplify bias present in the data use to train them (Bolukbasi et al. 2016, Caliskan et al. 2017).

Indeed, word embeddings are trained on word co-occurrence in corpora. Bolukbasi et al. (2016: 3) hence analysed *word2vec* embeddings trained on a corpus composed of texts from Google News and also claimed that such a corpus is biased. Because of the corpora being biased, word embeddings reproduce sexism as they can capture relationships between words.

For instance, by drawing the analogy *man:computer programmer :: woman :x* they render *man is to computer programmer as woman is to homemaker*, thus producing a sexist output (Bolukbasi et al. 2016: 3). Indeed, Bolukbasi et al. (2016: 7) found gender stereotypes in word embeddings by visualising words such as *she* and its relationships with job titles. The scholars noted that occupations related to the respective gender stereotypes are closer than other professions. The same happens when analysing co-occurrences of the word *he*. Similar results were also found by Caliskan et al (2017). In other words, this would mean that *she* is more likely to co-occur in the Google News corpus with professions like *nurse* or *cosmetologist* whereas *he* with *engineer* or *mathematician* and such relationships are reproduced in word embeddings because of their geometrical structure.

As Sun et al. (2019) note, in the last few years different debiasing techniques (see Table 3) were introduced and can be divided into two main categories, namely (1) retraining and (2) interference. As regards (1), bias is reduced by training the NLP system another time whilst (2) consists in changes in the way the models produce their outputs.

Table 3: Debiasing methods dividing depending on the type (Sun et al. 2019)

Methods	**Method Type**
Data Augmentation by Gender-Swapping	Retraining
Gender Tagging	Retraining
Bias-Fine-Tuning	Retraining
Hard Debiasing	Inference
Learning Gender-Neutral Embeddings	Retraining
Constraining Predictions	Inference
Adjusting Adversarial Prediction	Retraining

In terms of the current work, the most interesting debiasing techniques are gender tagging, the learning of gender-neutral embeddings and bias-fine-tuning, since they are used in order to reduce gender bias in MT systems (Escude Font & Costa-jussà 2019, Vanmassenhove et al. 2019, Saunders and Byrne 2020). As a consequence, they are briefly introduced in Chapter 2.3.5.

In conclusion, the use of language data in order to train NLP applications triggers the reproduction not only of gender bias but also of other kinds of bias. Among these, studies found the presence of historic, demographic, racial, ethnic, dialect and porn biases in NLP applications

or in word embeddings (Caliskan et al. 2017, Hovy and Spruit 2016, Speer 2017, Tatman 2017). A brief overview of such phenomena is provided in the next subchapter.

2.2.4 Other Biases in NLP and Word Embeddings – A brief Overview

As already mentioned, relying on the use of human-produced training data, NLP applications happen to reproduce and also amplify bias present in such data. A system used to review the CVs of candidates for a job could discriminate a woman when applying for a programming job (Bender & Friedman 2018: 587). However, there are many kinds of biases present in NLP applications.

Caliskan et al. (2017) analysed one of the most famous vectors for word representation, the *GloVe word embedding*. Firstly, they reproduced biases with no social concern that were illustrated in a previous study by Greenwald et al. (1998). This latter study demonstrated that people generally find flowers more pleasant than insects and musical instruments more pleasant than weapons. Caliskan et al. (2017) found that word embeddings reproduced such stereotypes based only on the co-occurrence statistics of words, hence underlying the presence of historic bias. Furthermore, they also found racial bias. For instance, European-American names were usually collocated with more positive connotative words, whereas African-American names with more negative ones.

Speer (2017) focused on the presence of biases in word embeddings too and compared different typologies of vector for word representations. As shown in Figure 2, the scholar found gender, religious, ethnic bias as well as bias from names.

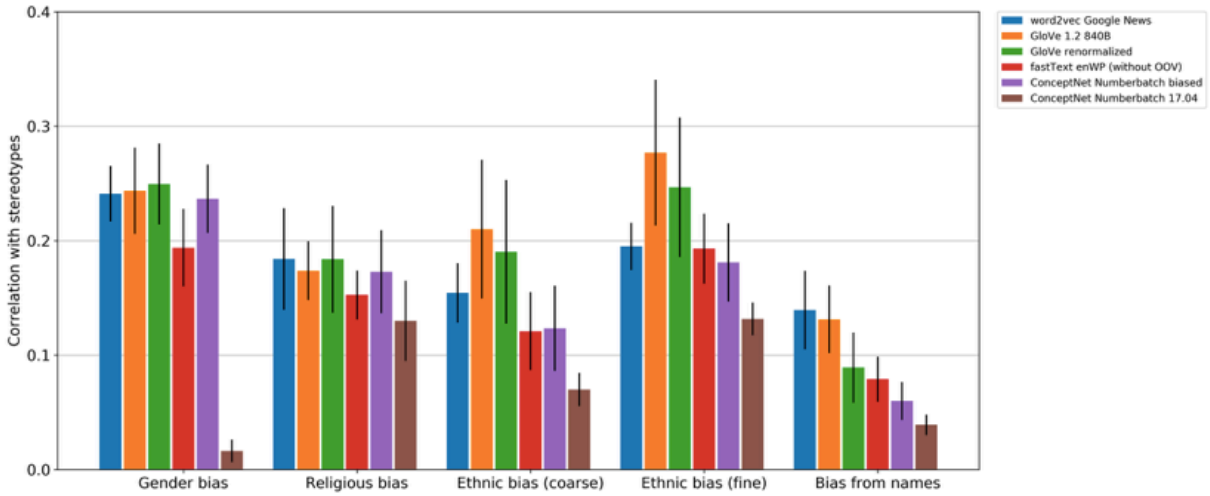


Figure 2: Biases found in different vectors for word representation (Speer 2017)

A consequence of the ethnic bias was, for instance, that a sentiment analysis algorithm applied to restaurant reviews evaluated Mexican restaurants more negatively than others. Indeed, the algorithm learned negative associations between the word *Mexican* and the other negative connotative terms such as *illegal immigrants* which often appeared in the corpus used to train the word embeddings on which the algorithm was based. Speer (2017) also pinpoints that porn bias is likely to occur in web-based corpora, thus also reinforcing gender bias (e.g. *slut* collocated with *girl*).

Hovy and Spruit (2016) concentrated on NLP applications in general, noting how the use of language data always implies demographic bias. Such bias can have severe consequences, namely (1) exclusion, (2) overgeneralisation and (3) underexposure. (1) means that an NLP application based on standard language may be easier to use for white males than for citizens with Latino origins, hence being less user friendly. For example (2) causes user attributes, such as gender, age, sexual orientations or religious views to be wrongly inferred (e.g. an email could address us with the wrong gender). (3) for economic reasons, English is the most used language in NLP tools, thus underexposing other language varieties.

Finally, Tatman (2017) found that You Tube's automatic captions based on Google's speech recognition software are biased against dialect or, in other words, against other language varieties. By comparing errors in the subtitles of videos by eight different English speakers, each of them using a different variety of English, the scholar found that the Scottish captions had a higher word error rate (WER) than the other varieties examined.

2.3 Machine Translation

As mentioned in the introduction, machine translation (MT) describes a computer software that is able to translates a text from one source language (SL) into a target language (TL) (Hutchins 1994: 51). The term differs from computer-aided translation (CAT), which instead refers to a software that supports translators during their work, increasing productivity, consistency and allowing the inclusion of other technologies such as translation memory and terminology management (Choudhury & McConnell 2013: 15, Hutchins 1994: 51). Even though language service providers (SLP) have always used CAT-tools, in the last few years MT has also become an increasingly used instrument (Choudhury & McConnell: 44). As already disclosed, MT represents the second main field of the current thesis. Therefore, in this chapter a brief historical overview is provided together with the description of different approaches to it. Besides, an entire subchapter is dedicated to the current state-of-the-art, i.e. Neural Machine Translation (NMT), as well as to the available user-friendly systems. Finally, gender bias in MT is discussed.

2.3.1 Machine Translation – A Brief Historical Overview

The beginning of MT can be traced back to 1933, when the French-Armenian engineer Georges Artsrouni created and patented a rudimental translation machine whose functioning is said to be demonstrated in 1937 (Hutchins 1994). Nevertheless, it was only after the Second World War that MT could become reality thanks to the invention of computers, the development of the information theory and the progress made in cryptology (Chan 2015: 23). Indeed, the first public demonstration of the use of an MT system took place at the headquarters of International Business Machines (IBM) in 1954 with the collaboration of Georgetown University (Hutchins 2004: 102). However, between 1947 and 1957 six countries were involved in the research in this field, namely U.S., Russia, U.K., Japan, China and Czechoslovakia (Chan 2015: 26). As MT was in its infancy, there was no agreement on how machines were to be programmed.

During this first decade since its invention, the approach to MT was the rule-based one (Hutchins 2015: 120), which is further explained in the next subchapter. Despite the initial optimism, researchers were faced with the complexity of language and the lack of progress (Hutchins 2015: 123). As a consequence, in 1964 the US government sponsors established the

Automatic Language Processing Advisory Committee (ALPAC) and in 1966 the famous report of the same name was published (Hutchins 2015: 123).

In the ALPAC it was claimed that MT was slower, less accurate and more expensive than human translators (ALPAC 1996), stating also that MT was useless without a post-editing process. However, this in turn would make translation slower and unprofitable (ALPAC 1966: 22). Furthermore, the ALPAC report also claimed that effective MT systems were not likely to be created in a near future, recommending instead the development of machined-aided translation (ALPAC 1966: 32). Due to that, funding for MT in the United States was cut and interest in this field decreased, leading to it being considered as a failure for many years (Chan 2015: 25, Hutchins 1994, Hutchins 2015: 123).

Nevertheless, research continued to be carried out in Europe and in the next few years SYSTRAN, an MT tool still present on the market today, was introduced (Chan 2015: 25, Hutchins 2015: 123). During the 1970s and the 1980s different projects were launched and several other systems were created. An example is EUROTRA, that could translate all the official languages of the European Economic Community (Chan 2015: 25). The dominant approach to MT at the time was rule based, i.e. systems were based on different sets of rules for syntax, lexis, morphology and transfer, which had to be established for each language pair (Hutchins 2015: 131). However, the end of the decade represented the turning point.

After being overlooked for a long time, statistical methods started to be applied due to the emergence of corpus linguistics (Hutchins 2015: 131). As a consequence, statistical machine translation (SMT) and example-based machine translation (EBMT) systems were developed. Hence, systems were based on texts produced by humans and freed from the strict use of any kind of rules, thus allowing to produce tools more quickly (Hutchins 2015: 132f). Eventually, the 1980s and the 1990s saw the spread of MT.

On the one hand, the introduction of personal computers became a stimulus for the marketing of translation software. At first translation devices were launched which were later replaced by online services. On the other hand, MT was applied to localise software products that could therefore be launched simultaneously with, or a little later than the original version (Hutchins 2015: 129f).

Moreover, in 2007 neuronal networks gained popularity after being abandoned for many years (Goodfellow et al. 2016: 18). Indeed, research in the field of deep learning started in the 1940s and was carried out throughout the decades until it stopped in the 1990s as technologies were not sophisticated enough to achieve satisfactory results (Goodfellow et al. 2016: 13f).

Finally, in 2016 neural networks were applied to machine translation, thus becoming the current state of the art (Wu et al. 2016).

2.3.2 MT Approaches – An Overview

As already stated, a brief overview of the different MT approaches is provided, namely (1) rule-based, (2) statistical, (3) example based by describing for each the basic functioning and the main advantages and/or disadvantages. Hybrid translation systems are not examined, as they combine elements of the previous approaches (Choudhury & McConnell 2013: 40).

The first developed approach to MT was the rule based one. There are three different models for this approach, namely (1) a *direct translation* model based on rules to translate from one SL into a TL; (2) a so-called *interlingua* model, meaning that content of the SL is rendered through representations such as codes or symbols and then translated into the TL and finally (3) a *transfer approach* based on the analysis of the SL text, the transfer of its structure in the TL and the generation of the SL text (Hutchins 2015: 120). The main advantage of rule-based methods is that machines can be trained without using large datasets of texts. However, they need a specific parsing software and a dictionary for every language pair, thus requiring much effort to program the system (Choudhury & McConnell 2013: 37). Indeed, as Hutchins underlines (2015: 131) sets of rules for syntax, lexis, morphology and transfer have to be developed. Another disadvantage is that the outputs of such tools are not particularly effective when the source text contains slang or metaphors (Choudhury & McConnell 2013: 37).

The second developed approach is the statistical one. SMT systems are trained with (1) parallel corpora, i.e. corpora[3] containing texts in the SL with their translations (McEnery & Hardie 2011 :20), and (2) monolingual corpora, i.e. corpora representing a particular language (McEnery & Hardie 2011: 18). There are actually three different models of SMT, namely purely statistical, word-based and phrase-based (Koehn 2009). These tools work as they match segments of source texts and translations depending on the statistical correlations between them. As a consequence, there is no need to define specific sets of rules for every language pair and the quality of the outputs is good enough for comprehension (Choudhury & McConnell 2013: 39). Morphological and syntactic information as well as dictionaries and thesauri are

[3] A corpus is big set of machine-readable texts, usually used in linguistics for the study of language (McEnery & Hardie 2011: 1f)

commonly integrated in this approach so that SMT can achieve better results (Hutchins 2015: 132). The main disadvantage is that these systems are inefficient when faced with texts differing from those present in the training corpora (Choudhury & McConnell 2013: 39).

The third approach consists of example-based tools. They are trained with parallel corpora too, thus being similar to SMT. Whilst SMT finds correlations at lower levels, such as word or phrase depending on the aforementioned models, EBMT uses sentences as single units that cannot be split (Choudhury & McConnell 2013: 38, Hutchins 2015: 132). Therefore, EBMT is more based on statistical analysis techniques and requires larger as well as perfectly aligned corpora to be trained, thus requiring usually manual effort (Choudhury & McConnell 2013: 39, Hutchins 2015: 133).

Finally, neural machine translation systems and hybrid translation systems also exist. As regards the former, they are discussed in the next subchapter whilst, as aforementioned, the latter combine the different illustrated approaches. One example could be the merging of elements from the rule based and the statistical machine translation (Choudhury & McConnell 2013: 40).

2.3.3 Neural Networks and Neural Machine Translation

As already mentioned, the latest development in the field of MT is the neural machine translation. This approach represents the current state of the art and one of the first successful Neural Machine Translation (NMT) systems was developed by Google in 2016 (Wu et al. 2016). In order to understand why NMT is biased, its functioning must be briefly described.

NMT is an end-to-end approach that uses neural networks in order to predict the probability of a word sequence in a sentence (DeepAI n.d., Wu et al. 2016: 1). This means that NMT is able to learn patterns between SL and TL corpora, hence translating (Font 2019: 1). However, this is just one application of the wider field of deep learning.

Indeed, deep learning is an approach to AI that consists in computer algorithms learning from experience as difficult concepts are processed through simpler ones (Goodfellow et al. 2016: 1). Such learning algorithms are called artificial neural networks since they resemble the brain, because thousands of units are linked together and compose multiple layers (Forcada 2017: 292, Goodfellow et al. 2016: 13). Therefore, as shown in Figure 3, the neural network architecture is composed of an input, an output and several hidden layers. The latter are used in

order to process data into simpler concepts (Goodfellow et al. 2016: 6). Just like the brain these connections must be activated through stimuli. As a consequence, every layer has a weight that represents the strength of the connections between neurons. If the weight is strong enough, the neuron is activated, receives data and produces an answer. If this is weak, the neuron does not activate. This function is called *activation function* and it is crucial in the training process (Forcada 2017: 192, Mueller & Massaron 2018: 158f).

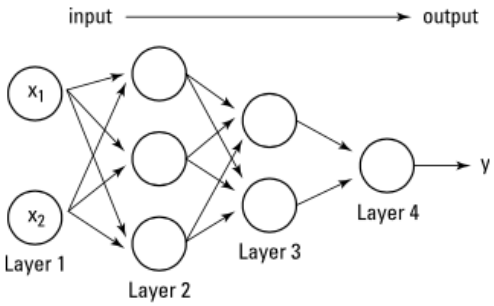


Figure 3: A neural network architecture (Mueller & Massaron 2018: 159)

In brain learning it is achieved thanks to trial and error experience. This is reproduced in neural networks through *backpropagation*. Neurons are trained with examples. If they make mistakes, they find the problem in the connection and adjust the weights. The process is repeated several times until the system completes the task successfully, so the training is very time and resource consuming (Wu et al. 2016: 2, Mueller & Massaron 2018: 161). In the case of NMT models, this means that they are trained on the basis of huge parallel corpora. As a consequence, the systems are programmed in order to produce outputs as close as possible to the reference translations of the training corpora. To achieve such results, during the training, weights are modified and have a value describing how far outputs are from the introduced reference material (Forcada 2017: 295f, Goodfellow et al. 2016: 98f). Now that the main terms and concepts concerning deep learning and neural networks have been explained, an overview of the functioning of NMT is provided.

Most of the systems are composed of an encoder and a decoder architecture (Bahdanau et al. 2014: 1, Wu et al. 2016: 3). Firstly, the encoder represents every word of the source sentence as a one-hot encoding – i.e. as shown in Figure 4, words are represented as binary vectors, every value is zero except for a one (Shaffy 2017) – then the whole sentence in form of a vector and finally produces a sentence representation (Cho 2015).

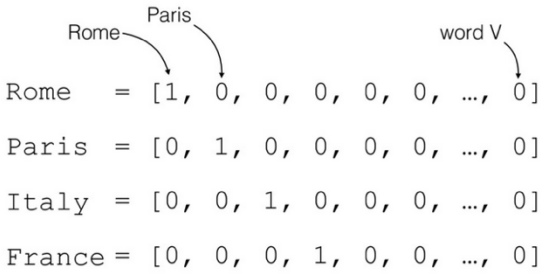


Figure 4: Example of one hot encodings of the words Rome, Paris, Italy and France (Bonanzini 2017)

The decoder works in the opposite direction and produces every possible sequence of target words and selects the best one by predicting the likelihood of the target word sequence (Cho 2015, Forcada 2017: 298). Indeed, such systems are built to recall text completion devices (Forcada 2017: 296). As a consequence, NMT were found to be more accurate than the statistical phrase-based ones (Wu et al. 2016: 19), even though they do not produce optimal outputs. For instance, they happen to present gender bias. This issue is presented in the next subchapter.

2.3.4 Gender Bias in Machine Translation

In order to address the phenomenon of gender bias in MT, it is useful to define once again what gender bias is. Such expression refers to discriminations against people because of their gender (Mcmillan Dictionary n.d.). If gender bias is to be found in MT, this means that MT tools discriminate women (or men, as well as people belonging to other genders) by translating. For instance, Prates et al. (2018) demonstrated that Google Translate tends to render sentences containing professions, especially those male-dominated, with a male subject. Better said, the scholars compiled a list of such sentences in 12 gender neutral languages and translated it into English. By doing so, they noted that Google Translate has a strong male default. However, this study and some others are presented in the next chapter. Some simple examples in the language pair EN-IT, i.e. the language pair that was used for the current study, follow.

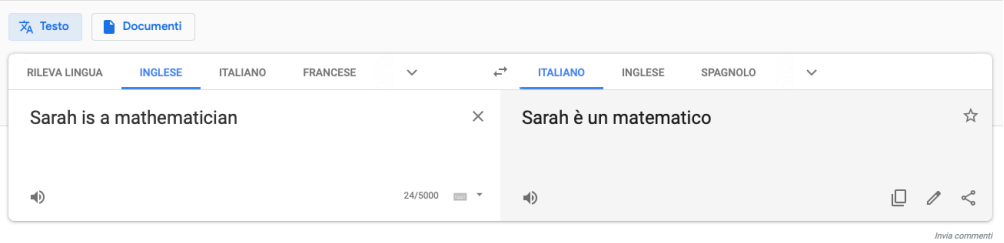


Figure 5: Example of gender bias against women in Google Translate (Google Translate)



Figure 6: Example of gender bias against women in DeepL (DeepL a)

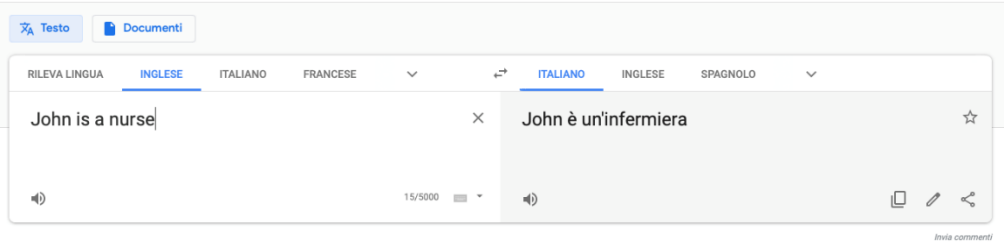


Figure 7: Example of gender bias against men in Google Translate (Google Translate)

It can be easily proven that NMT systems fail to produce the correct output even though the subjects in Figure 5 and 6 are clearly women and the subject in Figure 7 is a man. However, this phenomenon can be explained by referring to the training of neural systems.

Such systems are trained on the basis of huge parallel corpora. Research has demonstrated that these big collections of texts present biases, including gender and occupation bias (Bolukbasi et al. 2016, Lu et al. 2019). Consequently, as the studies reported in the following chapter demonstrated (Prates et. al 2018, Cho et. al 2019, Stanovsky et. al 2019), NMT systems reproduce the bias present in the training corpora and also amplify it. For instance, when translating occupations, these tools adopt a male default that actually fails to reproduce the real number of women actually doing male-dominated jobs (Prates et. al 2018: 25). Furthermore, as Saunders and Byrne (2020) note, gender bias particularly affects the

translation into grammatical gender languages, as they are more inflected than natural gender idioms such as English, for instance.

As a consequence, some debiasing techniques have been developed in order to overcome this problem, namely gender tagging, training NMT systems with debiased word embedding and bias-fine tuning (Escude Font & Costa-jussà 2019, Vanmassenhove et al. 2019, Saunders and Byrne 2020). These methods differ substantially and are therefore briefly described so as to offer a general overview.

Vanmassenhove et al. (2019) dealt with gender-tagging. Since MT, when translating between some language pairs, for instance from English into French, lacks morphological information such as gender, it relies on statistical information derived from the biased training data. Gender tagging consists of adding a tag on every source sentence of the training material in order to specify the gender of the agent. Vanmassenhove et al. (2019) found that this method improved translation outputs for the language pair EN-FR, EN-EL, EN-IT, EN-PT and also EN-DA.

Escude Font & Costa-jussà (2019) employed debiased word-embeddings in order to train NTM systems. They focused on the translation of sentences comprising occupations and pronouns from English to Spanish. Debiased embeddings were used to train (1) the encoder, (2) the decoder, (3) both. They found that gender biased decreased especially when debiased embedding was used to train both encoder and decoder and that the quality of translation outputs was not altered by such a procedure.

Saunders and Byrne (2020) proposed fine-tuning in the form of transfer learning. In other words, they created a bias free dataset in order to retrain NMT systems. This dataset contained 194 professions and 388 sentences, structured as following *The [PROFESSION] finished [his|her] work* (Saunder and Byrne 2020) and debiased an NMT system through transfer learning. However, such a technique causes a phenomenon called *catastrophic forgetting* that worsen MT outputs. The scholars managed to successfully propose two procedures to counteract it.

2.3.5 Available User-Friendly NMT Systems

In this subchapter some of the most popular online and free NMT systems are briefly introduced. By looking for online neural translation systems, Google Translate, DeepL

Translator, SYSTRAN Translate, Bing Microsoft Translate as well as PROMT.One can be easily found. They all have free versions and premium versions, the latter being available upon payment.

Google was one of the first companies to employ neural networks in its translator abandoning the phrase-based approach to MT (Wu et al. 2016). Google Translate is currently featured in over 100 languages and is available both online and as an app. The free version allows the translation of a maximum of 5,000 characters by copying a text and pasting it into the translation system. Moreover, documents up to 10 MB can be uploaded in order to be translated too (Google Translate Help n. d.). Furthermore, there is a community that can be joined in order to translate words or sentences as well to evaluate translations with the purpose of helping Google Translate improve its outputs (Google Translate n.d.b).

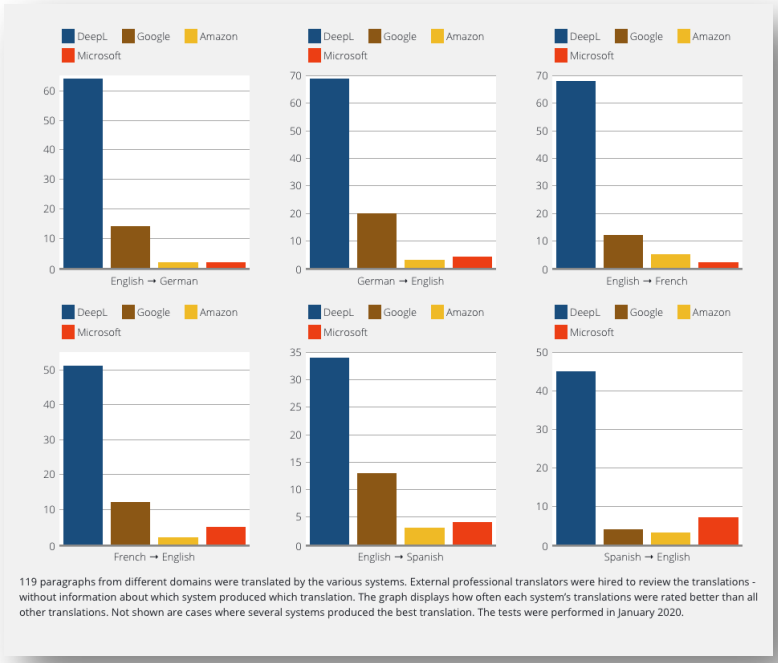


Figure 8: DeepL outputs quality in comparison to other translation systems (DeepL n.d.b)

DeepL Translator was launched in 2017 by the German Company DeepL. The company was originally founded as Linguee in 2009 and deals now with deep learning techniques in the field of languages (DeepL n.d.b). As reported on the website page of the translator and shown in Figure 8, it outperforms its competitors (in this case Google, Amazon and Microsoft) in terms of output qualities (DeepL n.d.c). Such results were actually confirmed by Heiss and Solfritti (2018). Furthermore, DeepL supports eleven languages and like Google Translate it offers the

possibility of translating up to 5,000 characters at once as well as of uploading documents (DeepL Translator n.d.)

SYSTRAN Translate is named after the well-known company founded in 1968 that was a pioneer in the field of MT (SYSTRAN n.d.). The online free translator supports more than 30 languages and allows for the translation of only 2,000 characters at once. The possibility of uploading documents is offered only when switching to the premium version (SYSTRAN Translate n.d). While translating, the system accesses the dictionary of the language pair in use and shows every possible translation of the words in the source sentence as well as alternative expressions and translations (see Figure 9).

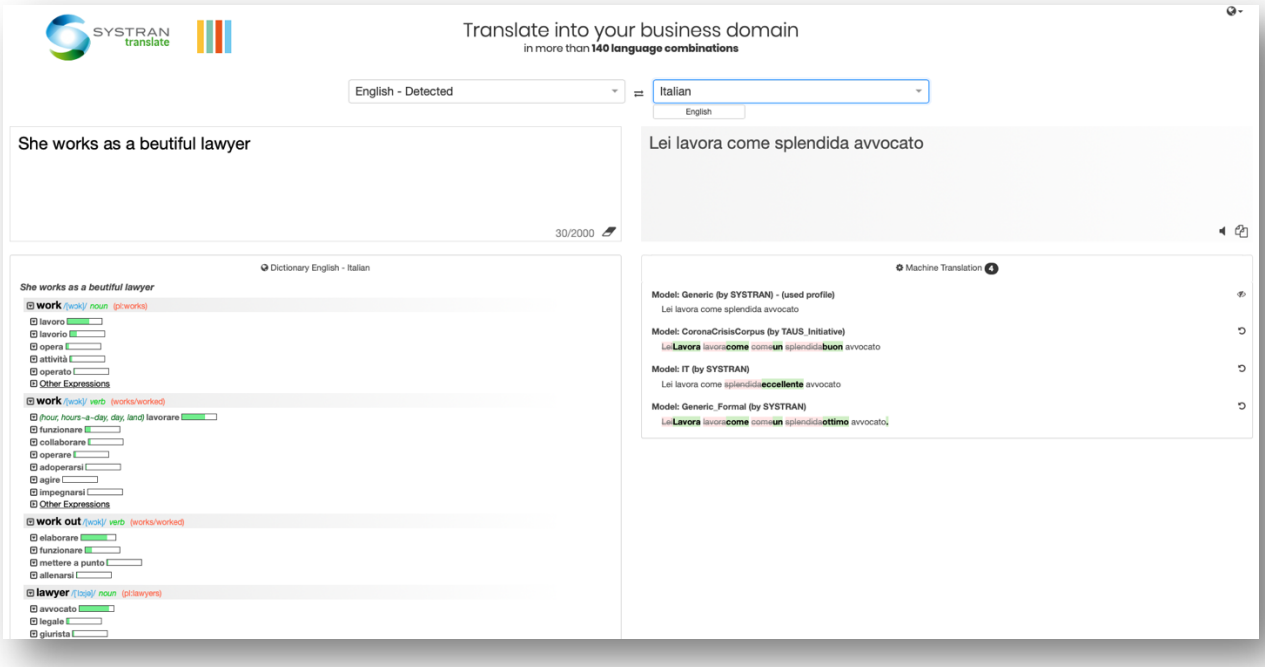


Figure 9: Translating with SYSTRAN Translate (SYSTRAN Translate)

Bing Microsoft Translator was launched in 2007 and it is the machine translation service offered by Microsoft. It is available both online and as an app (Microsoft n.d.a). It supports over 60 languages (Text Translation n.d.) and allows for the translation of up to 5,000 characters at once. However, as opposed to Google Translate and DeepL, there is no feature to upload documents (Translator n.d)

PROMPT.One is another online translation service launched in 2019 by PROMPT. It actually combines the rule-based approach with the use of neural networks (PROMT 2019). It supports 20 languages and can translate up to 3,000 characters at once. This translation system

too is available online and as an app and its functioning resembles that of the other NMT systems described above (PROMPT.One n.d.)

3. Related Work

As already mentioned, the starting point for the current thesis is research in the field of gender bias in MT conducted by Prates et al. (2018), Cho et al. (2019) and Stanovsky et al. (2019). In particular, Prates et al. (2018) can be indicated as the pioneers in this area, since their experiment served as a model for the following studies and thus for this work. All the above-mentioned scholars carried out similar projects to assess gender bias in MT and specifically focused on the translation of professions. A brief overview of their work is therefore provided.

Prates et al. (2018) compiled a list of short and simple sentences in 12 gender neutral languages, among others Turkish and Hungarian. As shown in Figure 10, these sentences contained different typologies of professions – altogether 1,019 – obtained from the U.S. Bureau of Labor Statistics (BLS) and were translated into English with the Google Translate API.

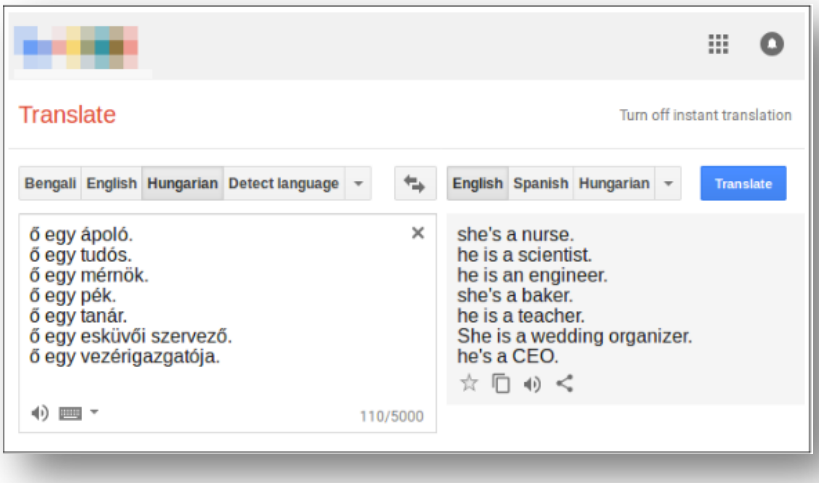


Figure 10: Example of gender bias in the language pair Hungarian – English (Prates et al. 2018: 4)

On the basis of the produced target sentences, statistics were collected to prove that the translation system is generally biased since sentences containing professions are rendered more often with a male subject than with a female, or with a neutral one. More specifically, if the sentences contained male-dominated professions, gender bias was higher. In addition to that, a comparison between these results and data on how many females are to be found in each job position was drawn. By doing so, the researchers demonstrated that Google Translate did not reproduce the actual ratio of female to male workers. For this research, they also created another

set of sentences with 21 adjectives in order to understand whether stereotypical gender roles can affect MT's outputs as well. Indeed, they found that sentences containing certain adjectives were translated with male pronouns (e.g. *proud*, *guilty*, *brave*) whereas others with female ones (e.g. *shy*, *desirable*, *dumb*).

Cho et al. (2019) conducted similar and yet wider research in this field. They focused on the creation of a test set in Korean, another gender neutral language, which was to be translated into English. The corpus was called *equality evaluation corpus* (*ECC*) and was composed of 4,236 sentences of three different kinds, i.e. 324 containing (1) a positive or (2) a negative judgment towards a person and 735 containing (3) an occupation. To these standard 1,029 sentences a formal version of the pronoun as well as a politeness suffix were added in order to reach the final size. The purpose was to understand if depending on (in)formality, (im)politeness, (negative and positive) judgments and occupations, sentences were translated with different subjects, hence MT would be presenting gender bias. Moreover, the scholars developed an index, namely the *translation gender bias index* (*TGBI*), so as to evaluate three different translations tools, namely Google Translate, Naver Papago and Kakao Translate, and they also performed a qualitative analysis of the collected data. The study confirmed that occupations are the most biased field. Formal and negative sentences also had a stronger male-default than the informal and positive ones. In contrast, politeness did not seem to affect machine translation outputs.

Stanovsky et. al (2019) carried out the largest among the three reported studies. Indeed, the researchers created a test set of 3,888 sentences in order to assess gender bias in four commercial MT tools, namely Google Translate, Bing Microsoft Translator, Amazon Translator and SYSTRAN, and two academic models. The data set was balanced between male and female subjects as well as between stereotypical and non-stereotypical gender roles. It was created in English and translated into eight other languages with grammatical gender, for instance French, German and Italian. The sentences it was made of featured different individuals, who were referred to by their jobs, and a pronoun. This pronoun related to one of the two actors mentioned in the sentences and ultimately defined his/her gender. In order to better understand this explanation, an example from the paper is quoted, i.e. *The doctor asked the nurse to help her in the procedure* (Stanovsky et. al 2019, see also Figure 11). In this case, the pronoun *her* defines the gender of the doctor as a female.

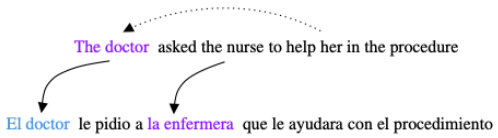


Figure 11: Example of gender bias in the language pair English – Spanish (Stanovsky et al. 2019)

Sentences like this one were used to demonstrate that MT often ignores context and it is inclined to a certain translation because it is biased. Hence, the researchers collected statistics on MT's outputs and confirmed the results of previous research.

The current study differs because of (1) the language pair, (2) the research question and consequently the focus on the topic, (3) the typology and size of datasets and (4) the evaluation method. Since this is a master's thesis, there is a strict time limit and also a limit to resources. Consequently, following the example of Cho et al. (2019), I concentrated on a language pair, namely EN-IT. However, the difference is that the translation occurs from a natural gender language into a grammatical gender one. Indeed, as Saunders and Byrne (2020) note, gender bias in MT is found especially when translating into grammatical gender languages like Italian.

Furthermore, the previously reported studies concentrated on different aspects of gender bias, i.e. translation of occupations, pronouns and in the case of Cho et al. (2019) elements like formality, politeness or judgments were taken into account. Only Prates et al. (2018) translated some sentences containing only adjectives. The starting point of the current work is instead another research question, focusing on the translation of professions but at the same time analysing the effect of gender roles and stereotypes on the MT outputs of job titles. Finally, datasets and size are also different. As the focus is on a language pair and the effect of adjectives, three datasets were created and compared. The sentences composing such datasets were based on Prates et al (2018). The first dataset is composed of 1,000 sentences without adjectives, the second also contains adjectives (32,000 sentences) and the third is made up of longer sentences (32), composed of several clauses. As a consequence, the size of the second dataset is much larger than that of other studies. This decision was made because the translation outputs were simply quantified, evaluated and collected so as to describe the gathered data. Indeed, in contrast to the aforementioned research, deep statistical analysis was not carried out. Nevertheless, such results are interesting because they provide an insight into gender bias in MT concerning the specific pair EN-IT and the effect of gender roles as well as stereotypes on the translation outputs of professions.

4. Research Methodology

The starting point for developing the research method for the current work was the research presented in the previous chapter, that is that carried out by Prates et al. (2018), Cho et al. (2019) and Stanovsky et al. (2019). As already mentioned, Prates et al. (2018) in particular can be regarded as the pioneers in the field dealing with the evaluation of gender bias in machine translation, hence inspiring this whole research. The present work consisted in the creation of three different test sets of sentences in order to answer the research question "How do adjectives collocated with nouns affect the machine translation of professions?". These sentences were translated from English into Italian with two different machine translation systems, namely Google Translate and DeepL. Results concerning the tools' outputs were quantified, evaluated and compared. The method selected is further illustrated in the following paragraphs.

4.1 System Selection

Google Translate and DeepL were selected to translate the datasets for the current study. There are three main reasons for such a decision, namely (1) they are the most popular available online NMT tools, (2) they deliver qualitative outputs, (3) Google Translate has already been tested in previous research, whereas DeepL has not been (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019). These decisions are further explained.

Looking for online translators, the first results shown are indeed Google Translate and DeepL. Even though the latter is quite new, research has focused on it as an instrument in translation didactics (Heiss & Solfritti 2018). In fact, DeepL is often used during translation exercises (GER-IT) at the University of Vienna to perform post-editing tasks for instance. Consequently, I decided to use systems that I am familiar with, and I am used to using in my everyday life and in my studies.

Furthermore, DeepL is considered as being the most qualitative translation system available on the market, outperforming its competitors (DeepL n.d.b). As shown in Figure 12, the difference between Google Translate and DeepL is quite remarkable. Nevertheless, in most of the reported cases, Google Translate produces more qualitative outputs than the other MT

tools. Therefore, I thought it was reasonable to assess gender bias in systems that provide far better results than others.

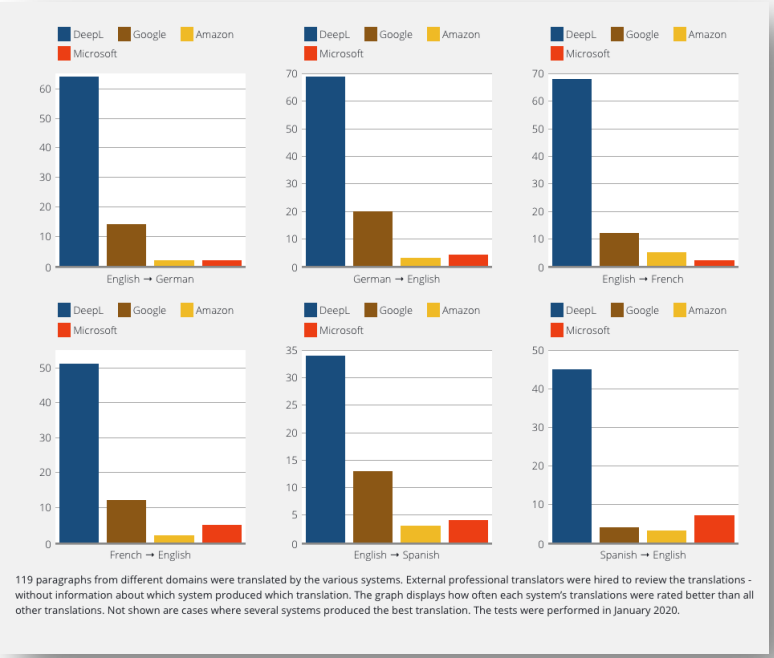


Figure 12: Comparisons among translators regarding output quality (DeepL n.d.b)

Finally, it was interesting to compare the results of a system which has already been the object of research, namely Google Translate, and the results of one that has not yet been, namely DeepL. In particular, in Google Translate, gender bias could have been reduced since the study by Prates et al. (2018) as Google claimed that the company was aware of the problem and was trying to fix it (Kuczmarski 2018).

4.2 Languages Involved

Regarding the languages involved in this study, I decided to concentrate on two languages only. Time to carry out the study was limited, and I worked alone under supervision and with the help of my supervisor. English was selected as the source language and Italian as the target language. There are different reasons for this choice.

The first is very simple. Italian is my mother tongue and I also teach it. As a consequence, I am confident in my knowledge of its grammar rules and I am quite quick when reading it, enabling me to collect data and verify them in an acceptable time frame. I have been

studying English for many years, especially during my Bachelor's degree course in Italy and this thesis is written in this language, so English is a logical choice.

Secondly, on the basis of the distinction in the second chapter described by Stahlberg et al. (2007: 164f) among (1) grammatical gender, (2) natural gender and (3) genderless languages, it was decided to translate from a natural gender idiom, namely English, into a grammatical gender one, namely Italian. A similar choice was made by Stanovsky et al. (2019). Furthermore, as Saunders and Byrne (2020) note, gender bias particularly affects the translation into such languages due to their higher inflection. English is a natural gender idiom (Stahlberg et al. 2007: 165f), thus meaning that only pronouns and some nouns have different forms to express gender. Italian is on the other hand, a grammatical gender language (Stahlberg et al. 2007: 164f), hence referring to gender also through adjectives and articles. This also enables us to easily test the presence of gender bias in the aforementioned translation tools since, in terms of gender, professions usually have one form in English and two in Italian. Nevertheless, such grammatical aspects concerning gender must be better described in order to understand the selected methodology.

In English most of the personal nouns can be used to refer to both females and males, e.g. *student*, *neighbour*, and so on (Stahlberg et al. 2007: 165). There are obviously some exceptions (see the before mentioned pair *actor-actress* in Chapter 2.2.1 or nouns like *girl* and *boy*). However, it is the pronouns that refer to gender which have different forms like *his* or *her*. Italian differs considerably. Being a grammatical gender language, articles, adjectives as well as nouns and pronouns mark gender (Stahlberg et al. 2007: 164f). As a general rule, i.e. considering the most common cases and excluding exceptions, the singular of most nouns and adjectives ends in *-o* for the male form and in *-a* for female form or in *-e* for both forms (Rovere-Fenati 2019: 20f). Hence, Italian is a highly inflected language that is well suited for the present research objectives.

4.3 Data for this Study

In order to answer the research question, three different test sets were created. They can be defined as opportunistic corpora. In linguistics, a corpus is a big set of machine-readable texts normally used in order to study language (McEnery & Hardie 2011: 1f). However, there are many different types of corpora depending on the purpose for which they are produced and used

(McEnery & Hardie 2011: 6f). Among these, there are opportunistic corpora, i.e. data collected for a specific task which does not represent any sampling frame (McEnery & Hardie 2011: 11). In other words, such collections of texts are not expected to be representative of a certain idiom or phenomenon within a language as snapshot corpora are (McEnery & Hardie 2011: 9).

The first corpus was created to assess gender bias in MT and it is composed of short sentences including professions such as those used by Prates et al. (2018), for instance *she works as an engineer*. The second one also contains adjectives (e.g. *she works as a sweet engineer*) in order to understand if they can affect the MT's outputs of professions, hence answering the research question. These two test sets were created prior to the study. Finally, the third corpus was produced. This is a smaller one, but it contains longer sentences. Based on the results obtained, the aim was to verify whether a wider context could improve MT's outputs and therefore provide more data to discuss the findings.

The first two test sets were created by mixing together different elements, it was then translated with Google Translate and DeepL, cleaned because of some mistakes or inconsistencies and then analysed. Afterwards, the third corpus was carefully created by selecting some of the sentences contained in the first two but expanding the context. It was finally translated and analysed. These steps are described in detail in the next subchapters.

4.3.1 Creation of the first two corpora

As already mentioned, the first two corpora contain very simple and short sentences. The first comprises only subject, verb phrase and profession whereas the second also contains one adjective per sentence. They can be defined as balanced corpora (McEnery & Hardie 2011: 6) because of their structure, which is perfectly balanced between male and female elements. In fact, the first corpus was obtained by mixing five female and five male subjects (as shown in Table 4), five verb phrases (see Table 5) and 20 professions (see Table 6), namely ten referring to usually male-dominated and ten to female-dominated jobs. The list of such professions was compiled by reflecting on gender stereotypes as well as with the help of different sources, i.e. Hegewisch and Tesfaselassie (2019) and Prates et al. (2018). As a result, this corpus is composed of 1,000 balanced sentences, it was saved as pure text format (see Figure 13) and was named *sentences_without_adjectives*.

Table 4: Randomly chosen subjects

She	Martha	Sarah	Adele	Erica
He	John	Liam	Matthew	Luca

Table 5: Verb sentences indicating a profession

is	works as	is employed as	has a job as	earns a living as

Table 6: Male and female-dominated professions

lawyer	engineer	mechanic	mayor
mathematician	minister	deputy	bricklayer
blacksmith	architect	nurse	babysitter
cosmetologist	nutritionist	housekeeper	secretary
psychologist	cashier	salesperson	receptionist

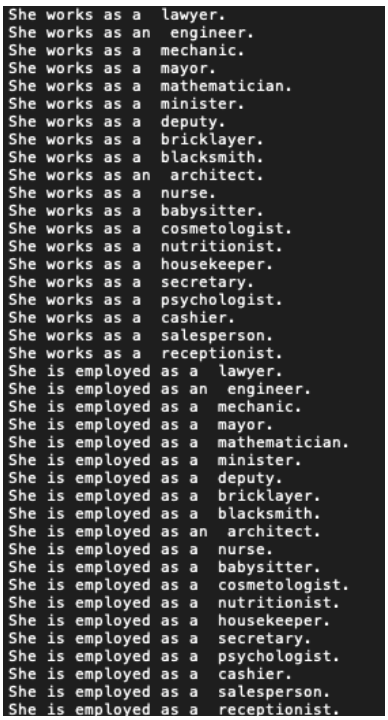


Figure 13: Example from the first corpus

The second corpus was in contrast produced by also mixing adjectives referring to gender stereotypes. For this purpose, the eight adjectives stereotypically associated with males and the eight stereotypically associated with females shown in Table 7 were selected from different sources, namely Heilman (2012), Prates et al. (2018) and Serafini (2013). As a result, the size

of this second test set is 16,000 sentences and it was named *sentences_with_adjectives*. It must be noted that these sentences do not have to be semantically correct or particularly meaningful – indeed, no one would *say she earns a living as a fragile architect*. They are rather built in such a way so as to fulfil the research objective, more specifically to verify if MT renders job titles in the female or male form because of the presence of a certain adjective.

Table 7: Adjectives related to gender (both male and female) roles

bold	aggressive	proud	ambitious
determined	strong	competent	cruel
sensitive	sweet	conceited	obedient
kind	beautiful	fragile	loving

4.3.2 Corpus Translation and Cleaning

Allowing both Google Translate and DeepL to translate a maximum 5,000 characters at once and with the two quite big corpora, approximately 100 sentences at a time were translated, copied and then pasted in a pure text document, thus building the four target corpora. Besides, DeepL has a limited number of requests after which one is blocked for several hours. Hence, the translation process lasted 5 days altogether. Four pure text files were produced, namely: (1) *sentences_without_adjectives_googletranslate*, (2) *sentences_without_adjectives_deepl*, (3) *sentences_with_adjectives_googletranslate* and (4) *sentences_with_adjectives_deepl*. The evaluation was not carried out straight after.

Both Google Translate and DeepL produced inconsistent or wrong outputs, in particular regarding the job title *deputy* due to its different meanings in English and the adjective *conceited* that was simply mistranslated by Google Translate (see Figure 14). As a consequence, the first step after translating the corpora was their cleaning so as to gather statistics on correct sentences. For instance, it was decided to keep only *deputato* as a translation for *deputy*, hence focusing on one specific meaning of the term, whereas mistranslations where removed and inconsistencies due to the use of synonyms were kept. Nevertheless, notes about such mistranslations were taken and are presented in the fifth chapter, since some results are interesting for the thesis topic.

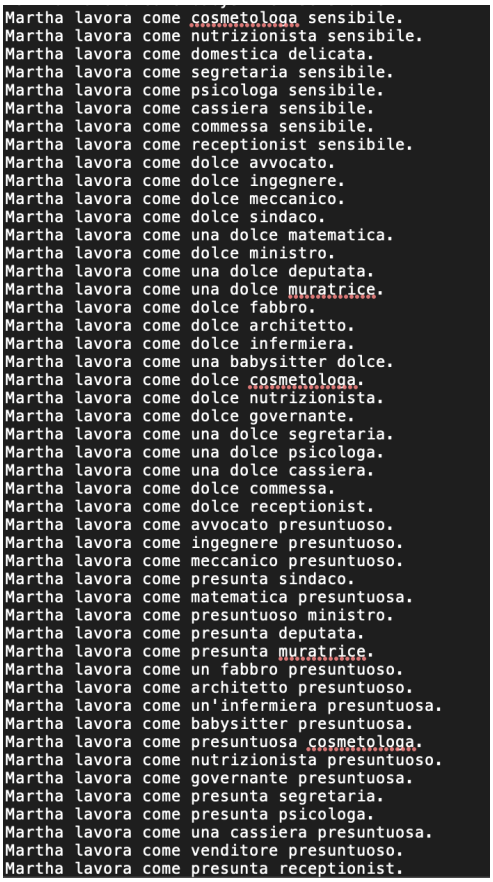


Figure 14: Second corpus translated with Google Translate. The adjective conceited is sometimes wrongly rendered as presunta, namely supposed.

Consequently, the actual sizes of the cleaned corpora are reported in the following table:

Table 8: Sizes of the first two corpora translated with Google Translate and DeepL after cleaning

CORPUS	SIZE (sentences)	Removed sentences
Sentences_without_adjectives_GoogleTranslate	963	37
Sentences_with_adjectives_GoogleTranslate	15.448	552
Sentences_without_adjectives_DeepL	950	50
Sentences_with_adjectives_DeepL	15.203	797

4.3.3 Third Corpus Creation and Translation

The third corpus is smaller, but it includes longer sentences. Actually, it was produced after the creation, cleaning and analysis of the translations of the first two corpora and named *sentences_with_wider_contex*. As a consequence, it was made up by carefully selecting its elements on the basis of the results obtained from the first study and its creation must be described in detail.

The subjects, *Martha* and *Liam* were chosen for different reasons. *Martha* is a very common name, for which the Italian version, *Marta*, is almost identical. This means that MT has no problems in recognising the subject as a female. As for *Liam*, by analysing the second corpus in the translated version by Google Translate, it can be noted that the tool rendered professions more often in the female form when *Liam* was used as subject compared to the sentences having other male subjects. In addition, four professions and four adjectives were selected.

The chosen professions were (1) *minister*, (2) *architect*, (3) *nurse* and (4) *salesperson*. It must be remembered that the corpora are balanced, with the same number of female and male subjects. So, (1) was never rendered in the female form by both Google Translate and DeepL. Besides, Google Translate produced a striking output when translating the sentence *Martha is a beautiful minister*, i.e. *Martha è una bella pastorella* (lit.: *Martha is a beautiful shepherdess*). Hence, it was interesting to test whether such output would be repeated or not. (2) was rendered a few times in the female form when translated by Google Translate. As a consequence, it was included so as to understand whether a wider context could increase the occurrence of the female form when translating. (3) was instead almost always rendered by Google Translate in the female form and sometimes also by DeepL, i.e. when the subject was a male, but the job title was collocated with the adjectives *sweet* and *beautiful*. Finally, (4) was often rendered in the male form by Google Translate whereas DeepL translated it in the female form when the subject was a male but collocated with the adjective *beautiful*.

Selected adjectives were (1) *beautiful*, (2) *sweet*, (3) *proud* and (4) *determined*. As previously mentioned, adjectives were selected on the basis of some literature (Heilman 2012, Prates et. al 2018, Serafini 2013) and also divided into female (1, 2) and male (3, 4) stereotyped. These four adjectives were included in the third corpus because they resulted as being among the most effective when collocated with professions. Indeed, when the subject of a sentence was a female, typically male considered job titles were more likely to be translated in the female

form if occurring with the aforementioned adjectives than when occurring with others. Moreover, when translated with DeepL, sentences with male subjects and female-dominated occupations were sometimes rendered in the female form if they contained the adjectives *beautiful* and *sweet*.

As a consequence, the corpus comprises 32 sentences that are structured as follows: *Her name is Martha and she is from London. She is the most beautiful minister of her city and really enjoys her job. She also gets on well with her colleagues*. Pronouns are found many times and the adjective is in the superlative form, thus also making it clearer where adjectives ending in *-e* (actually *sweet*) are used in the male or in the female form in the target language.

4.4 Evaluation Process

The evaluation process took about three weeks. It consisted of different steps: (1) analysis of the two translated versions of the first corpus (2) analysis of two translated versions of the second corpus, including the creation of five subcorpora, (3) analysis of the two translated versions of the third corpus and (4) comparison among the collected data. (1) and (3) consisted in collecting numerical data about the amount of correct translations both in general and for each of the aforementioned professions. In order to carry out such tasks, different tables were created and are to be found in the fifth chapter.

During the whole evaluation process, data were analysed with the help of Sketch Engine. Sketch Engine is both a "corpus query tool" and a "corpus web service" which enables users to analyse pre-installed corpora or to upload one's own (Kilgariff et al. 2014: 7f). This tool has different functions, among them there are (1) word sketches, i.e. clearly structured tables that summarise a word's grammatical and collocational properties, (2) concordances and (3) a thesaurus based on the collocations present in the selected corpus (Kilgariff et al. 2014: 9f). As shown in Figure 15, Sketch Engine's concordance function was used in order to speed up the analysis of the translated corpora, since it can look out for a word and groups it every time this occurs in the corpus. As a result, terms are clearly represented, and it is very easy to gather data regarding the translation of professions.

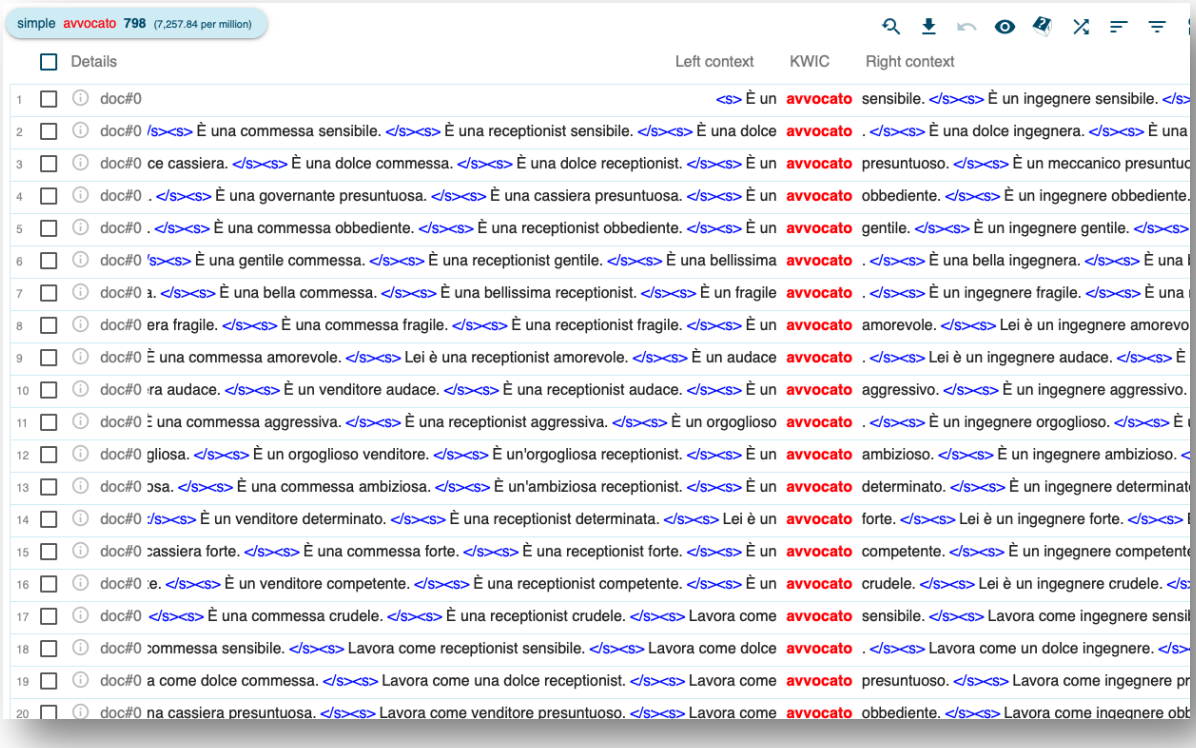


Figure 15: Example of the use of Sketch engine during the study (Sketch Engine)

Moreover, in order to double-check the results, the corpora were also opened as pure text format and the search function was used. In order to understand better the whole evaluation process, the aforementioned steps (2) and (4) must be discussed with greater attention.

Regarding the analysis of the translated corpora containing adjectives, the focus was firstly on the translation of professions. Therefore, a table was created in order to gather data on how many times a certain job title was rendered in the female or male form depending on the subject of the sentence and on the typology of adjective co-occurring (see Table 9). Furthermore, such a table included a box for taking notes. For instance, striking results as well as adjectives affecting translation outputs were reported. Results about outputs of every profession were then added together in order to obtain more general data.

Table 9: Example of how the analysis of the second corpus was conducted. Results about the translation outputs for lawyer by DeepL

PROFESSION	SUBJECT	RESULTS ACCORDING TO ADJECTIVES	NOTES:
Lawyer	Female subject	24/200 (female stereotyped adjectives)	DeepL renders 40 times the form avvocatessa. Among them, 23 with the adjective *beautiful*, 6 with *determined*, 5 with *cruel*.
		16/200 (male stereotyped adjectives)	
	Male subject	200/200 (female stereotyped adjectives)	
		200/200 (male stereotyped adjectives)	

Secondly, five subcorpora were also created in order to examine in depth the effect of adjectives on translation outputs. Two were produced by selecting sentences from the Google Translate version of the second corpus and three from the DeepL version. Indeed, it was noted that MT is particularly biased against women, as some professions, i.e. mainly male-dominated professions, are rarely translated in the female form when sentences contain a female subject. As a consequence, sentences comprising a female subject and such correctly translated professions were selected, thus creating two subcorpora (one for Google Translate and one for DeepL). More specifically, the first Google Translate subcorpus includes all the correct translations of the male-dominated professions listed in chapter 4.3.1 as well as those of the occupation *salesperson* in the female form. Bias against women was also found in sentences containing this job title. The total size of this test set is 826 sentences. The DeepL corpus contains, on the other hand, only male-dominated professions and a total of 375 sentences. These datasets made it possible to quantitatively determine which adjective influenced translation outputs the most. Furthermore, the same process was carried out for the female-dominated occupations that presented bias against men. Hence, two additional corpora containing sentences with male subjects and such correctly translated job titles in the male form were created. The second Google Translate subcorpus contains therefore 82 sentences whilst the DeepL one comprises 478. Finally, a fifth subcorpus was created, since it was noted that DeepL generally translated female-dominated professions in the male form when sentences had a male subject except when they were co-occurring with certain adjectives. This last dataset contains 199 sentences. The following table sums up the content of these subcorpora:

Table 10: Overview of the subcorpora content and size

Subcorpus	Professions with occurrence	Size
Translation of professions biased against women (Google Translate)	*Engineer* (5), *Mechanic* (73), *Mathematician* (235), *Deputy* (134), *Bricklayer* (191), *Architect* (14), *Salesperson* (174)	826
Translation of professions biased against women (DeepL)	*Lawyer* (40), *Mechanic* (184), *Mathematician* (92), *Deputy* (31), *Bricklayer* (28)	375
Translation of professions biased against men (Google Translate)	*Nurse* (11), *Babysitter* (30), *Housekeeper* (41)	82
Translation of professions biased against men (DeepL)	*Housekeeper* (119), *Secretary* (99), *Cashier* (260)	478
Exceptions biased against men (DeepL)	*Nurse* (32), *Babysitter* (30), *Cosmetologist* (18), *Nutritionist* (19), *Psychologist* (22), *Salesperson* (14), *Receptionist* (64)	199

The next step involved the comparison of the data. Results of the Google Translate versions of the first two corpora were compared in order to make observations on the effect of adjectives on translation outputs of professions. The same was also carried out for the DeepL versions of the first two corpora. Furthermore, performances delivered by the two tools were also compared and finally, on the basis of the collected results, the third corpus was created, translated and evaluated, thus marking the end of the study.

To conclude, the three corpora provided large test sets for the analysis of the automated translation of professions as well as of the impact of adjectives on gender bias in MT. Descriptive statistics were used to gather and describe data on the occurrence of (mis)translations, both when occupations were collocated with adjectives and when they were

not. Descriptive statistics are used in order to organize, display and describe data, which can be either qualitative or quantitative (Shafer & Zhang 2012: 9f). As a consequence, the translation outputs were simply quantified, evaluated and collected so as to describe the gathered data. As explained in Chapter 3.1, this differs from other studies in the field of gender bias in MT (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019), as deep statistical analysis was not carried out.

4.5 Research Validity: Advantages and Disadvantages of the Selected Methodology

After having accurately described the selected methodology, the advantages and disadvantages of such are presented. On the one hand, similar studies in this field have been carried out and such methodology allows for a concrete approach to a likewise concreate research object. Besides, the size of the datasets was considerable. On the other hand, the study is quite brief, focusing on a language pair and on two NMT systems. The data was not always clearly evaluable and a simple statistical analysis was carried out.

As already mentioned, the current work is inspired by previous research conducted in this field (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019), which is one of the main advantages. Indeed, the existence of such studies, which are all similar, can be regarded as a proof of the research validity. Obviously, the present work has not the same validity claims, since it is a master's thesis and I have worked for a limited time, with the help of my supervisor. Another advantage is that it is concrete in nature. The research object is not a subjective matter. In fact, it can be described through quantitative data. In this sense, the steps reported in this chapter demonstrate how the described method enables us to answer the research question concretely because the research problem is likewise concretely approached. Besides, the creation of three different corpora allowed for the examination of the research problem from different perspectives, namely whether MT is biased in general, whether adjectives do have an impact on the translation of professions or whether the context is actually crucial. It should also be underlined that the second corpus contains 16,000 sentences, thus being far bigger than the test set used, for instance, by Cho et al. (2019) which counted little more than 4,000 sentences.

As for the disadvantages, since this is a brief study, it must be noted that it exclusively focuses on the language pair EN-IT, hence not providing a general overview of gender bias in

MT in the field of the Romance languages or of the most spoken grammatical gender idioms. Moreover, only two NMT systems were tested, which are however considered as the best on the market (DeepL n.d.b). The main disadvantage of the selected methodology lies in the selected data for the creation of the corpora. Indeed, some professions as well as adjectives included in the corpora have only one form for male and for female in Italian, thus making a perfect evaluation of the data impossible. However, datasets were big enough so that it was possible to state that many sentences were clearly translated either in the female or in the male form and to make generalisations about the errors of the NMT systems when translating and it must be noted that professions and adjectives were selected on the basis of literature (Heilman 2012, Prates et al. 2018, Serafini 2013, Tesfaselassie 2019) because they were related to gender roles and/or stereotypes. Finally, the selected methodology foresees no deep statistical analysis (for instances, no index to assess gender bias was created nor null hypothesis was verified as opposed to the before mentioned studies). The sample was simply described.

Nevertheless, due to the similarity to previous research, it can still be assumed that the selected methodology is valid because the phenomenon of gender bias in MT has actually been studied by translating sets of sentences with occupations and/or adjectives. Furthermore, the studies mentioned as well as the current work obtained similar results, namely that MT are to some extent biased.

5. Results

In the current chapter results of the conducted study are objectively presented. As the analysis carried out involved different test sets it is useful to describe the structure of this chapter in detail. It is composed of three main sections, namely the first dedicated to Google Translate, the second to DeepL and the third to the comparison between their performances. More specifically, in the first and second section data collected by translating with Google Translate and DeepL are presented. This means that results from the cleaning of the first corpus, namely the one without adjectives, and the second containing adjectives are presented. Furthermore, data on how often professions were rendered in the correct gender form are provided and described from different points of view. Ultimately, a focus on the effect of adjectives on the translation outputs is provided through the analysis of some subcorpora, as detailed in Chapter 4.4. They were created by gathering data from both Google Translate and DeepL output versions of the corpus containing adjectives. In addition, findings on the third corpus, namely the corpus with longer sentences, are reported so as to present data on the influence of a wider context on the NMT systems employed in this study. Finally, the third section deals with an in-depth comparison between the performances delivered by Google Translate and DeepL.

5.1 Gender Bias in Google Translate

As mentioned in Chapter 4.3, Google Translate and DeepL did not always produce optimal outputs. Inconsistencies or mistranslations were found quite frequently in the first and especially in the second corpus, thus requiring them to be cleaned in order for the dataset to present a structure as coherent as possible. When inconsistencies were due to English terms referring to different concepts and therefore being translated into different Italian terms, the most common Italian term was selected and the others were removed from the translated versions of the corpora. This is the case of the profession *deputy* that has a double meaning. Otherwise, i.e. when inconsistencies occurred because some professions as well as some adjectives can be translated by using synonyms (e.g. *housekeeper*, *bold*, *proud*), all the alternatives were kept. Mistranslations in the outputs were removed. The cleaning of the translated version of the corpora was carried out in order to analyse only correct sentences. A summary of such inconsistencies and mistranslation is therefore provided.

In the case of occupations, only inconsistencies were found. As mentioned above, the occupation *deputy* was translated differently due to its double meaning. Indeed, a *deputy* is both the substitute of a person fulfilling a certain role or his/her second as well as a political representative (Merriam-Webster n.d.). Hence, in Italian different terms are used, namely, *vice*, *sostitut** and *deputat**. In the corpus without adjectives, *vice* appeared 37 times out of 50. Actually, it was also the most recurring translation also in the corpus with adjectives, but *sostitut** was also found 46 times. Nevertheless, *deputat** was chosen for the analysis as (1) in the everyday language it is the most common occupation among the three alternative translations, (2) it ends in *-o* or *-a* depending on its gender, thus easily allowing for the analysis of the correct forms used and (3) the decision was to focus on a specific translation of the profession. Furthermore, *housekeeper* was rendered both as *governante* (respectively 33 and 770 times) and *domestica* (respectively 17 and 30). However, since such terms are synonyms, one was not selected over the other.

Not only were adjectives inconsistently translated, but they were also mistranslated. As a consequence, the translations of the second corpus presented different problematic outputs that were therefore removed. The adjective *conceited* was rendered 272 times as *presunt** (*alleged*), three times as *supposta* (*supposed*) and once *supponente* (*opinionated*), instead of *presuntuos**. More specifically, *presunta* appeared 145 times and it was often collocated with *mayor* (19), *deputy* (19) as well as with *receptionist* (25 + 2 with a male subject, namely *Liam*). *Presunto* appeared instead 127 times, often collocated with *mayor* (24 + 3 with a female subject), *deputy* (13 + 2 with a female subject), *bricklayer* (11 + 4 with a female subject), *receptionist* (14). *Supposta* appeared three times whilst *supponente* once. It was collocated with *deputy* all the times it appeared and the sentences contained female subjects. However, in the case of *supponente*, deputy was rendered in the male form. The adjective *loving* too was mistranslated three times as *amata* (*beloved*) – namely when collocated with *deputy* in a sentence with a female subject – and 27 times as *amante*. *Amante* means actually *loving*, but it is a gerund form and is very uncommon in Italian. Therefore, such output sentences were removed. However, *amante* was used only when sentences included female subjects and occupations like *mathematician* (twice in the female form and twelve times in the male form), *blacksmith* (nine times out of nine as *fabbro*, i.e. the male form) and *bricklayer* (four times out of four as *muratore*, namely male form). Surprisingly, when collocated with the job title *receptionist*, it was translated three times as *accogliente* (*welcoming*). I personally thought this was a striking and particularly optimal output. Nevertheless, I decided to erase those output sentences as the meaning differs from the source as well as from the other translations produced

by Google Translate. It must be noted that the subjects were two females, namely *Martha* and *Adele*, and a male, namely *Liam*. However, in this latter case, the sentence was translated in the female form. Some other adjectives and professions as well were mistranslated or inconsistently rendered, even though to a lesser extent. *Sensitive* was rendered six times as *delicata* (*delicate*) when collocated with the profession *housekeeper*. Subjects were five times female and only once a male, namely Liam. In this case as well, the sentence was translated in the female form. *Beautiful* was rendered 6 times as *meraviglioso* (*wonderful*) when collocated with *engineer* (3) and with *mechanic* (3) whilst once as *splendido* (*gorgeous*) when collocated with *receptionist*. Subjects were males. *Engineer* was mistranslated six times as *ingegnos** (*clever*) when collocated with the adjective *conceited*. It was rendered in the female form only once out of three times when in a sentence with a female subject. *Bold* was translated 9 times as *grassetto*, meaning *bold* (the font). It was once collocated with *deputy* – this sentence had a female subject – and eight times when collocated with the job title *housekeeper* and male subjects. However, it was rendered 990 times as *audace* and only once as *coraggiosa*, but such output was not removed from the translated version of the corpus containing adjectives, since *audace* and *coraggios** are synonyms. Besides, the sentence *Martha is a beautiful minister* was translated as *Martha è una bella pastorella* (*shepherdess*). Such result is particularity striking as the profession was rendered with a totally different term. Finally, *kind* was once rendered as *specie di* meaning *kind of* in the sentence *Sarah is a kind blacksmith* and the adjective *proud* was inconsistently translated, namely as *fier** (75) and *orgoglios** (925), which are synonyms. As summarised in Table 11, 552 sentences were removed from the Google Translate version of the second corpus, i.e. the 15,448 remaining sentences were then analysed.

Table 11: Data on the cleaning of the corpus containing adjectives translated by Google Translate

Word	Mistranslations or inconsistency	Removed sentences
Deputy	Inconsistency: vice, sostitut*	213
Conceited	Mistranslations: presunt*, supposta, supponente	276
Loving	Mistranslations: amata, amante, accogliente	33
Sensitive	Mistranslations: delicata	6
Beautiful	Mistranslations: meraviglioso, splendido	7

Bold	Mistranslation: in grassetto	9
Engineer	Mistranslation: ingenos*	6
Minister	Mistranslation: sheperdess	1
Kind	Mistranslation: specie di	1

Consequently, having had detailed data filtered from the first and the second corpus, the actual results of the study can be presented in detail. However, for the moment, the focus is on these corpora whilst the third is presented later on in this chapter.

Professions occurring in the sentences without adjectives were translated correctly 543 times out of 820, i.e. 66.21% of the times. Even though the first corpus consisted of 963 sentences, only 820 were taken into account. This is due to the fact that some job titles, namely *babysitter*, *nutritionist*, *receptionist* and *housekeeper*, when rendered as *governante*, have only a single form for both male and female in Italian too. Furthermore, sentences did not always include an article, thus making the distinction between genders impossible. Since every profession occurred 50 times in the corpus (*deputy* excluded), the data set was not large enough to generalise about Google Translate presenting a female or male default. As a consequence, only sentences with an explicit gender form were assessed in the evaluation process. The same problems in terms of recognising gender forms were observed in the second corpus as well. In fact, also part of the adjectives did not present a recognisable form, as they end in *-e*. Nevertheless, every profession occurred about 800 times (the amount can vary because of cleaning), therefore the dataset was quite large, thus allowing more reliable generalisations. As a consequence, professions were translated correctly 10,689 times out of 15,448, i.e. 69.19% of the times. As shown in Table 12, the output's correctness rose by about 2.98 percentage points when using adjectives. Those results must be described in more details since quite interesting phenomena were noted during the study.

Table 12: General results on output correctness (Google Translate)

	Sentences without adjectives	Sentences with adjectives	Difference %
Correctly translated professions	543/820* (66.21%) *due to the small size of the corpus, only sentences with an explicit gender form were included	10,689*/15,448 (69.19%) *due to the large size of the corpus, generalisations were made	+2.98%

As explained in Chapter 4.3.1, the corpora were produced by mixing different elements, first of all female and male subjects. It was noted, as previous research also suggested (Prates et al. 2018), that Google Translate produces different outputs depending on the gender of the subjects. Indeed, as shown in Table 13, professions were translated correctly respectively 184 times out of 410 (44.87%) and 4,295 times out of 7,674 (55.96%) when the subject was a female. When the subject was a male, professions were translated correctly respectively 359 out of 410 (87.59%) and 6,394 times out of 7,774 (82.24%). By using adjectives, sentences with a female subject were translated correctly about 11.09% more often whereas the outputs correctness of sentences having male subjects decreased to about 5.32 percentage points.

Table 13: Output correctness for subject's gender (Google Translate)

	Sentences without adjectives	Sentences with adjectives	Difference %
Professions translated in the female form when the sentence has a female subject	184/410 (44.87%)	4,295/7,674 (56.96%)	+11.09%

Professions translated in the male form when the sentence has a male subject	359/410 (87.56%)	6,394/7,774 (82.24%)	-5.32%

Furthermore, it must be noted that one of the male subjects, namely *Liam*, was a source of bias as it was found that sentences containing such subject were more often rendered in the female form than those having other male subjects regardless of the occupations occurring. This means that Google Translate quite often rendered job titles for which the male form is known and used in the female one when the subject was *Liam*. As a consequence, as shown in Table 14, in the corpus without adjectives out of 820 sentences analysed, *Liam* appears 86 times and 21 of which were biased. In the corpus with adjectives, *Liam* appears 1,549 times, 443 are surely biased. Indeed, some sentences could not actually be recognised as female or male.

Table 14: Bias with Liam in the corpora (Google Translate)

	Sentences without adjectives	Sentences with adjectives
Liam appearing biased	21/86 (24.41%)	443*/1,549 (28.59%) *these are only the certain biased sentences. In the other cases it could not always be stated whether bias was present or not and therefore they were not included in the calculation of the biased sentences.

Nevertheless, it was also noted that results obtained varied across occupations, i.e. some job titles were generally translated better than others. As mentioned in Chapter 4.3.1, a distinction can be made between male and female-dominated professions.

Male-dominated professions tended to be translated less frequently in the female form than female-dominated ones. However, there was one exception, namely *salesperson,* which is considered as a female-dominated profession, was found to include bias against female

subjects. As a consequence, when reporting the data about male-dominated professions *salesperson* is always included and it is excluded from female-dominated occupations. The former typology of job titles was translated correctly respectively 287 times out of 513 (55.94%) for the output corpus without adjectives and 5,041 out of 8,402 (59.99%) for the output corpus with adjectives. As shown in Table 15, the difference in percentage points is quite remarkable, namely 4.05%. Female dominated professions were optimally rendered respectively 256 times out of 307 (83.38%) and 5,800 out of 7,046 (82.31%), thus resulting in a tiny negative difference, namely 1.07 percentage points, when adjectives were used. However, it is interesting to look at such results by dividing them according to the gender of the subjects.

Table 15: Output correctness for typology of occupation (Google Translate)

	Sentences without adjectives	Sentences with adjectives	Difference %
male-dominated professions (+ salesperson)	287/513 (55.94%)	5,041/8,402 (59.99%)	+4.05%
female-dominated professions (- salesperson)	256/307 (83.38%)	5,800/7,046 (82.31%)	-1.07%

When analysing sentences depending on whether the subjects are male or female and on the typology of occupation occurring (i.e. male or female-dominated), it was found that the use of adjective could improve Google Translate performances only in one case, i.e. when considering sentences containing female subjects and a male-dominated profession (always including *salesperson*). Such sentences were translated correctly 26 times out of 252 (10.31%) when they did not contain an adjective and 826 out of 4,160 (19.85%) when they were collocated with one. As a consequence, output correctness actually rose in 9.54% of cases. Results are rather different when such job professions were to be translated in the male form because of the presence of a male subject in the sentence. As shown in Table 16, translations were correct 261 out of 261 (100%) and 4,225 out of 4,242 (99.59%). In the case of female dominated professions, sentences were correctly translated in the female form respectively 158 times out

of 158 (100%) and 3,469 out of 3,514 (98.71%). The use of adjectives slightly worsened the output (-1.29 percentage points). When they were to be translated in the male form, the outputs were correct 98 times out of 149 (65.77%) and 2,221 out of 3,532 (62.88%) with quite a worsening, namely in 2.89% of cases when using adjectives. Finally, another distinction is to be made, i.e. between adjectives related to female and male stereotypes. Indeed, it was noted that some adjectives affected machine translation outputs more than others.

Table 16: Output correctness for subject's gender and typology of occupation (Google Translate)

		Sentences without adjectives	Sentences with adjectives	Difference %
Professions translated in the female form when the sentence has a female subject	male-dominated profession (+ *salesperson*)	26/252 (10.31%)	826/4,160 (19.85%)	+9.54%
	female-dominated profession (- *salesperson*)	158/158 (100%)	3,469/3,514 (98.71%)	-1.29%
Professions translated in the male form when the sentence has a male subject	male-dominated profession (+ *salesperson*)	261/261 (100%)	4,225/4,242 (99.59%)	-0.41%
	female-dominated profession (- *salesperson*)	98/149 (65.77%)	2,221/3,532 (62.88%)	-2.89%

The division between adjectives related to female and male roles proposed in Chapter 4.3.1 was found to be relevant for the current study. Firstly, when analysing sentences with a female subject and a male-dominated profession (including *salesperson*) it was found that adjectives

related to a female role – 478 correct sentences out of 2,050, i.e. 23.33% of the times – affected the translation output more than those related to a male role – 348 correct sentences out of 2,110, i.e. 16.49% of the times – hence with a difference of 6.84 percentage points. Nevertheless, the use of adjectives in general improved Google Translate output, as shown in Table 17. Furthermore, when it came to the translation of female-dominated professions, no difference in the results was found when using different kind of adjectives. Indeed, out of 1,714 containing a female-stereotyped adjective, 1,691 were translated correctly (98.65%) whereas out of 1,800 containing a male-stereotyped adjective, 1,778 were correct (98.77%).

Table 17: Output correctness for female subjects, typology of profession and typology of adjective (Google Translate)

FEMALE SUBJECT	Without adjectives	Adjective related to a female role	Adjective related to a male role
Male-dominated profession (+ salesperson)	26/252 (10.31%)	478/2,050 (23.33%)	348/2,110 (16.49%)
Female-dominated profession (- salesperson)	158/158 (100%)	1,691/1,714 (98.65%)	1,778/1,800 (98.77%)

Secondly, when analysing sentences with male subjects, male-dominated professions and adjectives, no particular difference was found with regard to the effect of female or male stereotypes. In the first case, 2,095 sentences out of 2,109 (99.33%) were translated in the male form whereas in the second the correct sentences were 2,130 out of 2,133 (99.85%). As shown in Table 18, differences were found when it came to the translation of female-dominated professions. Out of 1,740 sentences containing a female-stereotyped adjective, 1,031 were translated correctly (59.25%) whereas out of 1,792 sentences including a male-stereotyped adjective, the correct ones were 1,138 (63.50%). Hence, differing in 4.25 percentage points.

Table 18: Output correctness for male subjects, typology of profession and typology of adjective (Google Translate)

MALE SUBJECT	Without adjectives	Adjective related to a female role	Adjective related to a male role
Male-dominated profession (+ salesperson)	261/261 (100%)	2,095/2,109 (99.33%)	2,130/2,133 (99.85%)
Female-dominated profession (- salesperson)	98/149 (65.77%)	1,031/1,740 (59.25%)	1,138/1,792 (63.50%)

Furthermore, as explained in Chapter 4.4, two subcorpora were created. The first subcorpus was created by selecting correctly translated sentences in the female form from the output corpus containing adjectives. Two were the requirements for the selection, namely that sentences had to contain (1) a female subject and (2) one of the professions that turned out to be biased against women. Indeed, as already explained in the course of the chapter, some professions, i.e. male-dominated professions and the occupation *salesperson*, were always or very often rendered in the male form, thus revealing the presence of bias in Google Translate. Table 19 shows the frequencies of the extent to which Google Translate rendered these job titles in the female form and such data, namely 826 sentences, were used to analyse the effect of adjectives on the translation outputs.

Table 19: Frequencies of correctly translated male-dominated professions in the female form (Google Translate)

Professions with data on the correct outputs in the female form when sentences contained a female subject		Deputy 134/242		
Lawyer 0/400	Mayor 0/376	Engineer 5/397	Mechanic 73/400	Mathematician 235/382
Blacksmith 0/390	Bricklayer 191/393	Architect 14/400	Minister 0/396	Salesperson 174/394

The purpose of this analysis was to understand which adjectives could make Google Translate use the female form for the above listed professions that were generally found to be biased against women, thus having an impact on the translation outputs and providing additional data in order to answer the research question and evaluate the resulting hypothesis. As a consequence, a histogram in Figure 16 shows the frequencies of adjectives occurring in the subcorpus as number whereas a pie chart in Figure 17 depicts the percentage of their recurrence.

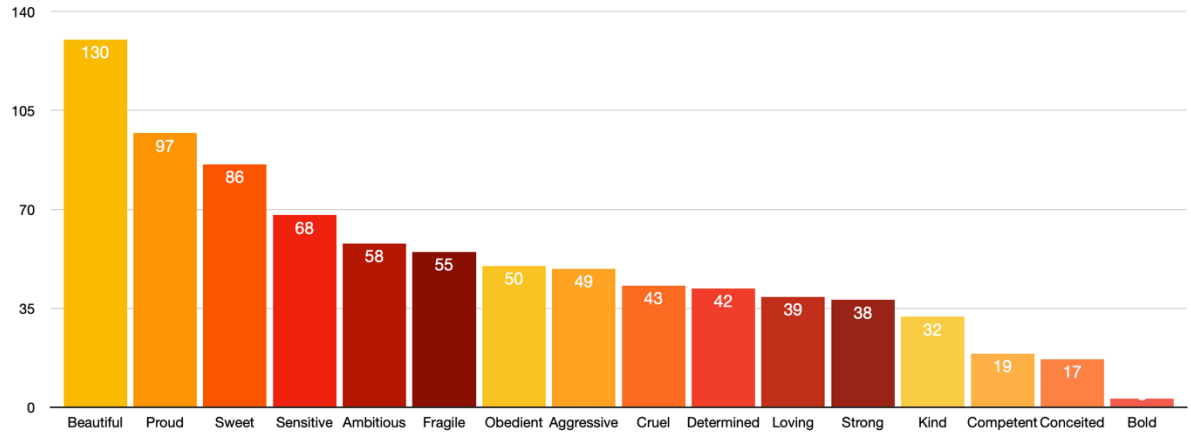


Figure 16: Frequencies (number) of adjectives occurring when male-dominated professions were translated in the female form (Google Translate)

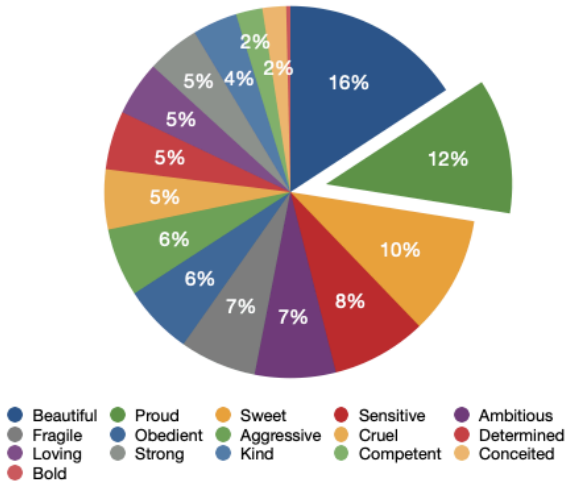


Figure 17: Frequencies (%) of adjectives occurring when male-dominated professions were translated in the female form (Google Translate)

It was found that *beautiful* was the most recurring adjective when a male-dominated profession was translated in the female form, namely 130 times out of 826 (about 16% of cases) as depicted in Figure 16 and 17. The second most recurring adjective was *proud* (97, about 12%), followed by *sweet* (86, about 10%) and *sensitive* (68, about 8%). As a consequence, it was also found that adjectives relating to a female role occurred more often, namely 477 times, than those relating to a male role (349 times).

To allow for a deeper investigation, a second subcorpus containing correctly translated sentences in the male form was created. In this case, the requirements for the selection of the sentences were different. They had to contain (1) a male subject and (2) one of the professions

that presented bias against men. On the basis of the analysis carried out, the professions biased against men were *nurse*, *babysitter* and *housekeeper*, hence this second subcorpus being smaller than the first one and made up of 82 sentences (11 with *nurse*, 30 with *babysitter* and 41 with *housekeeper*). Furthermore, it does not contain all of the adjectives present in the main corpus. This is also due to the fact that *babysitter* and *housekeeper* also have a form in Italian too (*babysitter* or *baby-sitter* and *governante*) for male and female, thus requiring adjectives ending in *-o* or in *-a* in order to assess the form of the professions (when these professions where collocated with the other adjectives ending in *-e*, the gender form was hence not explicit recognisable as neither the adjective nor the job title were marked in terms of gender). A histogram in Figure 18 shows the frequencies of the adjectives recurring in this second subcorpus.

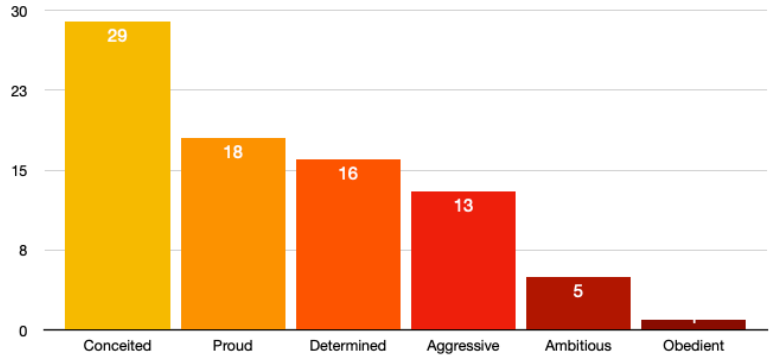


Figure 18: Frequencies of adjectives occurring when professions presenting bias against men were correctly translated in the male form (Google Translate)

The most influential adjective was found to be *conceited*, it appeared namely 29 times, followed by *proud*, *determined* and *aggressive* which occurred respectively 18, 16 and 13 times. Having presented such results, findings on the third corpus can be now discussed. This time adjectives related to a male role were found more often, i.e. 52 times, than those relating to a female stereotype, i.e. 30 times.

The third corpus was created in order to understand whether a wider context would affect the translation output and it was therefore created by carefully selecting its elements, as explained in Chapter 4.3.3. In the second corpus, *minister* was never rendered in the female form and the same result was also obtained with a wider context. *Architect* was rendered 14 times out of 400 in the female form (seven times when collocated with the adjective *beautiful*, five with *sweet*, one with *obedient* and *cruel*). However, results did not improve when enlarging the sentences and *architect* was never rendered in the female form in the third corpus. Better

outputs were produced when it came to the translation of *salesperson* (4 times out of 4 correct). As aforementioned, Google Translate presented bias against women when sentences contained this job title, namely only 174 sentences out of 394 were rendered in the female form (25 times when collocated with *beautiful*, 23 *cruel*, 22 *sweet*, 20 *proud*, 17 *sensitive*, 15 *loving*, 14 *ambitious*). Besides, it was also always rendered in the male form (4/4) with Liam as the subject in the third corpus. In the second corpus, mistranslations of *salespersons* concerning the male gender occurred only 5 times out of 391, but always when Liam was the subject of such sentences. Finally, bias was still found in the translation of *nurse*, which was rendered in the male form only once out of four (in the second corpus eleven times out of 400, but never with *Liam* as subject). Results of the analysis of the third corpus are summarised in Table 20.

Table 20: Results from the analysis of the third corpus by Google Translate

Profession	Martha	Liam
Minister	0/4	4/4
Architect	0/4	4/4
Nurse	4/4	1/4
Salesperson	4/4	4/4

5.2 Gender Bias in DeepL

DeepL too did not always produce optimal outputs. Hence, they were sometimes inconsistent or wrong. Also in this case, mistranslations were removed from the output corpora and when inconsistencies were due to the double meaning of a term, only one Italian referent was selected for the analysis and the other filtered from the results. More specifically, *deputy* was wrongly translated or not by using different terms due to its double meaning, whilst *sweet* was sometimes mistranslated when combined with salesperson and *receptionist*. In general, inconsistencies and mistranslations occurred far less than in the corpora translated by Google Translate. A brief overview of them is presented.

The professions, *deputy*, *cosmetologist*, *salesperson* and *blacksmith* were mistranslated. *Deputy* was rendered respectively 13 and 37 times as *vicesceriffo* (*undersheriff*) and *vice* in the translation of the first corpus. It was rendered 393 times as *vicesceriffo*, 268 as *vice*, four as

delegato (*representative*), three as *sostituto* and once as *villano* (*peasant*) in the second corpus, hence being also mistranslated. *Cosmetologist* was rendered ten times as *estetista* (always when collocated with the adjective *competent*, seven times with a female subject), which is actually the correct translation for the professions, as a *cosmetologo* is actually a *cosmetic chemist*. Nevertheless, both terms were kept since also Google Translate always rendered the professions as *cosmetologo* and does not seem to know the terminological difference between English and Italian. *Salesperson* was translated five times (always when the subject was female, four times when collocated with the adjective *competent*) as *addetta alle vendite*, which is a synonym. Finally, *blacksmith* was once rendered as *maniscalco* (*farrier*) in the sentence *she is a kind blacksmith*.

Regarding the adjectives, DeepL mistranslated *sweet* 20 times when combined with *salesperson* and twice with *receptionist* respectively as *venditore/venditrice di dolci* (someone who sells sweets) and *receptionist di dolci*. Loving was translated five times as *affettuoso* and twice as *amoroso* which are synonyms of the most used term *amorevole*. Bold was rendered as audace (972 times) or as coraggios* (23 times) but also mistranslated five times as the font *grassetto*. Proud was translated as *orgoglios** (848) or *fier** (152) whilst *competente* was once rendered as *bravo* (*capable*). Consequently, 697 sentences were removed from the first corpus, hence the actual size being 15,303 sentences. The following table sums up the mistranslations or inconsistencies that were removed from the corpus including adjectives before it was analysed:

Table 21: Data on the cleaning of the second corpus translated by DeepL

Word	Mistranslations or inconsistency	Removed sentences
Deputy	Inconsistencies: vice, vicesceriffo, sostituto, delegato Mistranslation: villano	669
Blacksmith	Mistranslation: maniscalco	1
Sweet	Mistranslation: di dolci	22
Bold	Mistranslation: in grassetto	5

Having cleaned both the DeepL versions of the first and second corpora, the results of their analysis can be now detailed. As already explained in the case of Google Translate, not every

profession could be clearly recognised as being translated in the female or male form since some job titles have only one form, namely *babysitter*, *nutritionist*, *housekeeper* and *receptionist*. As a consequence, the sentences analysed in the corpus without adjectives were 790 out of 950. Out of them, 557 were correctly translated, i.e. 70.50% of cases. As regards the corpus with adjectives, generalisations about DeepL presenting a female or male default were made when the co-occurrence of the aforementioned professions with adjectives ending in *-e* made the distinction between the two analysed genders impossible. This was done due to the large size of the corpus. Out of 15,303 sentences, 10,982 were correct, i.e. 71.76% of cases. As shown in Table 22, when using adjective outputs quality increased by 1.26 percentage points, a slight difference. However, the results are completely different if the gender of the subjects is taken into account.

Table 22: General results on the output correctness (DeepL)

	Sentences without adjectives	Sentences with adjectives	Difference %
Correctly translated professions	557/790* (70.50%) **deputy* was never translated as *deputato* + due to the small size of the corpus, only sentences with a clear form were included	10,982*/15,303 (71.76%) *due to the large size of the corpus, generalisations were made	+1.26%

When sentences had a female subject, they were translated correctly respectively 184 times out of 395 (46.58%) and 4,253 out of 7,654 (55.56%). Consequently, the use of adjectives improved the output quality considerably, namely 8.98 percentage points. However, the opposite was verified when analysing sentences containing male subjects (see Table 23). Translations of the first corpus were very precise, namely 373 correct sentences out of 395 (94.43%). Regarding the second corpus, 6,729 professions out of 7,649 (87.97%) were rendered in the male form. This means that adjectives have a negative impact on the translation when the subject is a male,

with a worsening in 6.46% of cases. Furthermore, also this time differences were found in the translation of female and male-dominated professions, which confirmed the stereotypical categorisation of job titles presented in Chapter 4.3.1.

Table 23: Output correctness for subject's gender (DeepL)

	Sentences without adjectives	Sentences with adjectives	Difference %
Professions translated in the female form when the sentence has a female subject	184/395 (46.58%)	4,253/7,654 (55.56%)	+8.98%
Professions translated in the male form when the sentence has a male subject	373/395 (94.43%)	6,729/7,649 (87.97%)	-6.46%

Firstly, DeepL performed worse when translating male-dominated professions. Indeed, it translated respectively 293 sentences out of 450 (53.11%) and 4,040 out of 7,326 (55.14%) correctly. In this case, the use of adjectives slightly improved translation outputs, namely in 2.03% of cases (see Table 24). When it came to female-dominated professions, DeepL performed much better, i.e. it translated 318 sentences out of 340 (93.25%) and 6,942 out of 7,997 (86.80%) correctly. A remarkable worsening of the outputs was observed with the use of adjectives, namely in 6.72% of cases. Nevertheless, it was also noted that such results differed substantially when also taking subjects into account.

Table 24: Output correctness for typology of occupation (DeepL)

	Sentences without adjectives	Sentences with adjectives	Difference %
male-dominated professions	239/450 (53.11%)	4,040/7,326 (55.14%)	+2.03%
female-dominated professions	318/340 (93.52%)	6,942/7,997 (86.80%)	-6.72%

Indeed, DeepL performed poorly when male-dominated professions were to be translated in sentences containing a female subject, i.e. 14 sentences out of 225 (6.22%) and 375 out of 3,661 (10.24%) were rendered in the male form. The use of adjectives improved translation outputs, namely in 4.02% of cases. When male-dominated professions were to be translated in sentences containing a male subject, DeepL did not make any mistakes, as shown in Table 25. On the contrary, female-dominated professions were translated quite correctly when the subject of the sentences was a female, namely 170 out of 170 (100%) and 3,878 out of 3,993 (97.11%). Nevertheless, when using an adjective DeepL performances slightly worsened (-2.89 percentage points). When the subject was a male, 148 sentences out of 170 (87.05%) and 3,064 out of 3,984 (76.90%) were rendered in the male form. The decrease is remarkable, i.e. in 10.15% of cases. In one case, also the typology of adjectives played an important role concerning translation outputs.

Table 25: Output correctness for subject's gender and typology of occupation (DeepL)

		Sentences without adjectives	Sentences with adjectives	Difference %
Professions translated in the female form when the sentence has a female subject	male-dominated profession	14/225 (6.22%)	375/3,661 (10.24%)	+4.02%
	female-dominated profession	170/170 (100%)	3,878/3,993 (97.11%)	-2.89%
Professions translated in the male form when the sentence has a male subject	male-dominated profession	225/225 (100%)	3,665/3,665 (100%)	-
	female-dominated profession	148/170 (87.05%)	3,064/3,984 (76.90%)	-10.15%

As previously noted, male-dominated professions were rarely translated in the female form when the sentences had a female subject, but the use of adjectives improved translation outputs. However, the distinction presented in Chapter 4.3.1 between adjectives related to female and male roles was only partially confirmed. Indeed, 189 sentences containing female-stereotyped adjectives out of 1,818 (10.39%) were translated correctly. Those containing male-stereotyped adjectives were correct 186 times out of 1,843 (10.09%). A slight difference (4.64 percentage points) was observed when it came to female-dominated professions, as shown in Table 26. Sentences containing such professions and an adjective related to a female role were rendered 1,982 times out of 1,993 (99.44%) in the female form whilst those including an adjective related to a male role 1,896 out of 2,000 (94.80%).

Table 26: Output correctness for female subjects, typology of profession and typology of adjective (DeepL)

FEMALE SUBJECT	Without adjectives	Adjective related to a female role	Adjective related to a male role
Male-dominated profession	14/225 (6.22%)	189/1,818 (10.39%)	186/1,843 (10.09%)
Female-dominated profession	170/170 (100%)	1,982/1,993 (99.44%)	1,896/2,000 (94.80%)

Male-dominated professions were always correctly translated when the sentences had male subjects (see Table 27). A greater difference (14.62 percentage points) with regards to the effect of the typology of co-occurring adjective was noted in the translation of female-dominated professions with male subjects. Indeed, when an adjective related to a female role was present, DeepL performed worse (1,380 correctly translated sentences out of 1,985, namely 69.62% of cases) than when an adjective related to a male role was found (1,684 correctly translated sentences out of 1,999, namely 84.24% of cases). As a consequence, three subcopora were created in order to study better the effect of adjectives on the translation outputs.

Table 27: Output correctness for male subjects, typology of profession and typology of adjective (DeepL)

MALE SUBJECT	Without adjectives	Adjective related to a female role	Adjective related to a male role
Male-dominated profession	225/225 (100%)	1,821/1,821 (100%)	1,844/1,844 (100%)
Female-dominated profession	148/170 (87.05%)	1,380/1,985 (69.62%)	1,684/1,999 (84.24%)

As explained in Chapter 4.4, the subcorpora were created in order to assess the effect of adjectives on the translation outputs of DeepL. As carried out with the Google Translate data,

the first subcorpus was created by selecting correctly translated sentences in the female form from the DeepL output version of the corpus containing adjectives. The selection was carried out on the basis of two requirements: (1) the sentences had to contain a female subject and (2) one of the professions that were found to be biased against women. An overview of how often every profession was rendered in the female form when occurring with a female subject is to be found in Table 28. Such data, namely 375 sentences, make up the first DeepL subcorpus.

Table 28: Frequencies of correctly translated male-dominated professions in the female form (DeepL)

Professions with data on the correct outputs in the female form				
Lawyer 40/400	Mayor 0/400	Engineer 0/400	Mechanic 184/400	Mathematician 92/400
Blacksmith 0/399	Bricklayer 28/400	Architect 0/400	Minister 0/400	Deputy 31/62

Again, the purpose was to find which adjectives could make the translation system, in this case DeepL, use the female form for the above listed professions that were generally found to be biased against women. The histogram in Figure 19 and the pie chart in Figure 20 show the frequencies of the adjectives occurring with the aforementioned professions when they were correctly rendered in the female form in numbers and in percentages, respectively

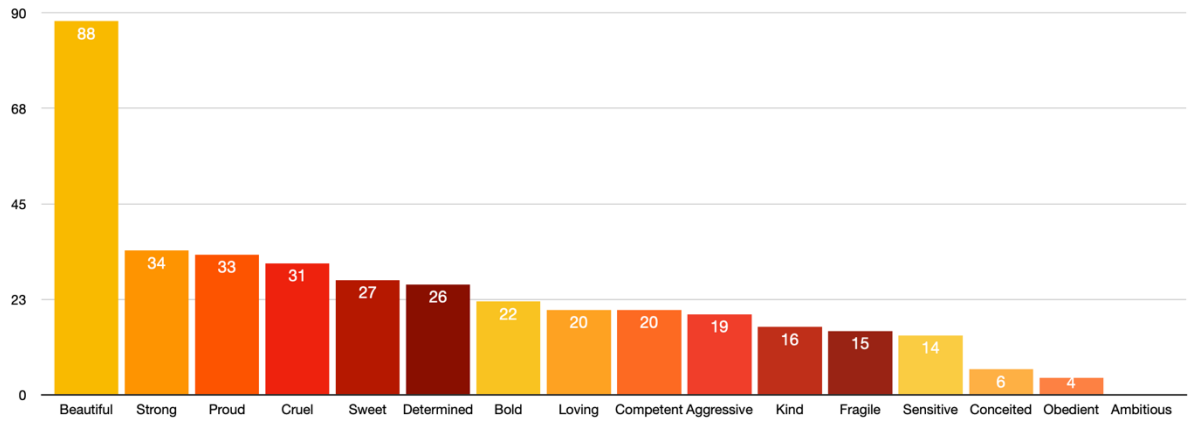


Figure 19: Frequencies of adjective (numbers) when male-dominated professions were correctly translated in the female form (DeepL)

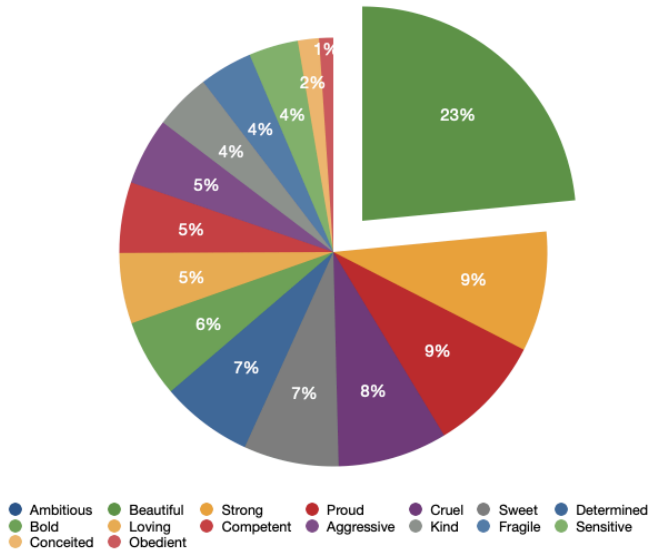


Figure 20: Frequencies of adjective (%) when male-dominated professions where correctly translated in the female form (DeepL)

It was found that *beautiful* was by far the adjective with the most effect when male-dominated professions were translated in the female form (88 times out of 375, about 23% of cases) as can be seen in Figure 19 and 20. It was followed by *strong* (34, about 9%), *proud* (33, 9%), *cruel* (31, 8%), *sweet* (27, 7%). *Ambitious* was never found in this subcorpus. It was also found that adjectives relating to female roles were slightly more frequent (195 times) than those relating to male roles (190 times).

Furthermore, it was noted that DeepL as well as Google Translate very often rendered some female-dominated professions in the female form even though the subject of the sentence was male. As a consequence, a second subcorpus was created by selecting sentences from the DeepL output version of the corpus including adjectives. This second subcorpus contained 478 sentences collected on the basis of two requirements: they had to include a male subject and one of the professions that were biased against men, namely *housekeeper* (119), *secretary* (99) and *cashier* (260) that were correctly rendered in the appropriate gender form. The analysis concentrated once again on the occurrence of adjectives in such sentences so as to understand which of them had an impact on translation outputs. Figures 21 and 22 show a histogram and a pie chart with the frequencies of the adjectives recurring in this subcorpus.

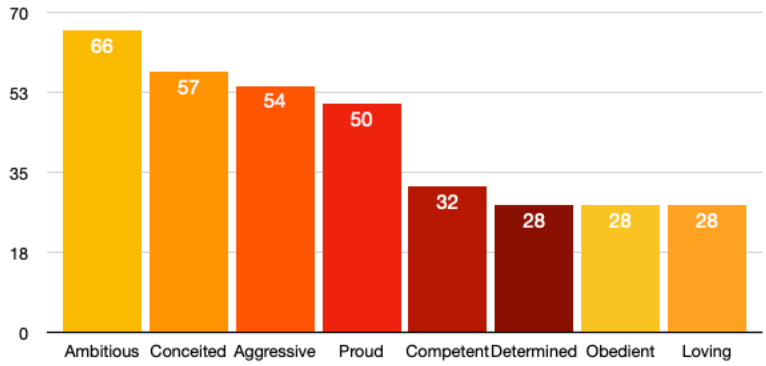


Figure 21: Frequencies of adjective (numbers) when professions presenting bias against women where correctly translated in the male form (DeepL)

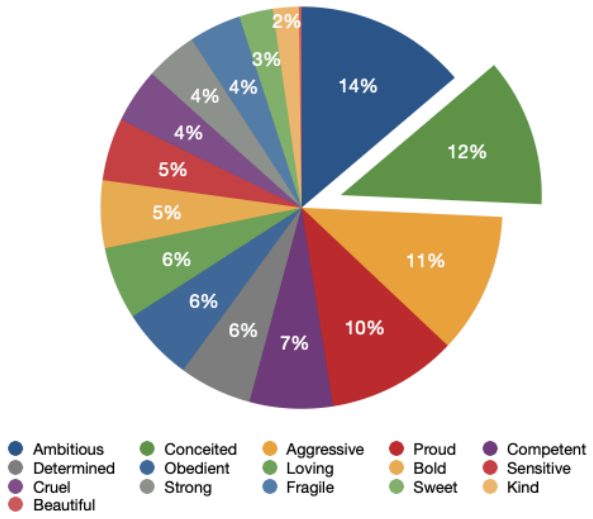


Figure 22: Frequencies of adjective (%) when professions presenting bias against women where correctly translated in the male form (DeepL)

The most recurring adjective when sentences included a female-dominated profession and a male subject was *ambitious* (66 times out of 478, i.e. about 14% of cases), followed by *conceited* (57, about 12%), *aggressive* (54, about 11%) and *proud* (50, about 10%). *Beautiful* was found only once. It was found that adjectives relating to a male role occurred much more frequently, namely 297 times, than those relating to a female role, namely 181 times.

Finally, as explained in Chapter 4.4, it was noted that DeepL generally rendered female-dominated professions, except for the already discussed *housekeeper*, *secretary* and *cashier*, in the correct form when the subject of the sentence was male. Nevertheless, it was also observed that such professions were sometimes rendered in the female form when collocated with certain

adjectives. As a consequence, a third subcorpus was created on the basis of the following requirements: sentences had to include (1) a male subject and (2) one of the female-dominated professions that were normally rendered in the appropriate form. However, (3) only the sentences that were rendered in the female form were selected in order to create the third subcorpus. Table 29 shows the extent to which DeepL mistranslated the abovementioned professions when the subject of the sentences was male.

Table 29: Frequencies of mistranslations of female-dominated professions that did not generally present bias against men

Mistranslations (199/2,785)	Nurse 32/400	Babysitter 30/400	Cosmetologist 18/400
Nutritionist 19/400	Psychologist 22/400	Salesperson 14/386	Receptionist 64/399

The histogram in Figure 23 shows the frequencies of the adjectives occurring in this subcorpus. *Beautiful* was the most recurring adjective when such professions were mistranslated, namely 144 times out of 199 (72% of cases). Other adjectives found in the corpus were *sweet* (14), *aggressive* (14), *conceited*, *determined*, *obedient* (9), *proud* (5), *ambitious* and *kind* (2). This means that adjectives relating to a female role occurred much more frequently, namely 174 times, than those relating to a male role, namely 25 times.

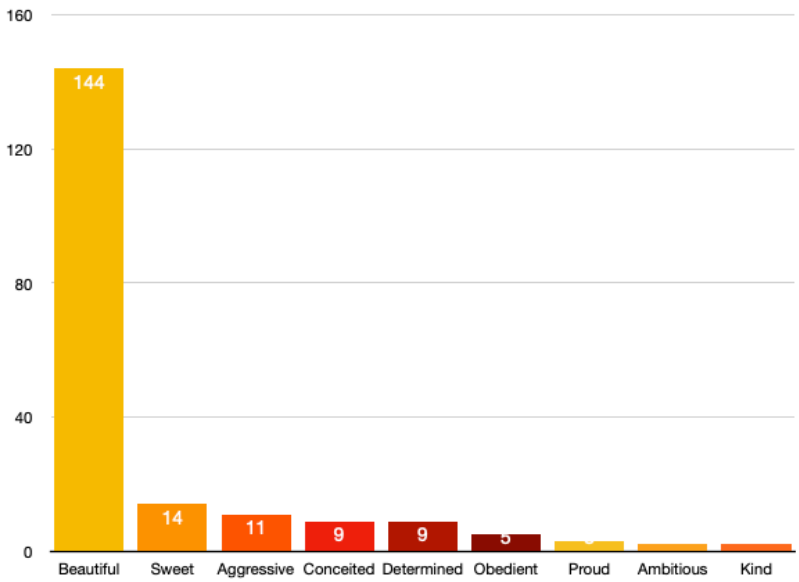


Figure 23: Results from the analysis of the third DeepL subcorpus

After the creation and analysis of the three subcorpora, the DeepL version of the third corpus was also analysed. As regards the sentences with *Martha* as subject, the wider context did not play any role in changing the translation outputs. Indeed, *minister* and *architect* were not rendered in the female form as was found in the analysis of the second corpus. *Nurse* and *salesperson* were instead translated correctly. As regards Liam, *minister* and *architect* were rendered as usual in the male form. However, when it came to the translation of *nurse* and *salesperson*, the inclusion of longer sentences improved the outputs as they were translated correctly. On the contrary, in the second corpus *nurse* was rendered in the female form every time when collocated with the adjective *beautiful* (25/25) and sometimes also when collocated with *sweet* (7/25). *Salesperson* as well was rendered in the female form every time when collocated with *beautiful* (14/14). Results from the analysis of the third corpus are reported in the following table:

Table 30: Results from the analysis of the third corpus by DeepL

Profession	Martha	Liam
Minister	0/4	4/4
Architect	0/4	4/4
Nurse	4/4	4/4
Salesperson	4/4	4/4

5.3 Results Compared

In Chapter 5.1 and 5.2 the results obtained from the analysis of the three corpora and the five subcorpora created for the current study with both Google Translate and DeepL were presented. As a consequence, a comparison between the performances delivered by the two NMT systems is provided. In order to do so, the data collected from Google Translate and DeepL are juxtaposed in this section and they are finally discussed in the next chapter.

Regarding the translation of professions in general, DeepL delivered better performances than Google Translate. As shown in Table 31, the system rendered job titles in the correct form 70.50% of the time when sentences did not include an adjective, thus outperforming its competitor in 4.29% of cases. However, such a difference declined when sentences included adjectives at 2.57 percentage points. Actually, Google Translate was more likely to produce better outputs with the use of adjectives (+2.98 percentage points) than DeepL (+1.26 percentage points).

Table 31: General results for output correctness - comparison between Google Translate and DeepL

Correctly translated professions	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	543/820 (66.21%)	10,689/15,448 (69.19%)	+2.98
DeepL	557/790 (70.50%)	10,982/15,303 (71.76%)	+1.26
Difference between performances (pp)	+4.29 D	+2.57 D	+1.72 GT

DeepL also slightly outperformed Google Translate in the translation of professions in the female form when sentences included an adjective, namely in 1.71% of cases. Nevertheless, Google Translate outputs remarkably improved when an adjective was included in the sentences (+11.09 percentage points), hence producing better results than the competitor in 1.40% of times. As shown in Table 32, when it came to the translation of job titles in the female form, DeepL as well benefitted from the use of adjectives, i.e. in 8.98% of cases.

Table 32: Output correctness for female subjects – comparison between Google Translate and DeepL

Professions translated in the female form + female subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	184/410 (44.87%)	4,295/7,674 (56.96%)	+11.09
DeepL	184/395 (46.58%)	4,253/7,654 (55.56%)	+8.98
Difference between performances (pp)	+1.71 D	+1.40 GT	+2.11 GT

DeepL also correctly translated professions in the male form more often than its competitor did. More precisely, the system performed better both when the corpora did not (+6.87 percentage points) and when they contained adjectives (+5.73 percentage points). However, as shown in Table 33 the use of adjectives negatively affected both translation tools for the male form with male subject, but this time the decline was more remarkable in DeepL outputs, namely -1.14 percentage points in comparison with Google Translate.

Table 33: Output correctness for male subjects – comparison between Google Translate and DeepL

Professions translated in the male form + male subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	359/410 (87.56%)	6,394/7,774 (82.24%)	-5.32
DeepL	373/395 (94.43%)	6,729/7,649 (87.97%)	-6.46
Difference between performances (pp)	+6.87 D	+5.73 D	-1.14 D

If the translation of male-dominated professions is considered, results differ. Google Translate rendered them more often in the correct form when the sentences did not contain adjectives, namely in 2.83% of cases. Furthermore, this difference increased when adjectives were used (+4.85 percentage points). As shown in Table 34, both translation systems gained advantage from the use of adjectives, but Google Translate was found to be more receptive to them (+2.02% of cases).

Table 34: Output correctness for male-dominated professions - comparison between Google Translate and DeepL

Male-dominated professions	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	287/513 (55.94%)	5,041/8,402 (59.99%)	+4.05
DeepL	239/450 (53.11%)	4,040/7,326 (55.14%)	+2.03
Difference between performances (pp)	+2.83 GT	+4.85 GT	+2.02 GT

DeepL, on the other hand, rendered female-dominated professions correctly more often than Google Translate did, namely in +10.14% of cases when adjectives were not present in the corpus. Nevertheless, as shown in Table 35, adjectives caused many mistranslations in terms of gender, especially in DeepL (-5.65% of times more than Google Translate), hence the difference between the results produced by the two systems falling to 4.49% of cases.

Table 35: Output correctness for female-dominated professions - comparison between Google Translate and DeepL

Female-dominated professions	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	256/307 (83.38%)	5,800/7,046 (82.31%)	-1.07
DeepL	318/340 (93.52%)	6,942/7,997 (86.80%)	-6.72
Difference between performances (pp)	+10.14 D	+4.49 D	-5.65 D

When male-dominated professions were to be translated in the female form because sentences included a female subject, Google Translate performed better than DeepL, i.e. in +4.09% of cases when sentences did not include adjectives. As shown in Table 36, in the second corpus the difference increased into 9.61 percentage points. In fact, the use of adjectives improved the Google Translates outputs in 9.54% of cases whilst DeepL results got better only in 4.02% of times.

Table 36: Output correctness for female subjects and male-dominated professions - comparison between Google Translate and DeepL

Male-dominated professions in the female form with a female subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	26/252 (10.31%)	826/4,160 (19,85%)	+9,54
DeepL	14/225 (6.22%)	375/3,661 (10,24%)	+4,02
Difference between performances (pp)	+4.09 GT	+9.61 GT	+5.52 GT

When female-dominated professions were to be translated in the female form because sentences included a female subject, both translation systems produced optimal outputs. However, as shown in Table 37, the use of adjectives negatively affected the results, especially those produced by DeepL (-2.89 percentage points) that were slightly worse, namely in 1.60% of cases.

Table 37: Output correctness for female subjects and female-dominated - comparison between Google Translate and DeepL

Female-dominated professions in the female form with a female subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	158/158 (100%)	3,469/3,514 (98.71%)	-1,29
DeepL	170/170 (100%)	3,878/3,993 (97.11%)	-2,89
Difference between performances (pp)	-	+1.60 GT	-1,60 D

When male-dominated professions were to be translated in the male form because sentences included a male subject, both systems produced optimal outputs as well. As shown in Table 38, Google Translate produced rarely worse results when sentences included adjectives, namely in 0.41% of cases, whilst DeepL was not affected.

Table 38: Output correctness for male subjects and male-dominated - comparison between Google Translate and DeepL

Male-dominated professions in the male form with a male subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	261/261 (100%)	4,225/4,242 (99.59%)	-0.41
DeepL	225/225 (100%)	3,665/3,665 (100%)	-
Difference between performances (pp)	-	+0,41% D	-0.41 GT

When female-dominated professions were to be translated in the male form because sentences included a male subject, there were remarkable differences between DeepL and Google Translate. The former outperformed its competitor in 21.28% of cases. However, its results were affected much more significantly, and they got worse in 10.15% of cases. As shown in Table 39, the outputs produced by Google as well were negatively affected but to a far lesser extent, namely in 2.89% of cases.

Table 39: Output correctness for male subjects and female-dominated - comparison between Google Translate and DeepL

Female-dominated professions in the male form with a male subject	Sentences without adjectives	Sentences with adjectives	Difference (pp)
Google Translate	98/149 (65,77%)	2,221/3,532 (62,88%)	-2,89
DeepL	148/170 (87,05%)	3,064/3,984 (76,90%)	-10,15
Difference between performances (pp)	+21,28 D	+14,02 D	-7,26 D

As already mentioned, the use of adjectives affected Google Translate outputs more than DeepL when male-dominated professions were to be translated in the female form. Furthermore, there was also a difference in the typology of adjectives that affected such outputs. Indeed, when male stereotyped job titles were rendered in the female form, adjectives related to a female role were more likely to co-occur (23.33% of cases) than those related to a male role (16.49% of cases). As shown in Table 40, this difference was not observed in DeepL results.

Table 40: Output correctness for female subjects, male-dominated professions and typology of adjectives - comparison between Google Translate and DeepL

Male-dominated professions in the female form with a female subject	Sentences without adjectives	Adjective related to a female role	Adjective related to a male role
Google Translate	26/252 (10.31%)	478/2,050 (23.33%)	348/2,110 (16.49%)
DeepL	14/225 (6.22%)	189/1,818 (10.39%)	186/1,843 (10.09%)

In the cases of female-dominated professions being translated in the female form, the use of adjectives had a slightly negative impact on both NMT systems. This time, the use of adjectives related to a male role seemed to impact DeepL performances more. As shown in Table 41, the typology of adjectives co-occurring with the professions did not play any role when working with Google Translate. The correctness of DeepL outputs marked a decrease of 4.24 percentage points when female-dominated job titles co-occurred with adjectives related to a male role.

Table 41: Output correctness for female subjects, female-dominated professions and typology of adjective form - comparison between Google Translate and DeepL

Female-dominated professions in the female form with a female subject	Sentences without adjectives	Adjective related to a female role	Adjective related to a male role
Google Translate	158/158 (100%)	1,691/1,714 (98.65%)	1,778/1,800 (98.77%)
DeepL	170/170 (100%)	1,982/1,993 (99.44%)	1,896/2,000 (94.80%)

As shown in Table 42 and already discussed, both NMT systems correctly translated male-dominated professions in the male form and the use of adjectives did, to a very small extent, partially affect Google Translate outputs. Nevertheless, no difference regarding the typology of adjective co-occurring was observed.

Table 42: Output correctness for male subjects, male-dominated professions and typology of adjectives - comparison between Google Translate and DeepL

Male-dominated professions in the male form with a male subject	Sentences without adjectives	Adjective related to a female role	Adjective related to a male role
Google Translate	261/261 (100%)	2,095/2,109 (99.33%)	2,130/2,133 (99.85%)
DeepL	225/225 (100%)	1,821/1,821 (100%)	1,844/1,844 (100%)

Greater differences were noted when it came to the translation of female-dominated professions in the male form. As aforementioned, both translation tools were affected by the use of adjectives and, as shown in Table 43, both were influenced especially by those related to a male role. In the case of Google Translate, such adjectives were found 4.25% more times than those related to a female role. In DeepL, the difference was greater, namely, 14.62 percentage points.

Table 43: Output correctness for male subjects, female-dominated professions and typology of adjectives - comparison between Google Translate and DeepL

Female-dominated professions in the male form with a male subject	Sentences without adjectives	Adjective related to a female role	Adjective related to a male role
Google Translate	98/149 (65.77%)	1,031/1,1740 (59.25%)	1,138/1,792 (63.50%)
DeepL	148/170 (87.05%)	1,380/1,985 (69.62%)	1,684/1,999 (8424%)

A comparison of the co-occurrence of specific adjectives is also proposed by juxtaposing the results obtained from the analysis of the first two subcorpora created in order to determine the influence of adjectives on the translation outputs. For the first subcorpus, namely the one created by selecting sentences correctly rendered in the female form but including male-dominated professions that were found to be usually biased, it was observed that in both cases *beautiful* was the adjective which had the most effect, respectively in 16% and 23% of cases. *Proud* also recurs quite often, being the second and the third most influential adjective. As shown in Table 44, the co-occurrence of adjectives differs quite considerably for the rest.

Table 44: Adjective co-occurrence in the first subcorpus - comparison between Google Translate and DeepL

	Google Translate (826 sentences)		DeepL (375 sentences)	
Most recurring adjectives in the first subcorpus	Beautiful (16%)	Cruel (5%)	Beautiful (23%)	Competent (5%)
	Proud (12%)	Determined (5%)	Strong (9%)	Aggressive (5%)
	Sweet (10%)	Loving (5%)	Proud (9%)	Kind (4%)
	Sensitive (8%)	Strong (5%)	Cruel (8%)	Fragile (4%)
	Ambitious (7%)	Kind (4%)	Sweet (7%)	Sensitive (4%)
	Fragile (7%)	Competent (2%)	Determined (7%)	Conceited (2%)
	Obedient (6%)	Conceited (2%)	Bold (6%)	Obedient (1%)
	Aggressive (6%)	Bold (>1%)	Loving (5%)	Ambitious (0%)

In the second corpus, namely the one with correctly translated sentences in the male form that included generally biased female-dominated professions, *conceited* was respectively the most and the second most recurring adjective. As shown in Table 45, *ambitious* was also found quite often in both cases. Besides, it must be noted that *proud* was very effective, but this adjective was also found in the first subcorpus.

Table 45: Adjective co-occurrence in the second subcorpus - comparison between Google Translate and DeepL

	Google Translate (82 sentences)	DeepL (478 sentences)	
Most recurring adjectives in the second subcorpus	Conceited (35%)	Ambitious (14%)	Bold (5%)
	Proud (22%)	Conceited (12%)	Sensitive (5%)
	Determined (20%)	Aggressive (11%)	Cruel (4%)
	Aggressive (16%)	Proud (10%)	Strong (4%)
	Ambitious (6%)	Competent (7%)	Fragile (4%)
	Obedient (1%)	Determined (6%)	Sweet (3%)
		Obedient (6%)	Kind (2%)
		Loving (6%)	Beautiful (>1%)

Finally, it should be pointed out that the adjective *beautiful* was found to be very effective in DeepL. As a matter of fact, it was found 144 times out of 199 when an otherwise not biased female-dominated profession was rendered in the female form regardless of the male subject. Such a phenomenon was not observed in Google Translate.

6. Discussion

In this chapter, the most meaningful results are finally discussed. In particular, general gender bias was found in both Google Translate and DeepL. However, the latter presented more bias, especially against women, as shown in Chapter 5 and explained in this chapter. In addition, it was found that adjectives either improved, i.e. when male-dominated professions were to be rendered in the female form, or worsened, i.e. when female-dominated professions were to be rendered in the male form, translation outputs. Tendencies regarding the main assumptions of the current work, namely that the gender role related to the adjective impact on how MT renders occupations, were partially observed. Furthermore, it was also noted that a wider context could improve translation outputs only if the NMT systems show bias but are quite familiar with both gender forms of the occupations which are to be translated. Such implications are further illustrated in the next subchapter along with some reflections on the research process.

6.1 Gender Bias in Google Translate and DeepL

In order to interpret the results presented in the fifth chapter, it is useful to discuss the data gathered from different perspectives, i.e. depending on the gender of the subjects, on the typology of occupations and adjectives included in the sentences. Firstly, it was noted that both Google Translate and DeepL failed to render occupations in the right gender to a different extent according to the subjects.

Since both NMT systems failed in almost a third of all cases, be it in the first or in the second corpus, to render the right gender of professions, it can be stated that they present general gender bias. However, as also previous research found (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019) women were much more often the object of linguistic discrimination than men. In fact, professions were translated in the female form only about half of the times and both NMT systems performed quite poorly. DeepL performed slightly better when adjectives were not included in the sentences whilst Google Translate performances improved when adjectives were used. On the contrary, on analysing the translation of professions in the male form, there was found to be less bias, especially in DeepL. However, it must be noted that Google Translate often mistranslated professions when sentences contained the subject *Liam* considering him a woman. Therefore, it is of relevance to note that not only did the subject play

a role in the correct translation of professions but the typology of job occupation (male and female dominated) did as well.

Indeed, a division into male- and female-dominated professions was proposed in Chapter 4.3.1 and it was found that such a division was actually reflected by the translation systems, with a little difference. *Salesperson* was considered a female-dominated profession but found to be biased against women in Google Translate, being correctly rendered in the female form only 174 times out of 394. As also Prates et al. (2018) highlighted, it was found that the translation systems were more likely to translate male-dominated professions in the male form, even though the subject was female, in almost 80-90% of cases depending on the system and on the corpus analysed. For instance, out of 400 times occurring with a female subject, *lawyer* was never found in the female form in the second Google Translate corpus and only 40 in the DeepL one. Since DeepL performs considerably worse (4.09% of all cases) on correctly translating sentences with female form and female subject when a male dominated profession occurs in the sentences, it could be considered more biased then Google Translate. In fact, it is only when taking such data into account that it can be stated that DeepL is more biased than Google Translate. Furthermore, both systems rendered female-dominated professions correctly in the female form without particular problems, except in isolated cases.

As for bias against men, both systems performed almost perfectly when it came to the translation of male-dominated professions. However, gender bias was found when female-dominated professions were to be rendered in the male form. In this case, both translation systems performed much worse (Google Translate delivered false outputs in about 30% of cases whilst DeepL from about 15% to 23% depending on the corpus, i.e. the one without adjectives and the one containing adjectives). In particular, only some of the female-dominated professions were the source of wrong outputs when sentences included a male subject, namely *nurse*, *babysitter*, *housekeeper* regarding Google Translate and *housekeeper*, *secretary* and *cashier* regarding DeepL. As a consequence, bias against men was found to a far lesser extent (i.e. in about a third of all cases) than bias against women. This is not surprising though. As a matter of fact, it was explained in Chapter 2.2.1 that a feature of sexist languages is their asymmetry (Litosseliti 2013: 15), namely that some words describing men or women have no counterpart. Nevertheless, as Pauwels (2003: 553) notes, the phenomenon of lexical gaps concerns more women than men, as the lack of words describing men in certain roles and/or occupations tends to be filled more quickly. This seems to be the reason why male forms for female-dominated occupations are more likely to be used by NMT systems than female forms for male-dominated occupations. Our language and hence the corpora used for the training of

machines are more biased against women than men. Another reason one could consider might be that gender bias still prevents women from occupying job positions which are considered male-dominated as women are actually discriminated against (Heilman 2012, Grover 2015, Rifkin 2015), hence Google Translate and DeepL reproduce a workplace bias. However, as described in Chapter 3.1, this hypothesis was rejected by Prates et al. (2018). The scholars demonstrated that Google Translate does not reproduce the actual ratio of female workers to male ones, but it actually exacerbates gender bias. As a consequence, it seems more reasonable to explain the phenomenon of gender bias against women by referring to a sexist use of language, so as described in Chapter 2.2.1. As a matter of fact, Telve (2011) notes that roles like *ministro* (*minister*), *sindaco* (*mayor*), *deputato* (*deputy*) have the female form ending in -*a*, but for historic reasons they have always been used in the male form, thus explaining the strong bias against women and towards men found in both the tested NMT systems.

6.2 The Influence of Adjectives on Translation Outputs

The main topic of the current work is obviously gender bias in the field of machine translation. Its main focus was to study the effect of adjectives on translation outputs, hence answering the research question "How do adjectives collocated with nouns affect the machine translation of professions?". The hypothesis formulated in Chapter 1.2 was "in general, MT correctly renders gender-dominated professions in the female or male form if the gender stereotype related to the adjectives preceding such job titles is consistent with the gender of the subject". This means that when the subject is a female and the adjective relates to a female role and/or stereotype a male-dominated profession is more likely to be translated in the female form. Tendencies regarding such assumptions were partially observed in the gathered data. However, in order to explain such a statement, the results must be carefully discussed.

First of all, adjectives collocated with professions actually affect machine translation outputs in two different ways: either professions are more likely to be rendered in the right gender or the opposite is to be found. Female forms linguistically benefit from the use of adjectives in machine translation. As a matter of fact, Google Translate rendered professions in the female form less than half of the times when the corpus did not include adjectives whereas the output correctness increased by 11.09 percentage points when the corpus included adjectives. DeepL performed similarly. Its outputs were also correct in less than half of all cases

and correctness increased by 8.98 percentage points when adjectives were used. More specifically and interestingly in terms of bias against women, adjectives had quite a strong positive impact solely on the translation of male-dominated professions whilst a slightly negative impact on those that are considered female-dominated was observed. Indeed, a general improvement was observed when sentences contained a female subject, a male-dominated profession plus an adjective. On the contrary, the NMT systems delivered slightly worse performances when it came to the translation of sentences having a female subject, a female-dominated profession plus an adjective.

When translating male-dominated professions in the male form, adjectives did not play any role. As men, mistranslations produced by Google Translate are not due to the co-occurrence of adjectives but to the subject *Liam*, often considered as a woman. Data were quite different in the case of female-dominated professions. As a matter of fact, the use of adjectives affected negatively both NMT systems when such professions were to be rendered in the female form because the sentences contained a female subject. Correctness of outputs fell by 2.89 percentage points in Google Translate and even by 10.15 percentage points in DeepL. As a consequence, different conclusions can be drawn.

The use of adjectives positively affected translation outputs when sentences contained a female subject and a male-dominated occupation and negatively affected outputs when sentences contained a male subject and a female-dominated profession. In the other two cases (male subject plus male-dominated occupation and female subject plus female-dominated profession) they had no or little influence. Furthermore, tendencies regarding the assumptions that gender roles are more likely to affect machine translation outputs if consistent with the gender of the subjects were partially observed.

It was noted that in sentences with a female subject and a male-dominated profession Google Translate used the female form more frequently (in about 7% more cases) if the adjective included in the sentence was related to a female role than to a male one, thus gender of the subject and the gender stereotype corresponded. Such a tendency was also confirmed when the frequencies of the adjectives occurring in the first Google Translate subcorpus (i.e. the subcorpus with sentences containing female subjects and male-dominated occupations rendered in the appropriate form) was analysed. On the other hand, this difference was not observed when translating with DeepL, with both typologies of adjectives occurring more or less with the same frequencies (those related to female roles were slightly more frequent but not to a meaningful extent).

When sentences included a male subject and a female-dominated profession, it was noted that Google Translate was more likely to render such job titles in the male form if the adjectives contained in the sentences were related to a male role (i.e. 4,25% more often) than when the adjectives were related to a female role. This tendency was even more marked when analysing DeepL results. Male stereotyped adjectives co-occurred namely about 15% more often than female stereotyped ones.

It is also interesting to underline that adjectives had an impact on Google Translate and on DeepL performances to far different extents. The correctness of Google Translate outputs (+9,54 percentage points) increased more than that of DeepL (+4,02 percentage points) when translating sentences with a female subject and a male-dominated profession in the female form. On the contrary, when translating sentences containing a male subject and a female-dominated profession in the female form, DeepL results were affected more negatively (-10,15 percentage points) than those produced by Google Translate (-2,89 percentage points).

To conclude, tendencies regarding the initial hypothesis that "in general, MT correctly renders gender-dominated professions in the female or male form if the gender stereotype related to the adjectives preceding such job titles is consistent with the gender of the subject" were therefore generally observed. The analysis of the subcorpora that were specifically created to determine which adjectives had the greatest effect provides an interesting insight into the topic.

Five subcopora were created in order to analyse the effect of adjectives on translation outputs. Two (one for Google Translate and one for DeepL) were created by selecting correctly translated sentences with female subjects and male-dominated professions; another two (one for Google Translate and one for DeepL) were created by selecting correctly translated sentences with a male subject and female-dominated professions that were generally source of bias; the last one was created selecting false sentences with a male subject and the other female-dominated professions that were generally correctly rendered. When analysing the first two subcorpora, it can be assumed that *beautiful*, namely a female-stereotyped adjective, was the adjective which had the most effect when male-dominated professions were translated in the female form in sentences with female subjects. It was found respectively in 16% and 23% of such sentences translated by Google Translate and DeepL. Another very effective adjective in this regard was *proud* (12% and 9% of cases) that is considered as a male stereotyped one (see Chapter 4.3.1). Furthermore, it was stated before that adjectives related to female roles were found to be more effective in Google Translate. Indeed, *sweet*, *sensitive* and *fragile* were found among the most recurring in the Google Translate version of the subcorpus discussed.

Furthermore, *beautiful* can be regarded as the most influential adjective in general. It was actually observed that female-dominated professions did not present particular bias, except for *nurse*, *babysitter*, *housekeeper* as regards Google Translate and *housekeeper*, *secretary* and *cashier* using DeepL. In other cases, the NMT systems performed far better than when translating male-dominated professions in the female form. However, DeepL wrongly rendered 199 times the gender form of the otherwise unbiased female stereotyped job titles. Out of them, *beautiful* appeared 144 times, hence DeepL associated beauty to femaleness. This is not surprising as Chang and McKeown (2019), for instance, note that female celebrities are much more often related to appearance than male in American magazines. Besides, Prates et al. (2018) found that, when translating adjectives from genderless languages into English, MT changed the subject depending on the gender stereotype related to the attribute used in the source language. Indeed, *beautiful* was also one of these adjectives.

In the case of female-dominated occupations translated in the male form, it was observed that adjectives related to a male role were the most effective. Thus, tendencies regarding the assumptions of this work were observed, with adjectives such as *proud* (22% of cases) or *determined* (20%) and *aggressive* (16%) often being found in the Google Translate version of the second subcorpus as well as *ambitious* (14%), *aggressive* (11%) and *proud* (10%) in the DeepL version. Nevertheless, *conceited* was found as being respectively the most (35%) and the second most (12%) adjective having an effect.

Finally, results of the third corpus must also be discussed. Such a dataset was created in order to understand whether a wider context could actually improve MT outputs. Some tendencies were also observed in this case. Firstly, male-dominated professions (i.e. *minister* and *architect*) were not rendered in the female form even though the sentences were expanded and, for instance, contained many pronouns clearly indicating that the subject of the sentence was a female. This could be explained because such roles and jobs are not, or only extremely rarely, found in the corpora used for the training of the NMT systems – rarely because *architect* was rendered 14 times out of 400 in the female form in the Google Translate version of the second corpus, therefore the translation system is aware of the form. As mentioned above, Telve (2011) noted that some roles and/or occupations do have the female form ending in *-a* but they are not used. The same trend was observed in Google Translate with the translation of *nurse* in the male form, namely bias was still found after enlarging the context. On the contrary, when Google Translate was more familiar with the female form of the occupation, i.e. in the case of *salesperson* that was found to be biased against women but still correctly rendered 174 times out of 394, outputs improved and were always right with sentences including a wider context.

It can be concluded that a wider context positively affects translation outputs of professions only if NMT systems are quite familiar with such, meaning that both gender forms appear in the training corpora with a certain frequency.

6.3 Reflection on the Research Process

As already stated, the study confirmed the results found by previous research (Prates et al. 2018, Cho et al. 2019, Stanovsky et al. 2019), namely that NMT systems present gender bias, especially against women. This is to be found in the form of occupational bias, i.e. professions are not translated in the correct form, hence linguistically discriminating against women and/or men, and in the reproduction and perpetuation of gender stereotypes, for example the fact that women are more likely to be associated with certain characteristics and men with others. However, as also explained throughout the fifth chapter, results were obtained generalising about Google Translate and DeepL presenting a female or male default in the case of the corpus including adjectives.

As a matter of fact, the data selected for the creation of the corpora was not always optimal as not every profession or adjective had different forms for the female and the male gender. This is partially due to the fact that such nouns and adjectives were selected on the basis of literature (Hegewisch & Tesfaselassie 2019, Heilman 2012, Prates et al. 2018, Serafini 2013), in order to validate the choices made and it was not always possible to find good examples that were also grammatically suitable for the study. When it came to the creation of the third corpus, the adjectives were included in the superlative form so as to force the translation systems to include the article in every sentence, thus helping reveal the gender form of the professions. This could have been done prior to the study when creating the second corpus in order to collect more reliable data. Nevertheless, in my opinion the dataset was big enough and that the generalisations made about certain professions being translated in the female or male form were adequate, even though there is no certainty due to the lack of a specific gender form of adjective or article.

Furthermore, it was interesting to note that both NMT systems were very inconsistent in their translations. Sometimes the sentences were rendered in the correct gender form whereas sometimes only the verb expression (employed was rendered in Italian as *impegat**, having two forms) and the adjective or sometimes just only the verb expression were correctly rendered.

Some examples of the inconsistencies in the gender form can be found in the appendix to the current work. In addition, some striking results such as *Martha è una bella pastorella* (*shepherdess*) as the translation of *Martha is a beautiful minister* were also found, where the profession was not only biased but completely changed into a different profession. Another interesting aspect of the study was to observe the influence of the adjective *beautiful*. It was quite shocking to note how Google Translate and especially DeepL rendered many professions in the female form that otherwise would be translated in the male one. This was particularly striking because this was by far the most influential adjective in both NMT systems.

Finally, it was sometimes quite difficult to conduct the study due to the size of the datasets used. Indeed, the analysis of the two translations of the second corpus was quite time-consuming and required concentration because not only it how many times every profession was rendered in the correct gender form was counted, but also data on the co-occurrence of the adjectives, namely amount and typology, were collected. As a consequence, calculations were repeated twice or even three times so as to ensure the correctness of the results presented herein.

7. Conclusion

The starting point for the current work was that MT results are potentially gender biased. The focus was on the effect that adjectives have on translation outputs of professions, i.e. the purpose of this master's thesis was to understand whether MT translates occupations differently when they are collocated with adjectives related to gender roles and/or stereotypes. As a consequence, three opportunistic corpora were created. They all included simple sentences in which either a female or a male subject does a job. The second corpus also included adjectives whilst the third comprised a wider context so as to determine the impact of more information on translation outputs. These datasets were translated with Google Translate and DeepL from English into Italian. Data about how often professions were rendered in the correct gender form were collected and comparisons between the corpora as well as between the performances delivered by both NMT systems were drawn in order to determine the impact of adjectives and context.

The results of the study revealed that both Google Translate and DeepL are generally biased. More specifically, gender bias is not only to be found in form of occupational bias but also in the reproduction and perpetuation of gender stereotypes. As a matter of fact, (1) male-dominated professions were rarely translated in the female form whilst (2) female-dominated professions were much more often rendered in the male one. Hence, gender bias is stronger against women than men. Furthermore, it was observed that the use of adjectives (1) improved the correctness of translation outputs when male-dominated professions were to be rendered in the female form and (2) worsened them when female-dominated professions were to be rendered in the male form. As regards (1), adjectives more likely to influence translation outputs tended to be related to female roles when translating with Google Translate. On the contrary, when using DeepL, the same results were obtained with all adjectives. For instance, *beautiful* was the most effective adjective by far in both NMT systems when male-dominated professions were rendered in the female form. As regards (2), adjectives more likely to influence translation outputs tended to be related to male roles when translating with both Google Translate and DeepL. Finally, the inclusion of a wider context improved translation outputs only when the NMT systems were quite familiar with both gender forms of the professions otherwise, i.e. when the female or male form of an occupation was not or was rarely rendered by the translation system, inputs did not benefit from the extra information included in the sentences.

There are different research approaches that could be further developed in the future. Firstly, I think that it could be quite interesting to study whether adjectives referring to aspect/beauty affect MT and/or are therefore a source of bias. Secondly, as this study focused on the language pair EN-IT, it could be also interesting to draw a comparison with other pairs such as EN-GER or EN-FR, i.e. always translating from a natural gender language into a grammatical gender language. I think this could be interesting not only in order to assess gender bias in machine translation, but also to determine how (differently) biased such languages are. Another interesting field, even though this is may still be premature, would be bias against other genders. The current work is developed on the basis of a (false) binary division of gender into female and male. As already stated in Chapter 2.2, at least 64 different gender categories such as non-binary or gender fluid are recognised nowadays. Finally, while writing this master's thesis, I noted that biases in general have been studied in the field of NLP applications but not in the field of MT, so this would be an interesting research area for future work.

8. Bibliography

ALPAC (1966) *Languages and Machines: Computers in Translation and Linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, D.C.: National Academy of Sciences, National Research Council, 1966.

Bahdanau, Dzmitry; Cho, KyungHyun & Bengio, Yoshua (2014). Neural machine translation by jointly learning to align and translate. In: Bengio, Yoshua & LeCun, Yann (Eds.), *Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, California.*

Batara, Meryl Anna; Ngo, Jan Micole; See, Kayley Anne & Erasga, Dennis (2018). Second Generation Gender Bias: The Effects of the Invisible Bias Among Mid-Level Women Managers. *Asia-Pacific Social Science Review*, 18(2), 138-151.

Bem, Sandra L. & Bem, Daryl J. (1973). Does Sex-biased Job Advertising "Aid and Abet" Sex Discrimination? 1. *Journal of Applied Social Psychology*, 3(1), 6-18.

Bender, Emily M. & Friedman, Batya (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. In Matsumoto, Yuji (Eds.), *Transactions of the Association for Computational Linguistics*, 6, 587-604

Blumberg, Rae Lesser (2008). The invisible obstacle to educational equality: Gender bias in textbooks. *Prospects*, 38(3), 345-361

Bolukbasi, Tolga; Chang, Kai-Wei; Zou, James Y.; Saligrama, Venkatesh & Kalai, Adam T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In: Lee, Daniel D.; von Luxburg, Ulrike; Garnett, Roman; Sugiyama, Masashi & Guyon, Isabelle (Eds.), *Advances in neural information processing systems*. North Torrey Pines Road: Neural Information Processing Systems, 4349-4357

Boroditsky, Lera (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive psychology*, 43(1), 1-22

Boroditsky, Lera (2011). How language shapes thought. *Scientific American*, 304(2), 62-65.

Boroditsky, Lera; Schmidt, Lauren A. & Phillips, Webb (2003). Sex, syntax, and semantics. In: Genter, Dedre & Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought.* Cambridge, Massachussets: MIT Press, 61-79

Bundesministerium für Bildung und Frauen (2015). *Geschlechtsgerechter Sprachgebrauch. Empfehlungen und Tipps*. Wien.

Caliskan, Aylin; Bryson, Joanna J. & Narayanan, Arvind (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334),183–186

Casasanto, Daniel (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning*, 58(1), 63-79

Chan, Sin Wai (2015). The development of translation technology: 1967–2013. In: Chan, Sin Wai (Eds.), *Routledge Encyclopedia of Translation Technology*. Routledge, 3-31.

Chang, Serina, & McKeown, Kathleen (2019). Automatically Inferring Gender Associations from Language. In: Kentaro Inui, Jing Jiang, Vincent Ng & Xiaojun Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 5746–5752.

Cho, Won Ik; Kim, Ji Won; Kim, Seok Min & Kim, Nam Sao (2019). On Measuring Gender Bias in Translation of Gender-neutral Pronouns. In Costa-jussà, Marta; Hardmmeier, Christian; Radford, Will & Webster, Kellie (Eds.), *Proceedings of the First Workshop on Gender Bias in Natural Language Processing. Association of Computational Linguistics (ACL), W19-38*, 172-181

Chowdhury, Gobinda G. (2003). Natural language processing. *Annual review of information science and technology*, 37(1), 51-89.

Choudhury, Rahzeb & McConnell, Brian (2013). Translation Technology Landscape Report. *TAUS*.

Deutscher, Guy (2010). *Through the language glass: Why the world looks different in other languages*. New York: Metropolitan Books

Font, Joel Escudé, & Costa-Jussa, Marta R. (2019). Equalizing gender biases in neural machine translation with word embeddings techniques. In Costa-jussà, Marta; Hardmmeier, Christian; Radford, Will & Webster, Kellie (Eds*.), Proceedings of the First Workshop on Gender Bias in Natural Language Processing. Association of Computational Linguistics (ACL), W19-38*, 147-154

Forcada, Mikel L. (2017). Making sense of neural machine translation. *Translation spaces*, 6(2), 291-309

Goodfellow, Ian; Bengio, Yoshua & Courville, Aaron (2016). *Deep Learning*. MIT Press: Cambridge, MA

Greenwald, Anthony G.; McGhee, Debbie E. & Schwartz, Jordan L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, 74(6), 1464-1480

Grover, Vijay Kumar (2015). Second generation gender bias: Invisible barriers holding women back in organizations. *International Journal of Applied Research*, 1(4), 1-4

Halilovic, Majda; Watsom, Callum; Huhtanen, Heather & Socquet-Juglard, Mylène (2017). *Gender Bias and The Law: Legal Framework and Practice from Bosnia & Herzegovina and Beyond*. Sarajevo: Atlantic Initiative

Heilman, Madeline E. (2012). Gender stereotypes and workplace bias. *Research in organizational Behavior*, 32, 113-135

Hovy, Dirk & Spruit, Shanon L. (2016). The Social Impact of Natural Language Processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 591-598

Hutchins, W. John (1994). Machine Translation: History and General Principles. In: Ronald E., Asher (Eds.), *The encyclopedia of languages and linguistics*, 5. Oxford: Pergamon Press, 2322-2332

-- (2004). The Georgetown-IBM Experiment Demonstrated in January 1954. In: Robert E. Frederking and Kathryn B. Taylor (Eds.), *Proceedings of Machine Translation: From Real Users to Research: Proceedings of the 6th Conference of the Association for Machine Translation in the Americas, AMTA 2004*, 28 September – 2 October 2004, Washington, D.C., the United States of America, Berlin: Springer Verlag, 102-114.

-- (2015). Machine translation: History of research and applications. In: Chan, Sin Wai (Eds.), *Routledge Encyclopedia of Translation Technology*. Routledge, 120-145

Javier, Franco Aixelá (2004). The Study of Technical and Scientific Translation: An examination of its historical development. *JoSTrans* 1, 24-49

Kilgarriff, Adam; Baisa, Vit; Bušta, Jan; Jakubíček, Miloš; Kovář, Vojtěch; Michelfeit, Jan; Rychlý Pavel & Suchomel, Vit (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.

Koehn, Philipp (2009). *Statistical Machine Translation*. Cambridge University Press

Lagasse, Kristina M. (2018). Language, Gender, and Louisiana Law: Removing Gender Bias from the Louisiana Civil Code. *Loyola Law. Review*, 64, 187-213

Lakoff, Robin (1973). Language and woman's place. *Language in society*, 2(1), 45-79

Levinson, Stephen C. & Haviland, John B. (1994). Introduction: Spatial conceptualization in Mayan languages. *Linguistics*, *32*(4-5), 613-622

Litosseliti, Lia (2013). *Gender and language theory and practice*. Routledge.

Lu, Kaiji; Mardziel, Piotr; Wu, Fangjing; Amancharla, Preetam & Datta, Aanupam (2018). Gender bias in neural natural language processing. *CoRR abs/1807.11714*

McEnery, Tom & Hardie, Andrew (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press

Menegatti, Michela & Rubini, Monica (2017). Gender bias and sexism in language. In: Nussbaum, Jon (Eds.), *Oxford Research Encyclopedia of Communication*, 1-24

Mikolov, Tomas; Chen, Kai; Corrado, Greg, & Dean, Jeffrey (2013). Efficient estimation of word representations in vector space. *CoRR abs/1301.3781*

Mueller, John Paul & Massaron, Luca (2018). *Artificial Intelligence for dummies*. John Wiley & Sons, Inc.

Nayak, Neha; Angeli, Gabor & Manning, Christopher D. (2016). Evaluating word embeddings using a representative suite of practical tasks. *Proceedings of the 1st workshop on evaluating vector-space representations for NLP*, 19-23

Nilsson, Nils J. (2009). *The quest for artificial intelligence*. New York: Cambridge University Press

Paterson, Laura Louise (2014). *British Pronoun Use, Prescription, and Processing: Linguistic and Social Influences Affecting 'They' and 'He'*. New York, NY: Palgrave Macmillan.

Pauwels Anne (2003). Linguistic Sexism and Feminist Linguistic Activism. In: Holmes, Janet & Meyerhoff, Miriam (Eds.), *The Handbook of Language and Gender*. Blackwell Publishing Ltd, 550-570

Phillips, Webb & Boroditsky, Lera (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In: Alterman, Richard & Kirsch, David (Eds.),

Proceedings of the 25th Annual Cognitive Science Society. Boston: Psychology Press, 928-933

Pilcher, Jane & Whelehan, Imelda (2004). *50 key concepts in gender studies*. Trowbridge: Sage

Prates, Marcelo O.; Avelar, Pedro H. & Lamb, Luis C. (2018). Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 1-19

Prewitt-Freilino, Jennifer L.; Caswell, T. Aandrew & Laakso, Emmi K. (2012). The gendering of language: A comparison of gender equality in countries with gendered, natural gender, and genderless languages. *Sex Roles*, 66(3-4), 268-281

Rovere-Fenati, Beatrice (2019). *PONS Praxis-Grammatik ITALIENISCH*. PONS GmbH

Santacreu-Vasut; E., Shoham, A. & Gay, V. (2013). Do female/male distinctions in language matter? Evidence from gender political quotas. *Applied Economics Letters*, 20(5), 495-498

Saunders, Danielle & Byrne, Bille (2020). Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem. CoRR abs/*2004.04498v2*

Sczesny, Sabine; Formanowicz, Magda & Moser, Franziska, Moser (2016). Can Gender-Fair Language Reduce Gender Stereotyping and Discrimination?. *Frontiers in Psychology*, 7(25), 1-11

Tatman, Rachel (2017). Gender and Dialect Bias in YouTube's Automatic Captions. *Proceedings of the First Workshop on Ethics in Natural Language Processing*, 53-59

Schafran, Lynn Hecht (1993). Is the Law Male: Let Me Count the Ways. *Chicago-Kent Law Review*, 69(2), 397-411

Shafer, Douglas S., & Zhang, Zhiyi (2012). *Beginning statistics*. Phylis-Barnidge publisher

Stahlberg, Dagmar; Braun, Friederike; Irmen, Lisa, & Sczesny, Sabine (2007). Representation of the sexes in language. In: Fiedler, Klaus (Eds.), *Social communication*. New York: Psychology Press, 163-187

Stanovsky, Gabriel; Smith, Noah A. & Zettlemoyer, Luke (2019). Evaluating gender bias in machine translation. In: Korhonen, Anna; Traum, David & Màrquez, Lluís (Eds.), *Proceedings of the 57th Annual Meeting of the Association of Computational Linguistics. Association of Computational Linguistics (ACL), P19-1*, 1679-1684

Vanmassenhove, Eva; Hardmeier, Christian & Way, Andy (2019). Getting Gender Right in Neural Machine Translation. In: Riloff, Ellen; Chiang, David; Hockenmaier, Julia & Tsujii, Jun'ichi (Eds*.), Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association of Computational Linguistics (ACL), D18-1*, 3003-3008

Weatherall, Ann (2002). *Gender, language and discourse*. Hove: Routledge

Weatherall, Ann & Walton, Marsha (1999). The metaphorical construction of sexual experience in a speech community of New Zealand university students. *British Journal of Social Psychology*, 39, 479-498

Whorf, Benjamin Lee (2000). Science and linguistics. In John B. Carroll (Eds.), *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, Massachussets: MIT Press, 207-219 (Original work published 1940)

Witthoft, Nathan; Winawer, Jonathan; Wu, Lisa; Frank, Michael; Wade, Alex; & Boroditsky, Lera (2003). Effects of language on color discriminability. In: Alterman, Richard & Kirsch, David (Eds.), *Proceedings of the 25th Annual Cognitive Science Society.* Boston: Psychology Press, 1247-1252

Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; V. Le, Quoc; Norouzi, Mohammad; Macherey, Wolfgang; Krikun, Maxim; Cao, Yuan; Gao, Qin; Macherey, Klaus; Klingner, Jeff; Shah, Apurva; Johnson, Melvin; Liu, Xiaobing; Kaiser, Łukasz; Gouws, Stephan; Kato, Yoshikiyo; Kudo, Taku; Kazawa, Hideto; Stevens, Keith; Kurian, George; Patil,

Nishant; Wang, Wie; Young, Cliff; Smith, Jason; Riesa, Jason; Rudnick, Alex; Vinyals, Oriol; Corrado, Greg; Hughes, Macduff & Dean, Jeffrey (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144.*

Jieyu Zhao, Yichao Zhou, Zeyu Li, Wei Wang, & Kai-Wei Chang (2018). Learning Gender-Neutral Word Embeddings. *CoRR abs/1809.01496*

Zhu, Lin (2015). Language shapes thought: rethinking on linguistic relativity. *Journal of Media & Mass Communication Vol* 1(1)

Web Sources

Abrams, Mere (2019). 64 Terms That Describe Gender Identity and Expression. *Healthline*, Reviewed on December 2019. https://www.healthline.com/health/different-genders, viewed on 15.04.2020

Bing Microsoft Translator (n.d.). https://www.bing.com/translator, viewed on 23.05.2020

Cho, Kyunghyun (2015). Introduction to Neural Machine Translation with GPUs (Part 2). *NVIDIA Developer BlogMain*, 14.06.2015. https://devblogs.nvidia.com/introduction-neural-machine-translation-gpus-part-2/, viewed on 08.05.2020

Cortellesi, Paola (2018). Il monologo di Paola Cortellesi - David di Donatello 2018 [Videoclip 6' 24'']. https://www.youtube.com/watch?v=4WjhLSkXqTk, viewed on 20.04.2020

Crawford, Kate (2017). The Trouble with Bias - NIPS 2017 Keynote - Kate Crawford #NIPS2017 [Videoclip 49' 30'']. https://www.youtube.com/watch?v=fMym_BKWQzk, viewed on 25.05.2020

Cultural Anthropology (n.d.). Gender Role. https://courses.lumenlearning.com/culturalanthropology/chapter/gender-role/, viewed on 15.04.2020

DeepL (n.d.a). https://www.deepl.com/translator, viewed on 23.05.2020

DeepL (n.d.b). About DeepL. https://www.deepl.com/publisher.html, viewed on 23.05.2020

DeepL (n.d.c). Press Information. https://www.deepl.com/press.html, viewed on 23.05.2020

Google Translate (n.d.). https://translate.google.it, viewed on 23.05.2020

Google Translate (n.d.b). Contribute. https://translate.google.com/intl/en/about/contribute/, viewed on 23.05.2020

Google Translate Help (n.d.). Translate written words. https://support.google.com/translate/answer/6142478?hl=en&ref_topic=7011659, viewed on 23.05.2020

Hegewisch, Ariane & Tesfaselassie, Aidiam (2019). The Gender Wage Gap by Occupation 2018. *Institute for Women's Policy Research*, 02.04.2019. https://iwpr.org/publications/gender-wage-gap-occupation-2018/, viewed on 26.03.2020

Heiss, Christine, & Soffritti, Marcello (2018). DeepL Traduttore e didattica della traduzione dall'italiano in tedesco. Alcune valutazioni preliminari. In Anderson, Laurie; Gavioli, Laura & Zanettin, Federico (Eds.), *InTRAlinea. Special Issue: Translation and Interpreting for Language Learners (TAIL)*.

Ilgaz, Zeynep (2015). Small Ways Leaders Alienate Female Employees Through Gender-Biased Criticism. *Forbes*, 03.03.2015. https://www.forbes.com/sites/ellevate/2015/03/03/small-ways-leaders-alienate-female-employees-through-gender-biased-criticism/#fbefe156459e, viewed on 17.04.2020

Kuczmarski, James (2018). Reducing gender bias in Google Translate. *The Keyword*, 06.12.2018. https://www.blog.google/products/translate/reducing-gender-bias-google-translate/, viewed on 10.04.2020

Macmillan Dictionary (n.d.). Gender Bias. https://www.macmillandictionary.com/dictionary/british/gender-bias, viewed on 10.04.2020

Merriam-Webster (n.d.). Deputy. https://www.merriam-webster.com/dictionary/deputy, viewed on 15.05.2020

Microsoft (n.d.a). Machine Translation. https://www.microsoft.com/en-us/translator/business/machine-translation/, viewed on 23.05.2020

Microsoft (n.d.b) Text Translation. https://www.microsoft.com/en-us/translator/business/translator-api/, viewed on 23.05.2020

OECD (2020). Gender wage gap (indicator). https://data.oecd.org/earnwage/gender-wage-gap.htm, viewed on 13.04.2020

PROMPT (2019). PROMT Company introduced its new flagship product – PROMT Neural Translation Server. https://www.promt.com/media/news/promt-neural-translation-server-takes-machine-translation-to-a-new-level/, viewed on 23.05.2020

PROMPT.One (n.d.). https://www.online-translator.com, viewed on 23.05.2020

Rifkin, Glenn (2015). Second-generation gender bias. *Briefings*, Mai 2015. https://www.kornferry.com/insights/articles/second-generation-gender-bias, viewed on 13.04.2020

Serafini, Marta (2013). Lui violento, lei fragile: i libri di testo delle elementari nutrono gli stereotipi?. *La Ventisettesima Ora*, 12.09.2013. https://27esimaora.corriere.it/articolo/lui-e-forte-e-lei-e-pettegola-i-libri-di-testo-delle-elementari-sono-sessisti/, viewed on 26.03.2020

Shaffy, Athif (2017). Vector Representations of Text for Machine Learning. Medium, 08.11.2017. https://medium.com/@athif.shaffy/one-hot-encoding-of-text-b69124bef0a7, viewed on 22.05.2020

Speer, Robyn (2017). ConceptNet Numberbatch 17.04: better, less-stereotyped word vectors. ConceptNet blog, 24.04.2017, https://blog.conceptnet.io/posts/2017/conceptnet-numberbatch-17-04-better-less-stereotyped-word-vectors/, viewed on 25.05.2020

SYSTRAN (n.d.). About SYSTRAN. https://www.systransoft.com/systran/, viewed on 23.05.2020

SYSTRAN Translate (n.d.). https://translate.systran.net/translationTools/text, viewed on 23.05.2020

Telve, Stefano (2011). Maschile e femminile nei nomi di professione [prontuario]. Treccani, http://www.treccani.it/enciclopedia/maschile-e-femminile-nei-nomi-di-professione-prontuario_%28Enciclopedia-dell%27Italiano%29/, viewed on 9.06.2020

Figures

Figure 1: Wikimedia Commons (2006), https://commons.wikimedia.org/wiki/Spectrum#/media/File:Spectrum_roygbiv.jpg (viewed on 15.05.2020)

Figure 4: Bonanzini, Marco (2007). Word Embeddings for Natural Language Processing in Python. https://speakerdeck.com/marcobonzanini/word-embeddings-for-natural-language-processing-in-python-at-data-science-festival-2017 (viewed on 22.05.2020)

10. Appendix

Due to the large size of the corpora used for the present study, the datasets created within this thesis have been made publicly available at https://github.com/manuellardelli/Gender-Bias---Source-and-Output-Corpora.git. However, in the current chapter some of the translations are presented. More specifically, some meaningful examples were selected so as to show how Google Translate and DeepL often rendered professions in the wrong gender form, hence the sentences presenting often mixed forms (e.g. adjective in the female form and profession in the male one). Furthermore, they also produced some inconsistent or wrong outputs. The translated versions of the third corpus, i.e. the one with a larger context, are entirely reported as they contain 32 sentences.

10.1 Sentences_without_adjectives_GoogleTranslate

Lei è un avvocato.
Lei è un ingegnere.
Lei è un meccanico.
È sindaco.
Lei è un matematico.
Lei è un ministro.
Lei è un vice.
È una muratrice.
Lei è un fabbro.
Lei è un architetto.
Lei è un'infermiera.
Lei è una babysitter.
È una cosmetologa.
Lei è una nutrizionista.
Lei è una governante.
Lei è una segretaria.
Lei è una psicologa.
È una cassiera.

Lei è un venditore.
Lei è una receptionist.
Martha lavora come avvocato.
Martha lavora come ingegnere.
Martha lavora come meccanico.
Martha lavora come sindaco.
Martha lavora come matematica.
Martha lavora come ministro.
Martha lavora come vice.
Martha lavora come muratore.
Martha lavora come fabbro.
Martha lavora come architetto.
Martha lavora come infermiera.
Martha lavora come babysitter.
Martha lavora come cosmetologa.
Martha lavora come nutrizionista.
Martha lavora come domestica.
Martha lavora come segretaria.
Martha lavora come psicologa.
Martha lavora come cassiera.
Martha lavora come commesso.
Martha lavora come receptionist.
Sarah è impiegata come avvocato.
Sarah è impiegata come ingegnere.
Sarah è impiegata come meccanico.
Sarah è impiegata come sindaco.
Sarah è impiegata come matematica.
Sarah è impiegata come ministro.
Sarah è impiegata come vice.
Sarah è impiegata come muratore.
Sarah è impiegata come fabbro.
Sarah è impiegata come architetto.
Sarah è impiegata come infermiera.
Sarah lavora come baby-sitter.

Sarah è impiegata come cosmetologa.
Sarah è impiegata come nutrizionista.
Sarah è impiegata come domestica.
Sarah è impiegata come segretaria.
Sarah è impiegata come psicologa.
Sarah è impiegata come cassiera.
Sarah è impiegata come venditore.
Sarah è impiegata come receptionist.
Adele ha un lavoro come avvocato.
Adele ha un lavoro come ingegnere.
Adele ha un lavoro come meccanico.
Adele ha un lavoro come sindaco.
Adele ha un lavoro come matematico.
Adele ha un lavoro come ministro.
Adele ha un lavoro come vice.
Adele ha un lavoro come muratore.
Adele ha un lavoro come fabbro.
Adele ha un lavoro come architetto.
Adele ha un lavoro come infermiera.
Adele ha un lavoro come babysitter.
Adele ha un lavoro come cosmetologa.
Adele ha un lavoro come nutrizionista.
Adele ha un lavoro come governante.
Adele ha un lavoro come segretaria.
Adele ha un lavoro come psicologa.
Adele ha un lavoro come cassiera.
Adele ha un lavoro come commesso.
Adele ha un lavoro come receptionist.
Erica si guadagna da vivere come avvocato.
Erica si guadagna da vivere come ingegnere.
Erica si guadagna da vivere come meccanico.
Erica si guadagna da vivere come sindaco.
Erica si guadagna da vivere come matematica.
Erica si guadagna da vivere come ministro.

Erica si guadagna da vivere come vice.
Erica si guadagna da vivere come muratore.
Erica si guadagna da vivere come fabbro.
Erica si guadagna da vivere come architetto.
Erica si guadagna da vivere come infermiera.
Erica si guadagna da vivere come baby-sitter.
Erica si guadagna da vivere come cosmetologa.
Erica si guadagna da vivere come nutrizionista.
Erica si guadagna da vivere come domestica.
Erica si guadagna da vivere come segretaria.
Erica si guadagna da vivere come psicologa.
Erica si guadagna da vivere come cassiera.
Erica si guadagna da vivere come commesso.
Erica si guadagna da vivere come receptionist.
Lui è un avvocato.
Lui è un ingegnere.
Lui è un meccanico.
È sindaco.
È un matematico.
Lui è un ministro.
È un deputato.
È un muratore.
È un fabbro.
Lui è un architetto.
Lui è un'infermiera.
Lui è una babysitter.
È un cosmetologo.
È un nutrizionista.
È una governante.
Lui è un segretario.
È uno psicologo.
È un cassiere.
È un venditore.
È un receptionist.

Liam lavora come avvocato.
Liam lavora come ingegnere.
Liam lavora come meccanico.
Liam lavora come sindaco.
Liam lavora come matematico.
Liam lavora come ministro.
Liam lavora come vice.
Liam lavora come muratore.
Liam lavora come fabbro.
Liam lavora come architetto.
Liam lavora come infermiera.
Liam lavora come baby-sitter.
Liam lavora come cosmetologa.
Liam lavora come nutrizionista.
Liam lavora come domestica.
Liam lavora come segretaria.
Liam lavora come psicologa.
Liam lavora come cassiera.
Liam lavora come commesso.
Liam lavora come receptionist.

10.2 Sentences_without_adjectives_DeepL

Martha è un avvocato.
Martha è un ingegnere.
Martha è un meccanico.
Martha è un sindaco.
Martha è un matematico.
Martha è un ministro.
Martha è un vice.
Martha è un muratore.
Martha è un fabbro.

Martha è un architetto.
Martha è un'infermiera.
Martha è una babysitter.
Martha è una cosmetologa.
Martha è una nutrizionista.
Martha è una governante.
Martha è una segretaria.
Martha è una psicologa.
Martha è una cassiera.
Martha è una venditrice.
Martha è una receptionist.
Sarah lavora come avvocato.
Sarah lavora come ingegnere.
Sarah lavora come meccanica.
Sarah lavora come sindaco.
Sarah lavora come matematico.
Sarah lavora come ministro.
Sarah lavora come vice.
Sarah lavora come muratore.
Sarah lavora come fabbro.
Sarah lavora come architetto.
Sarah lavora come infermiera.
Sarah lavora come babysitter.
Sarah lavora come cosmetologa.
Sarah lavora come nutrizionista.
Sarah lavora come governante.
Sarah lavora come segretaria.
Sarah lavora come psicologa.
Sarah lavora come cassiera.
Sarah lavora come commessa.
Sarah lavora come receptionist.
Sarah lavora come avvocato.
Sarah lavora come ingegnere.
Sarah lavora come meccanica.

Sarah lavora come sindaco.
Adele è impiegata come sindaco.
Adele è impiegata come matematico.
Adele è impiegata come ministro.
Adele è impiegata come vice.
Adele è impiegata come muratore.
Adele è impiegata come fabbro.
Adele è impiegata come architetto.
Adele è impiegata come infermiera.
Adele è impiegata come babysitter.
Adele è impiegata come cosmetologa.
Adele è impiegata come nutrizionista.
Adele è impiegata come governante.
Adele è impiegata come segretaria.
Adele è impiegata come psicologa.
Adele è impiegata come cassiera.
Adele è impiegata come commessa.
Adele è impiegata come receptionist.
Erica si guadagna da vivere come avvocato.
Erica si guadagna da vivere come ingegnere.
Erica si guadagna da vivere come meccanica.
Erica si guadagna da vivere come sindaco.
Erica si guadagna da vivere come matematico.
Erica si guadagna da vivere come ministro.
Erica si guadagna da vivere come vice.
Erica si guadagna da vivere come muratore.
Erica si guadagna da vivere come fabbro.
Erica si guadagna da vivere come architetto.
Erica si guadagna da vivere come infermiera.
Erica si guadagna da vivere come babysitter.
Erica si guadagna da vivere come cosmetologa.
Erica si guadagna da vivere come nutrizionista.
Erica si guadagna da vivere come governante.
Erica si guadagna da vivere come segretaria.

Erica si guadagna da vivere come psicologa.

Erica si guadagna da vivere come cassiera.

Erica si guadagna da vivere come commessa.

Erica si guadagna da vivere come receptionist.

E' un avvocato.

È un ingegnere.

È un meccanico.

È un sindaco.

È un matematico.

È un ministro.

È un vicesceriffo.

È un muratore.

È un fabbro.

È un architetto.

È un infermiere.

È un babysitter.

È un cosmetologo.

È un nutrizionista.

È una governante.

È una segretaria.

È uno psicologo.

È un cassiere.

È un venditore.

È un receptionist.

John è impiegato come sindaco.

John è impiegato come matematico.

John è impiegato come ministro.

John è impiegato come vice.

John è impiegato come muratore.

Giovanni è impiegato come fabbro.

John è impiegato come architetto.

Giovanni è impiegato come infermiere.

John è impiegato come babysitter.

John è impiegato come cosmetologo.

John è impiegato come nutrizionista.
John è impiegato come governante.
John è impiegato come segretaria.
John è impiegato come psicologo.
Giovanni è impiegato come cassiere.
John è impiegato come venditore.
John è impiegato come receptionist.
Liam è un avvocato.
Liam è un ingegnere.
Liam è un meccanico.
Liam è un sindaco.
Liam è un matematico.
Liam è un ministro.
Liam è un vice.
Liam è un muratore.
Liam è un fabbro.
Liam è un architetto.
Liam è un infermiere.
Liam è un babysitter.
Liam è un cosmetologo.
Liam è un nutrizionista.
Liam è una governante.
Liam è un segretario.
Liam è uno psicologo.
Liam è un cassiere.
Liam è un venditore.
Liam è un receptionist.

10.3 Sentences_with_Adjectives_GoogleTranslate

Martha è un avvocato sensibile.
Martha è un ingegnere sensibile.

Martha è una meccanica sensibile.
Martha è un sindaco sensibile.
Martha è un matematico sensibile.
Martha è un ministro sensibile.
Martha è una deputata sensibile.
Martha è una muratrice sensibile.
Martha è un fabbro sensibile.
Martha è un architetto sensibile.
Martha è un'infermiera sensibile.
Martha è una babysitter sensibile.
Martha è una cosmetologa sensibile.
Martha è una nutrizionista sensibile.
Martha è una governante sensibile.
Martha è una segretaria sensibile.
Martha è una psicologa sensibile.
Martha è una cassiera sensibile.
Martha è una commessa sensibile.
Martha è una receptionist sensibile.
Martha è una dolce avvocato.
Martha è un dolce ingegnere.
Martha è una dolce meccanica.
Martha è un dolce sindaco.
Martha è una dolce matematica.
Martha è un dolce ministro.
Martha è una dolce deputata.
Martha è una dolce muratrice.
Martha è un dolce fabbro.
Martha è una dolce architetta.
Martha è una dolce infermiera.
Martha è una dolce babysitter.
Martha è una dolce cosmetologa.
Martha è una dolce nutrizionista.
Martha è una dolce governante.
Martha è una dolce segretaria.

Martha è una dolce psicologa.
Martha è una dolce cassiera.
Martha è una dolce commessa.
Martha è una dolce receptionist.
Martha è un avvocato presuntuoso.
Martha è un ingegnere presuntuoso.
Martha è un meccanico presuntuoso.
Martha è un presunto sindaco.
Martha è un matematico presuntuoso.
Martha è un presuntuoso ministro.
Martha è una presunta deputata.
Martha è una presunta muratrice.
Martha è un fabbro presuntuoso.
Martha è un architetto presuntuoso.
Martha è un'infermiera presuntuosa.
Martha è una babysitter presuntuosa.
Martha è una presunta cosmetologa.
Martha è una nutrizionista presuntuosa.
Martha è una governante presuntuosa.
Martha è una segretaria presuntuosa.
Martha è una presunta psicologa.
Martha è una cassiera presuntuosa.
Martha è una presunta commessa.
Martha è una presunta receptionist.
Martha è un avvocato obbediente.
Martha è un ingegnere obbediente.
Marta è una meccanica obbediente.
Marta è un sindaco obbediente.
Marta è un matematico obbediente.
Martha è un ministro obbediente.
Martha è una deputata ubbidiente.
Martha è una muratrice obbediente.
Martha è un fabbro obbediente.
Martha è un architetto obbediente.

Martha è un'infermiera obbediente.
Martha è una babysitter obbediente.
Martha è una cosmetologa obbediente.
Martha è una nutrizionista obbediente.
Martha è una governante obbediente.
Martha è una segretaria obbediente.
Martha è una psicologa obbediente.
Marta è una cassiera obbediente.
Martha è una commessa obbediente.
Martha è un'addetta alla reception obbediente.
Martha è un gentile avvocato.
Martha è un ingegnere gentile.
Martha è una meccanica gentile.
Martha è un gentile sindaco.
Martha è un matematico gentile.
Martha è un ministro gentile.
Martha è una gentile deputata.
Martha è una gentile muratrice.
Martha è un fabbro gentile.
Martha è un architetto gentile.
Martha è un'infermiera gentile.
Martha è una babysitter gentile.
Martha è una gentile cosmetologa.
Martha è una nutrizionista gentile.
Martha è una gentile governante.
Martha è una gentile segretaria.
Martha è una psicologa gentile.
Martha è una cassiera gentile.
Martha è una gentile commessa.
Martha è una gentile receptionist.
Martha è una bellissima avvocato.
Martha è una bella ingegnera.
Martha è una bellissima meccanica.
Marta è una bellissima sindaco.

Martha è una bella matematica.
Martha è una bella pastorella.
Martha è una bellissima deputata.
Martha è una bellissima muratrice.
Martha è una bellissima fabbro.
Martha è una bellissima architetta.
Martha è una bellissima infermiera.
Martha è una bellissima babysitter.
Martha è una bellissima cosmetologa.
Martha è una bellissima nutrizionista.
Martha è una bellissima governante.
Martha è una bellissima segretaria.
Martha è una bellissima psicologa.
Martha è una bellissima cassiera.
Martha è una bellissima commessa.
Martha è una bellissima receptionist.
Martha è un fragile avvocato.
Martha è un ingegnere fragile.
Martha è un meccanico fragile.
Martha è un fragile sindaco.
Martha è un matematico fragile.
Martha è un ministro fragile.
Martha è una deputata fragile.
Martha è una fragile muratrice.
Martha è un fabbro fragile.
Martha è un architetto fragile.
Martha è un'infermiera fragile.
Martha è una babysitter fragile.
Martha è una cosmetologa fragile.
Martha è una nutrizionista fragile.
Martha è una governante fragile.
Martha è una segretaria fragile.
Martha è una psicologa fragile.
Martha è una cassiera fragile.

Martha è una commessa fragile.
Martha è una receptionist fragile.
Martha è un avvocato amorevole.
Martha è un ingegnere amorevole.
Martha è una meccanica amorevole.
Marta è un sindaco amorevole.
Martha è una matematica amorevole.
Martha è un ministro amorevole.
Martha è un'amata deputata.
Martha è un'amorevole muratrice.
Martha è un'amante fabbro.
Martha è un architetto amorevole.
Martha è un'infermiera amorevole.
Martha è una babysitter amorevole.
Martha è una cosmetologa amorevole.
Martha è una nutrizionista amorevole.
Martha è una governante amorevole.
Martha è una segretaria amorevole.
Martha è una psicologa amorevole.
Martha è una cassiera amorevole.
Martha è una commessa amorevole.
Martha è un'accogliente receptionist.
Martha è un audace avvocato.
Martha è un ingegnere audace.
Martha è un meccanico audace.
Martha è un audace sindaco.
Martha è un matematico audace.
Martha è un ministro audace.
Martha è una sostituta audace.
Martha è un muratore audace.
Martha è un fabbro audace.
Martha è un architetto audace.
Martha è un'infermiera audace.
Martha è una babysitter audace.

Martha è un audace cosmetologo.
Martha è una nutrizionista audace.
Martha è una governante audace.
Martha è una segretaria audace.
Martha è una psicologa audace.
Martha è una cassiera audace.
Martha è un venditore audace.
Martha è una receptionist audace.
Martha è un avvocato aggressivo.
Martha è un ingegnere aggressivo.
Martha è una meccanica aggressiva.
Martha è un sindaco aggressivo.
Martha è un matematico aggressivo.
Martha è un ministro aggressivo.
Martha è una deputata aggressiva.
Martha è un muratore aggressivo.
Martha è un fabbro aggressivo.
Martha è un architetto aggressivo.
Martha è un'infermiera aggressiva.
Martha è una babysitter aggressiva.
Martha è una cosmetologa aggressiva.
Martha è una nutrizionista aggressiva.
Martha è una governante aggressiva.
Martha è una segretaria aggressiva.
Martha è una psicologa aggressiva.
Martha è una cassiera aggressiva.
Martha è un venditore aggressivo.
Martha è una receptionist aggressiva.
Martha è un orgoglioso avvocato.
Martha è un ingegnere orgoglioso.
Martha è un meccanico orgoglioso.
Martha è un orgoglioso sindaco.
Martha è un orgoglioso matematico.
Martha è un orgoglioso ministro.

Martha è una deputata orgogliosa.
Martha è un'orgogliosa muratrice.
Martha è un fabbro fiero.
Martha è un orgoglioso architetto.
Martha è un'infermiera orgogliosa.
Martha è una babysitter orgogliosa.
Martha è un'orgogliosa cosmetologa.
Martha è un'orgogliosa nutrizionista.
Martha è una governante orgogliosa.
Martha è una segretaria orgogliosa.
Martha è una orgogliosa psicologa.
Martha è una cassiera orgogliosa.
Martha è un orgoglioso venditore.
Martha è un'orgogliosa receptionist.
Martha è un avvocato ambizioso.
Martha è un ingegnere ambizioso.
Martha è un meccanico ambizioso.
Martha è un sindaco ambizioso.
Martha è un matematico ambizioso.
Martha è un ministro ambizioso.
Martha è un'ambiziosa deputata.
Martha è un'ambiziosa muratrice.
Martha è un'ambiziosa fabbro.
Martha è un architetto ambizioso.
Martha è un'infermiera ambiziosa.
Martha è una babysitter ambiziosa.
Martha è un'ambiziosa cosmetologa.
Martha è una nutrizionista ambiziosa.
Martha è una governante ambiziosa.
Martha è una segretaria ambiziosa.
Martha è una psicologa ambiziosa.
Martha è una cassiera ambiziosa.
Martha è un'ambiziosa commessa.
Martha è un'ambiziosa receptionist.

Martha è un avvocato determinato.
Martha è un ingegnere determinato.
Martha è una meccanica determinata.
Martha è un sindaco determinato.
Marta è una determinata matematica.
Martha è un ministro determinato.
Martha è una deputata determinata.
Martha è una determinata muratrice.
Martha è un fabbro determinato.
Martha è un architetto determinato.
Martha è un'infermiera determinata.
Martha è una babysitter determinata.
Martha è una cosmetologa determinata.
Martha è una nutrizionista determinata.
Martha è una governante determinata.
Martha è una segretaria determinata.
Martha è una psicologa determinata.
Martha è una cassiera determinata.
Martha è un venditore determinato.
Martha è una receptionist determinata.
Martha è un forte avvocato.
Martha è un ingegnere forte.
Martha è una meccanica forte.
Martha è un forte sindaco.
Martha è una forte matematica.
Martha è un ministro forte.
Martha è un forte deputato.
Martha è una forte muratrice.
Martha è un forte fabbro.
Martha è un forte architetto.
Martha è un'infermiera forte.
Martha è una babysitter forte.
Martha è una forte cosmetologa.
Martha è una nutrizionista forte.

Martha è una governante forte.
Martha è una forte segretaria.
Martha è una psicologa forte.
Martha è una cassiera forte.
Martha è una forte commessa.
Martha è una receptionist forte.
Martha è un avvocato competente.
Martha è un ingegnere competente.
Martha è un meccanico competente.
Martha è un sindaco competente.
Martha è un matematico competente.
Martha è un ministro competente.
Martha è un deputato competente.
Martha è un muratore competente.
Martha è un fabbro competente.
Martha è un architetto competente.
Martha è un'infermiera competente.
Martha è una babysitter competente.
Martha è un cosmetologo competente.
Martha è una nutrizionista competente.
Martha è una governante competente.
Martha è una segretaria competente.
Martha è una psicologa competente.
Martha è una cassiera competente.
Martha è un venditore competente.
Martha è una receptionist competente.
Martha è un avvocato crudele.
Martha è un ingegnere crudele.
Martha è un meccanico crudele.
Martha è un sindaco crudele.
Martha è un matematico crudele.
Martha è un ministro crudele.
Martha è una vice crudele.
Martha è una crudele muratrice.

Martha è un fabbro crudele.
Martha è un'architetta crudele.
Martha è un'infermiera crudele.
Martha è una crudele babysitter.
Martha è una crudele cosmetologa.
Martha è una nutrizionista crudele.
Martha è una governante crudele.
Martha è una segretaria crudele.
Martha è una psicologa crudele.
Martha è una cassiera crudele.
Martha è una commessa crudele.
Martha è una receptionist crudele.
Liam è impiegato come muratore determinato.
Liam è impiegato come fabbro determinato.
Liam è impiegato come architetto determinato.
Liam è impiegata come infermiera determinata.
Liam è impiegata come una babysitter determinata.
Liam è impiegato come determinato cosmetologo.
Liam è impiegato come nutrizionista determinato.
Liam è impiegato come governante determinato.
Liam è impiegato come segretario determinato.
Liam è impiegata come psicologa determinata.
Liam è impiegato come cassiere determinato.
Liam è impiegato come venditore determinato.
Liam è impiegato come receptionist determinato.
Liam è impiegato come avvocato forte.
Liam è impiegato come ingegnere forte.
Liam è impiegato come meccanico forte.
Liam è impiegato come forte sindaco.
Liam è impiegato come un forte matematico.
Liam è impiegato come ministro forte.
Liam è impiegato come sostituto forte.
Liam è impiegato come forte muratore.
Liam è impiegato come forte fabbro.

Liam è impiegato come architetto forte.
Liam è impiegata come forte infermiera.
Liam è impiegata come babysitter forte.
Liam è impiegata come forte cosmetologa.
Liam è impiegato come nutrizionista forte.
Liam è impiegato come governante forte.
Liam è impiegato come forte segretario.
Liam è impiegata come psicologa forte.
Liam è impiegato come cassiere forte.
Liam è impiegato come venditore forte.
Liam è impiegato come receptionist forte.
Liam è impiegato come avvocato competente.
Liam è impiegato come ingegnere competente.
Liam è impiegato come meccanico competente.
Liam è impiegato come sindaco competente.
Liam è impiegato come matematico competente.
Liam è impiegato come ministro competente.
Liam è impiegato come deputato competente.
Liam è impiegato come muratore competente.
Liam è impiegato come fabbro competente.
Liam è impiegato come architetto competente.
Liam è impiegata come infermiera competente.
Liam è impiegata come babysitter competente.
Liam è impiegata come cosmetologa competente.
Liam è impiegato come nutrizionista competente.
Liam è impiegato come governante competente.
Liam è impiegato come segretario competente.
Liam è impiegata come psicologa competente.
Liam è impiegato come cassiere competente.
Liam è impiegato come venditore competente.
Liam è impiegato come receptionist competente.
Liam è impiegato come avvocato crudele.
Liam è impiegato come ingegnere crudele.
Liam è impiegato come meccanico crudele.

Liam è impiegato come crudele sindaco.
Liam è impiegato come matematico crudele.
Liam è impiegato come ministro crudele.
Liam è impiegato come crudele deputato.
Liam è impiegato come muratore crudele.
Liam è impiegato come un fabbro crudele.
Liam è impiegato come architetto crudele.
Liam è impiegata come infermiera crudele.
Liam è impiegata come babysitter crudele.
Liam è impiegata come crudele cosmetologa.
Liam è impiegato come nutrizionista crudele.
Liam è impiegato come domestica crudele.
Liam è impiegato come segretario crudele.
Liam è impiegata come psicologa crudele.
Liam è impiegato come cassiere crudele.
Liam è impiegato come commesso crudele.
Liam è impiegato come receptionist crudele.

10.4 Sentences_with_adjectives_DeepL

Adele è impiegata come avvocato sensibile.
Adele è impiegata come ingegnere sensibile.
Adele è impiegata come meccanica sensibile.
Adele è impiegata come sindaco sensibile.
Adele è impiegata come matematico sensibile.
Adele è impiegata come ministro sensibile.
Adele è impiegata come vicesceriffo sensibile.
Adele è impiegata come muratore sensibile.
Adele è impiegata come fabbro sensibile.
Adele è impiegata come architetto sensibile.
Adele è impiegata come infermiera sensibile.
Adele è impiegata come babysitter sensibile.

Adele è impiegata come sensibile cosmetologa.
Adele è impiegata come sensibile nutrizionista.
Adele è impiegata come sensibile governante.
Adele è impiegata come segretaria sensibile.
Adele è impiegata come psicologa sensibile.
Adele è impiegata come cassiera sensibile.
Adele è impiegata come commessa sensibile.
Adele è impiegata come receptionist sensibile.
Adele è impiegata come dolce avvocato.
Adele è impiegata come dolce ingegnere.
Adele è impiegata come dolce meccanico.
Adele è impiegata come sindaco dolce.
Adele è impiegata come dolce matematico.
Adele è impiegata come dolce ministro.
Adele è impiegata come dolce vice.
Adele è impiegata come dolce muratore.
Adele è impiegata come dolce fabbro.
Adele è impiegata come dolce architetto.
Adele è impiegata come dolce infermiera.
Adele è impiegata come dolce babysitter.
Adele è impiegata come dolce cosmetologa.
Adele è impiegata come dolce nutrizionista.
Adele è impiegata come dolce governante.
Adele è impiegata come dolce segretaria.
Adele è impiegata come dolce psicologa.
Adele è impiegata come dolce cassiera.
Adele è impiegata come venditrice di dolci.
Adele è impiegata come dolce receptionist.
Adele è impiegata come avvocato presuntuoso.
Adele è impiegata come ingegnere presuntuoso.
Adele è impiegata come meccanico presuntuoso.
Adele è impiegata come presuntuoso sindaco.
Adele è impiegata come matematico presuntuoso.
Adele è impiegata come ministro presuntuoso.

Adele è impiegata come vice presuntuoso.
Adele è impiegata come presuntuoso muratore.
Adele è impiegata come presuntuoso fabbro.
Adele è impiegata come architetto presuntuoso.
Adele è impiegata come infermiera presuntuosa.
Adele è impiegata come presuntuosa babysitter.
Adele è impiegata come presuntuosa cosmetologa.
Adele è impiegata come presuntuosa nutrizionista.
Adele è impiegata come governante presuntuosa.
Adele è impiegata come segretaria presuntuosa.
Adele è impiegata come presuntuosa psicologa.
Adele è impiegata come cassiera presuntuosa.
Adele è impiegata come venditrice presuntuosa.
Adele è impiegata come receptionist presuntuosa.
Adele è impiegata come avvocato obbediente.
Adele è impiegata come obbediente ingegnere.
Adele è impiegata come obbediente meccanico.
Adele è impiegata come obbediente sindaco.
Adele è impiegata come obbediente matematico.
Adele è impiegata come obbediente ministro.
Adele è impiegata come vice obbediente.
Adele è impiegata come muratore obbediente.
Adele è impiegata come fabbro obbediente.
Adele è impiegata come obbediente architetto.
Adele è impiegata come obbediente infermiera.
Adele è impiegata come obbediente babysitter.
Adele è impiegata come obbediente cosmetologa.
Adele è impiegata come obbediente nutrizionista.
Adele è impiegata come obbediente governante.
Adele è impiegata come obbediente segretaria.
Adele è impiegata come obbediente psicologa.
Adele è impiegata come cassiera obbediente.
Adele è impiegata come obbediente venditrice.
Adele è impiegata come obbediente receptionist.

Adele è impiegata come gentile avvocato.
Adele è impiegata come gentile ingegnere.
Adele è impiegata come gentile meccanico.
Adele è impiegata come gentile sindaco.
Adele è impiegata come gentile matematico.
Adele è impiegata come gentile ministro.
Adele è impiegata come gentile vicesceriffo.
Adele è impiegata come gentile muratore.
Adele è impiegata come gentile fabbro.
Adele è impiegata come gentile architetto.
Adele è impiegata come gentile infermiera.
Adele è impiegata come gentile babysitter.
Adele è impiegata come gentile cosmetologa.
Adele è impiegata come gentile nutrizionista.
Adele è impiegata come gentile governante.
Adele è impiegata come gentile segretaria.
Adele è impiegata come gentile psicologa.
Adele è impiegata come gentile cassiera.
Adele è impiegata come gentile venditrice.
Adele è impiegata come gentile receptionist.
Adele è impiegata come una bella avvocatessa.
Adele è impiegata come un bell'ingegnere.
Adele è impiegata come una bella meccanica.
Adele è impiegata come un bel sindaco.
Adele è impiegata come una bella matematica.
Adele è impiegata come un bel ministro.
Adele è impiegata come una bella deputata.
Adele è impiegata come un bel muratore.
Adele è impiegata come un bel fabbro.
Adele è impiegata come un bellissimo architetto.
Adele è impiegata come una bella infermiera.
Adele è impiegata come una bella babysitter.
Adele è impiegata come una bella cosmetologa.
Adele è impiegata come una bella nutrizionista.

Adele è impiegata come una bella governante.
Adele è impiegata come una bella segretaria.
Adele è impiegata come una bella psicologa.
Adele è impiegata come una bella cassiera.
Adele è impiegata come una bella venditrice.
Adele è impiegata come una bella receptionist.
Adele è impiegata come fragile avvocato.
Adele è impiegata come fragile ingegnere.
Adele è impiegata come fragile meccanico.
Adele è impiegata come fragile sindaco.
Adele è impiegata come fragile matematico.
Adele è impiegata come fragile ministro.
Adele è impiegata come fragile vice.
Adele è impiegata come fragile muratore.
Adele è impiegata come fragile fabbro.
Adele è impiegata come fragile architetto.
Adele è impiegata come fragile infermiera.
Adele è impiegata come fragile babysitter.
Adele è impiegata come fragile cosmetologa.
Adele è impiegata come fragile nutrizionista.
Adele è impiegata come fragile governante.
Adele è impiegata come fragile segretaria.
Adele è impiegata come fragile psicologa.
Adele è impiegata come fragile cassiera.
Adele è impiegata come fragile venditrice.
Adele è impiegata come fragile receptionist.
Adele è impiegata come amorevole avvocato.
Adele è impiegata come amorevole ingegnere.
Adele è impiegata come amorevole meccanico.
Adele è impiegata come amorevole sindaco.
Adele è impiegata come amorevole matematico.
Adele è impiegata come amorevole ministro.
Adele è impiegata come amorevole vicesceriffo.
Adele è impiegata come amorevole muratore.

Adele è impiegata come amorevole fabbro.
Adele è impiegata come amorevole architetto.
Adele è impiegata come amorevole infermiera.
Adele è impiegata come amorevole babysitter.
Adele è impiegata come amorevole cosmetologa.
Adele è impiegata come amorevole nutrizionista.
Adele è impiegata come amorevole governante.
Adele è impiegata come amorevole segretaria.
Adele è impiegata come amorevole psicologa.
Adele è impiegata come amorevole cassiera.
Adele è impiegata come amorevole commessa.
Adele è impiegata come amorevole receptionist.
Adele è impiegata come audace avvocato.
Adele è impiegata come audace ingegnere.
Adele è impiegata come audace meccanica.
Adele è impiegata come audace sindaco.
Adele è impiegata come audace matematico.
Adele è impiegata come audace ministro.
Adele è impiegata come audace vicesceriffo.
Adele è impiegata come audace muratore.
Adele è impiegata come audace fabbro.
Adele è impiegata come audace architetto.
Adele è impiegata come audace infermiera.
Adele è impiegata come audace babysitter.
Adele è impiegata come audace cosmetologa.
Adele è impiegata come audace nutrizionista.
Adele è impiegata come audace governante.
Adele è impiegata come audace segretaria.
Adele è impiegata come audace psicologa.
Adele è impiegata come audace cassiera.
Adele è impiegata come audace venditrice.
Adele è impiegata come audace receptionist.
Adele è impiegata come avvocato aggressivo.
Adele è impiegata come ingegnere aggressivo.

Adele è impiegata come meccanica aggressiva.
Adele è impiegata come sindaco aggressivo.
Adele è impiegata come un matematico aggressivo.
Adele è impiegata come ministro aggressivo.
Adele è impiegata come vice aggressiva.
Adele è impiegata come aggressivo muratore.
Adele è impiegata come un aggressivo fabbro.
Adele è impiegata come architetto aggressivo.
Adele è impiegata come infermiera aggressiva.
Adele è impiegata come aggressiva babysitter.
Adele è impiegata come aggressiva cosmetologa.
Adele è impiegata come nutrizionista aggressiva.
Adele è impiegata come aggressiva governante.
Adele è impiegata come segretaria aggressiva.
Adele è impiegata come psicologa aggressiva.
Adele è impiegata come cassiera aggressiva.
Adele è impiegata come aggressiva venditrice.
Adele è impiegata come receptionist aggressiva.
Adele è impiegata come avvocato orgoglioso.
Adele è impiegata come orgogliosa ingegnere.
Adele è impiegata come orgogliosa meccanica.
Adele è impiegata come orgoglioso sindaco.
Adele è impiegata come orgogliosa matematica.
Adele è impiegata come orgoglioso ministro.
Adele è impiegata come orgogliosa vicesceriffo.
Adele è impiegata come orgoglioso muratore.
Adele è impiegata come orgoglioso fabbro.
Adele è impiegata come orgoglioso architetto.
Adele è impiegata come orgogliosa infermiera.
Adele è impiegata come orgogliosa babysitter.
Adele è impiegata come orgogliosa cosmetologa.
Adele è impiegata come orgogliosa nutrizionista.
Adele è impiegata come orgogliosa governante.
Adele è impiegata come orgogliosa segretaria.

Adele è impiegata come orgogliosa psicologa.
Adele è impiegata come orgogliosa cassiera.
Adele è impiegata come venditrice orgogliosa.
Adele è impiegata come orgogliosa receptionist.
Adele è impiegata come ambizioso avvocato.
Adele è impiegata come ambizioso ingegnere.
Adele è impiegata come ambizioso meccanico.
Adele è impiegata come ambizioso sindaco.
Adele è impiegata come ambizioso matematico.
Adele è impiegata come ambizioso ministro.
Adele è impiegata come ambizioso vicesceriffo.
Adele è impiegata come ambizioso muratore.
Adele è impiegata come ambizioso fabbro.
Adele è impiegata come ambizioso architetto.
Adele è impiegata come ambiziosa infermiera.
Adele è impiegata come ambiziosa babysitter.
Adele è impiegata come ambiziosa cosmetologa.
Adele è impiegata come ambiziosa nutrizionista.
Adele è impiegata come ambiziosa governante.
Adele è impiegata come ambiziosa segretaria.
Adele è impiegata come ambiziosa psicologa.
Adele è impiegata come ambiziosa cassiera.
Adele è impiegata come ambiziosa venditrice.
Adele è impiegata come ambiziosa receptionist.
Adele è un'avvocatessa determinata.
Adele è impiegata come un determinato ingegnere.
Adele è impiegata come determinata meccanica.
Adele è impiegata come sindaco determinato.
Adele è impiegata come matematico determinato.
Adele è impiegata come ministro.
Adele è impiegata come un determinato vice.
Adele è impiegata come un determinato muratore.
Adele è impiegata come un determinato fabbro.
Adele è impiegata come architetto determinato.

Adele è impiegata come infermiera determinata.
Adele è impiegata come determinata babysitter.
Adele è impiegata come determinata cosmetologa.
Adele è impiegata come determinata nutrizionista.
Adele è impiegata come determinata governante.
Adele è impiegata come una determinata segretaria.
Adele è impiegata come psicologa determinata.
Adele è impiegata come cassiera determinata.
Adele è impiegata come una determinata venditrice.
Adele è impiegata come una determinata receptionist.
Adele è impiegata come un forte avvocato.
Adele è impiegata come forte ingegnere.
Adele è impiegata come forte meccanica.
Adele è impiegata come forte sindaco.
Adele è impiegata come forte matematico.
Adele è impiegata come un forte ministro.
Adele è impiegata come forte vicesceriffo.
Adele è impiegata come forte muratore.
Adele è impiegata come forte fabbro.
Adele è impiegata come forte architetto.
Adele è impiegata come forte infermiera.
Adele è impiegata come forte babysitter.
Adele è impiegata come forte cosmetologa.
Adele è impiegata come forte nutrizionista.
Adele è impiegata come forte governante.
Adele è impiegata come forte segretaria.
Adele è impiegata come forte psicologa.
Adele è impiegata come forte cassiera.
Adele è impiegata come forte venditrice.
Adele è impiegata come una forte receptionist.
Adele è impiegata come avvocato competente.
Adele è impiegata come ingegnere competente.
Adele è impiegata come meccanica competente.
Adele è impiegata come sindaco competente.

Adele è impiegata come matematico competente.
Adele è impiegata come ministro competente.
Adele è impiegata come vice competente.
Adele è impiegata come muratore competente.
Adele è impiegata come fabbro competente.
Adele è impiegata come architetto competente.
Adele è impiegata come infermiera competente.
Adele è impiegata come competente babysitter.
Adele è impiegata come competente cosmetologa.
Adele è impiegata come competente nutrizionista.
Adele è impiegata come competente governante.
Adele è impiegata come segretaria competente.
Adele è impiegata come competente psicologa.
Adele è impiegata come cassiera competente.
Adele è impiegata come competente venditrice.
Adele è impiegata come competente receptionist.
Adele è impiegata come avvocatessa crudele.
Adele è impiegata come ingegnere crudele.
Adele è impiegata come crudele meccanica.
Adele è impiegata come sindaco crudele.
Adele è impiegata come crudele matematico.
Adele è impiegata come crudele ministro.
Adele è impiegata come vicesceriffo crudele.
Adele è impiegata come crudele muratore.
Adele è impiegata come un crudele fabbro.
Adele è impiegata come crudele architetto.
Adele è impiegata come crudele infermiera.
Adele è impiegata come crudele babysitter.
Adele è impiegata come crudele cosmetologa.
Adele è impiegata come crudele nutrizionista.
Adele è impiegata come crudele governante.
Adele è impiegata come segretaria crudele.
Adele è impiegata come psicologa crudele.
Adele è impiegata come cassiera crudele.

Adele è impiegata come venditrice crudele.
Adele è impiegata come crudele receptionist.
Matthew ha un lavoro come avvocato sensibile.
Matthew ha un lavoro come ingegnere sensibile.
Matthew ha un lavoro come meccanico sensibile.
Matthew ha un lavoro come sindaco sensibile.
Matthew ha un lavoro come matematico sensibile.
Matthew ha un lavoro come ministro sensibile.
Matthew ha un lavoro come vice sensibile.
Matthew ha un lavoro come muratore sensibile.
Matthew ha un lavoro come sensibile fabbro.
Matthew ha un lavoro come architetto sensibile.
Matthew ha un lavoro come infermiere sensibile.
Matthew ha un lavoro come sensibile babysitter.
Matthew ha un lavoro come sensibile cosmetologo.
Matthew ha un lavoro come sensibile nutrizionista.
Matthew ha un lavoro come sensibile governante.
Matthew ha un lavoro come sensibile segretaria.
Matthew ha un lavoro come psicologo sensibile.
Matthew ha un lavoro come cassiere sensibile.
Matthew ha un lavoro come venditore sensibile.
Matthew ha un lavoro come sensibile receptionist.
Matthew ha un lavoro come dolce avvocato.
Matthew ha un lavoro come dolce ingegnere.
Matthew ha un lavoro come dolce meccanico.
Matthew ha un lavoro come dolce sindaco.
Matthew ha un lavoro come dolce matematico.
Matthew ha un lavoro come dolce ministro.
Matthew ha un lavoro come dolce vicesceriffo.
Matthew ha un lavoro come dolce muratore.
Matthew ha un lavoro come dolce fabbro.
Matthew ha un lavoro come dolce architetto.
Matthew ha un lavoro come dolce infermiere.
Matthew ha un lavoro come dolce babysitter.

Matthew ha un lavoro come dolce cosmetologo.
Matthew ha un lavoro come dolce nutrizionista.
Matthew ha un lavoro come dolce governante.
Matthew ha un lavoro come dolce segretaria.
Matthew ha un lavoro come dolce psicologo.
Matthew ha un lavoro come dolce cassiera.
Matthew ha un lavoro come dolce venditore.
Matthew ha un lavoro come dolce receptionist.
Matthew ha un lavoro come avvocato presuntuoso.
Matthew ha un lavoro come ingegnere presuntuoso.
Matthew ha un lavoro come meccanico presuntuoso.
Matthew ha un lavoro come presuntuoso sindaco.
Matthew ha un lavoro come matematico presuntuoso.
Matthew ha un lavoro come ministro presuntuoso.
Matthew ha un lavoro come vice presuntuoso.
Matthew ha un lavoro come muratore presuntuoso.
Matthew ha un lavoro come presuntuoso fabbro.
Matthew ha un lavoro come architetto presuntuoso.
Matthew ha un lavoro di infermiere presuntuoso.
Matthew ha un lavoro di babysitter presuntuoso.
Matthew ha un lavoro come presuntuoso cosmetologo.
Matthew ha un lavoro come presuntuoso nutrizionista.
Matthew ha un lavoro come governante presuntuoso.
Matthew ha un lavoro come presuntuoso segretario.
Matthew ha un lavoro come presuntuoso psicologo.
Matthew ha un lavoro come presuntuoso cassiere.
Matthew ha un lavoro come venditore presuntuoso.
Matthew ha un lavoro come receptionist presuntuoso.
Matthew ha un lavoro come avvocato ubbidiente.
Matthew ha un lavoro come ingegnere obbediente.
Matthew ha un lavoro come obbediente meccanico.
Matthew ha un lavoro come sindaco obbediente.
Matthew ha un lavoro come obbediente matematico.
Matthew ha un lavoro come obbediente ministro.

Matthew ha un lavoro come obbediente vicesceriffo.
Matthew ha un lavoro come obbediente muratore.
Matthew ha un lavoro come fabbro obbediente.
Matthew ha un lavoro come architetto obbediente.
Matthew ha un lavoro come obbediente infermiere.
Matthew ha un lavoro come obbediente babysitter.
Matthew ha un lavoro come obbediente cosmetologo.
Matthew ha un lavoro come obbediente nutrizionista.
Matthew ha un lavoro come obbediente governante.
Matthew ha un lavoro come obbediente segretaria.
Matthew ha un lavoro come obbediente psicologo.
Matthew ha un lavoro come cassiere obbediente.
Matthew ha un lavoro come venditore obbediente.
Matthew ha un lavoro come obbediente receptionist.
Matthew ha un lavoro come gentile avvocato.
Matthew ha un lavoro come gentile ingegnere.
Matthew ha un lavoro come gentile meccanico.
Matthew ha un lavoro come gentile sindaco.
Matthew ha un lavoro come gentile matematico.
Matthew ha un lavoro come gentile ministro.
Matthew ha un lavoro come gentile vicesceriffo.
Matthew ha un lavoro come gentile muratore.
Matthew ha un lavoro come gentile fabbro.
Matthew ha un lavoro come gentile architetto.
Matthew ha un lavoro come gentile infermiere.
Matthew ha un lavoro come gentile babysitter.
Matthew ha un lavoro come gentile cosmetologo.
Matthew ha un lavoro come gentile nutrizionista.
Matthew ha un lavoro come gentile governante.
Matthew ha un lavoro come gentile segretaria.
Matthew ha un lavoro come gentile psicologo.
Matthew ha un lavoro come gentile cassiera.
Matthew ha un lavoro come gentile venditore.
Matthew ha un lavoro come gentile receptionist.

Matthew ha un lavoro come un bell'avvocato.
Matthew ha un lavoro come un bellissimo ingegnere.
Matthew ha un lavoro come un bel meccanico.
Matthew ha un lavoro come un bel sindaco.
Matthew ha un lavoro come un bel matematico.
Matthew ha un lavoro come un bellissimo ministro.
Matthew ha un lavoro come bellissimo vicesceriffo.
Matthew ha un lavoro come bellissimo muratore.
Matthew ha un lavoro come un bel fabbro.
Matthew ha un lavoro come bellissimo architetto.
Matthew ha un lavoro come una bella infermiera.
Matthew ha un lavoro come una bella babysitter.
Matthew ha un lavoro come una bella cosmetologa.
Matthew ha un lavoro come bella nutrizionista.
Matthew ha un lavoro come bella governante.
Matthew ha un lavoro come bella segretaria.
Matthew ha un lavoro come una bella psicologa.
Matthew ha un lavoro come una bella cassiera.
Matthew ha un lavoro come bellissimo commesso.
Matthew ha un lavoro come una bella receptionist.
Matthew ha un lavoro come fragile avvocato.
Matthew ha un lavoro come fragile ingegnere.
Matthew ha un lavoro come fragile meccanico.
Matthew ha un lavoro come fragile sindaco.
Matthew ha un lavoro come fragile matematico.
Matthew ha un lavoro come fragile ministro.
Matthew ha un lavoro come fragile vicesceriffo.
Matthew ha un lavoro come fragile muratore.
Matthew ha un lavoro come fragile fabbro.
Matthew ha un lavoro come fragile architetto.
Matthew ha un lavoro come fragile infermiere.
Matthew ha un lavoro come fragile babysitter.
Matthew ha un lavoro come fragile cosmetologo.
Matthew ha un lavoro come fragile nutrizionista.

Matthew ha un lavoro come fragile governante.
Matthew ha un lavoro come fragile segretaria.
Matthew ha un lavoro come fragile psicologo.
Matthew ha un lavoro come fragile cassiere.
Matthew ha un lavoro come fragile commesso.
Matthew ha un lavoro come fragile receptionist.
Matthew ha un lavoro come avvocato amorevole.
Matthew ha un lavoro come ingegnere amorevole.
Matthew ha un lavoro come amorevole meccanico.
Matthew ha un lavoro come sindaco amorevole.
Matthew ha un lavoro come amorevole matematico.
Matthew ha un lavoro come amorevole ministro.
Matthew ha un lavoro come amorevole vicesceriffo.
Matthew ha un lavoro come amorevole muratore.
Matthew ha un lavoro come amorevole fabbro.
Matthew ha un lavoro come architetto amorevole.
Matthew ha un lavoro come amorevole infermiere.
Matthew ha un lavoro come amorevole babysitter.
Matthew ha un lavoro come amorevole cosmetologo.
Matthew ha un lavoro come amorevole nutrizionista.
Matthew ha un lavoro come amorevole governante.
Matthew ha un lavoro come amorevole segretaria.
Matthew ha un lavoro come amorevole psicologo.
Matthew ha un lavoro come amorevole cassiera.
Matthew ha un lavoro come amorevole venditore.
Matthew ha un lavoro come amorevole receptionist.
Matthew ha un lavoro come audace avvocato.
Matthew ha un lavoro come audace ingegnere.
Matthew ha un lavoro come audace meccanico.
Matthew ha un lavoro come audace sindaco.
Matthew ha un lavoro come audace matematico.
Matthew ha un lavoro come audace ministro.
Matthew ha un lavoro come audace vicesceriffo.
Matthew ha un lavoro come audace muratore.

Matthew ha un lavoro come audace fabbro.
Matthew ha un lavoro come audace architetto.
Matthew ha un lavoro come audace infermiere.
Matthew ha un lavoro come audace babysitter.
Matthew ha un lavoro come audace cosmetologo.
Matthew ha un lavoro come audace nutrizionista.
Matthew ha un lavoro come audace governante.
Matthew ha un lavoro come audace segretaria.
Matthew ha un lavoro come audace psicologo.
Matthew ha un lavoro come audace cassiera.
Matthew ha un lavoro come audace venditore.
Matthew ha un lavoro come audace receptionist.
Matthew ha un lavoro come avvocato aggressivo.
Matthew ha un lavoro come ingegnere aggressivo.
Matthew ha un lavoro come meccanico aggressivo.
Matthew ha un lavoro come sindaco aggressivo.
Matthew ha un lavoro come aggressivo matematico.
Matthew ha un lavoro come ministro aggressivo.
Matthew ha un lavoro come aggressivo vicesceriffo.
Matthew ha un lavoro come aggressivo muratore.
Matthew ha un lavoro come aggressivo fabbro.
Matthew ha un lavoro come aggressivo architetto.
Matthew ha un lavoro come aggressivo infermiere.
Matthew ha un lavoro come aggressivo babysitter.
Matthew ha un lavoro come aggressivo cosmetologo.
Matthew ha un lavoro come aggressivo nutrizionista.
Matthew ha un lavoro come governante aggressivo.
Matthew ha un lavoro come aggressivo segretario.
Matthew ha un lavoro come psicologo aggressivo.
Matthew ha un lavoro come aggressivo cassiere.
Matthew ha un lavoro come venditore aggressivo.
Matthew ha un lavoro come aggressivo receptionist.
Matthew ha un lavoro come orgoglioso avvocato.
Matthew ha un lavoro come orgoglioso ingegnere.

Matthew ha un lavoro come orgoglioso meccanico.
Matthew ha un lavoro come orgoglioso sindaco.
Matthew ha un lavoro da orgoglioso matematico.
Matthew ha un lavoro come orgoglioso ministro.
Matthew ha un lavoro come fiero vicesceriffo.
Matthew ha un lavoro come fiero muratore.
Matthew ha un lavoro come orgoglioso fabbro.
Matthew ha un lavoro come orgoglioso architetto.
Matthew ha un lavoro come orgoglioso infermiere.
Matthew ha un lavoro come fiero babysitter.
Matthew ha un lavoro come orgoglioso cosmetologo.
Matthew ha un lavoro come orgoglioso nutrizionista.
Matthew ha un lavoro come orgogliosa governante.
Matthew ha un lavoro come orgogliosa segretaria.
Matthew ha un lavoro come orgoglioso psicologo.
Matthew ha un lavoro come orgoglioso cassiere.
Matthew ha un lavoro come fiero commesso.
Matthew ha un lavoro come fiero receptionist.
Matthew ha un lavoro come ambizioso avvocato.
Matthew ha un lavoro come ambizioso ingegnere.
Matthew ha un lavoro come ambizioso meccanico.
Matthew ha un lavoro come ambizioso sindaco.
Matthew ha un lavoro come ambizioso matematico.
Matthew ha un lavoro come ministro ambizioso.
Matthew ha un lavoro come ambizioso vicesceriffo.
Matthew ha un lavoro come ambizioso muratore.
Matthew ha un lavoro come ambizioso fabbro.
Matthew ha un lavoro come ambizioso architetto.
Matthew ha un lavoro come ambizioso infermiere.
Matthew ha un lavoro come ambizioso babysitter.
Matthew ha un lavoro come ambizioso cosmetologo.
Matthew ha un lavoro come ambizioso nutrizionista.
Matthew ha un lavoro come ambizioso governante.
Matthew ha un lavoro come ambizioso segretario.

Matthew ha un lavoro come ambizioso psicologo.
Matthew ha un lavoro come ambizioso cassiere.
Matthew ha un lavoro come ambizioso venditore.
Matthew ha un lavoro come ambizioso receptionist.
Matthew ha un lavoro come un determinato avvocato.
Matthew ha un lavoro come ingegnere determinato.
Matthew ha un lavoro come meccanico determinato.
Matthew ha un lavoro come sindaco determinato.
Matthew ha un lavoro come matematico determinato.
Matthew ha un lavoro come ministro determinato.
Matthew ha un lavoro come determinato vicesceriffo.
Matthew ha un lavoro come determinato muratore.
Matthew ha un lavoro come un determinato fabbro.
Matthew ha un lavoro come architetto determinato.
Matthew ha un lavoro come infermiere determinato.
Matthew ha un lavoro come determinato babysitter.
Matthew ha un lavoro come determinato cosmetologo.
Matthew ha un lavoro come determinato nutrizionista.
Matthew ha un lavoro come governante determinato.
Matthew ha un lavoro come determinato segretario.
Matthew ha un lavoro come psicologo determinato.
Matthew ha un lavoro come cassiere determinato.
Matthew ha un lavoro come venditore determinato.
Matthew ha un lavoro come determinato receptionist.
Matthew ha un lavoro come un forte avvocato.
Matthew ha un lavoro come forte ingegnere.
Matthew ha un lavoro come forte meccanico.
Matthew ha un lavoro come forte sindaco.
Matthew ha un lavoro come forte matematico.
Matthew ha un lavoro come forte ministro.
Matthew ha un lavoro come forte vicesceriffo.
Matthew ha un lavoro come forte muratore.
Matthew ha un lavoro come forte fabbro.
Matthew ha un lavoro come forte architetto.

Matthew ha un lavoro come forte infermiere.
Matthew ha un lavoro come forte babysitter.
Matthew ha un lavoro come forte cosmetologo.
Matthew ha un lavoro come forte nutrizionista.
Matthew ha un lavoro come forte governante.
Matthew ha un lavoro come forte segretaria.
Matthew ha un lavoro come forte psicologo.
Matthew ha un lavoro come forte cassiera.
Matthew ha un lavoro come forte venditore.
Matthew ha un lavoro come forte receptionist.
Matthew ha un lavoro come avvocato competente.
Matthew ha un lavoro come ingegnere competente.
Matthew ha un lavoro come meccanico competente.
Matthew ha un lavoro come sindaco competente.
Matthew ha un lavoro come matematico competente.
Matthew ha un lavoro come ministro competente.
Matthew ha un lavoro come vicesceriffo competente.
Matthew ha un lavoro come muratore competente.
Matthew ha un lavoro come fabbro competente.
Matthew ha un lavoro come architetto competente.
Matthew ha un lavoro come infermiere competente.
Matthew ha un lavoro come babysitter competente.
Matthew ha un lavoro come estetista competente.
Matthew ha un lavoro come nutrizionista competente.
Matthew ha un lavoro come governante competente.
Matthew ha un lavoro come segretaria competente.
Matthew ha un lavoro come psicologo competente.
Matthew ha un lavoro come cassiere competente.
Matthew ha un lavoro come venditore competente.
Matthew ha un lavoro come receptionist competente.
Matthew ha un lavoro come avvocato crudele.
Matthew ha un lavoro come ingegnere crudele.
Matthew ha un lavoro come meccanico crudele.
Matthew ha un lavoro come sindaco crudele.

Matthew ha un lavoro come crudele matematico.
Matthew ha un lavoro come crudele ministro.
Matthew ha un lavoro come crudele vicesceriffo.
Matthew ha un lavoro come crudele muratore.
Matthew ha un lavoro come fabbro crudele.
Matthew ha un lavoro come crudele architetto.
Matthew ha un lavoro come crudele infermiere.
Matthew ha un lavoro come crudele babysitter.
Matthew ha un lavoro come crudele cosmetologo.
Matthew ha un lavoro come crudele nutrizionista.
Matthew ha un lavoro come crudele governante.
Matthew ha un lavoro come crudele segretaria.
Matthew ha un lavoro come crudele psicologo.
Matthew ha un lavoro come crudele cassiere.
Matthew ha un lavoro come venditore crudele.
Matthew ha un lavoro come crudele receptionist.

10.5 Sentences_with_wider_context_GoogleTranslate

Si chiama Martha ed è di Londra. È la più bella ministro della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È il ministro più dolce della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È la più orgogliosa ministro della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È il ministro più determinato della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'architetto più bello della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'architetto più dolce della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.

Si chiama Martha ed è di Londra. È l'architetto più orgoglioso della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'architetto più determinato della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'infermiera più bella della sua città e adora davvero il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'infermiera più dolce della sua città e adora davvero il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'infermiera più orgogliosa della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È l'infermiera più determinata della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È la commessa più bella della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È la commessa più dolce della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È la commessa più orgogliosa della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha ed è di Londra. È la commessa più determinata della sua città e gode davvero del suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam ed è di Londra. È il ministro più bello della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il ministro più dolce della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il ministro più orgoglioso della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il ministro più determinato della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'architetto più bello della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'architetto più dolce della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'architetto più orgoglioso della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.

Si chiama Liam ed è di Londra. È l'architetto più determinato della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'infermiera più bella della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'infermiera più dolce della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'infermiere più orgoglioso della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È l'infermiera più determinata della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il venditore più bello della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il commesso più dolce della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il venditore più orgoglioso della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.
Si chiama Liam ed è di Londra. È il venditore più determinato della sua città e gode davvero del suo lavoro. Va anche d'accordo con i suoi colleghi.

10.6 Sentences_with_wider_context_DeepL

Si chiama Martha e viene da Londra. È il più bel ministro della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È il ministro più dolce della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È il ministro più orgoglioso della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È il ministro più determinato della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'architetto più bello della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.

Si chiama Martha e viene da Londra. È l'architetto più dolce della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'architetto più orgoglioso della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'architetto più determinato della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'infermiera più bella della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'infermiera più dolce della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'infermiera più orgogliosa della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È l'infermiera più determinata della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È la più bella commessa della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È la più dolce commessa della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È la commessa più orgogliosa della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Martha e viene da Londra. È la commessa più determinata della sua città e le piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il più bel ministro della sua città e ama molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il ministro più dolce della sua città e ama davvero il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il ministro più orgoglioso della sua città e gli piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il ministro più determinato della sua città e ama il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'architetto più bello della sua città e gli piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'architetto più dolce della sua città e ama molto il suo lavoro. Va d'accordo anche con i suoi colleghi.

Si chiama Liam e viene da Londra. È l'architetto più orgoglioso della sua città e ama davvero il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'architetto più determinato della sua città e ama il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'infermiere più bella della sua città e ama molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'infermiere più dolce della sua città e gli piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'infermiere più orgoglioso della sua città e ama davvero il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È l'infermiere più determinato della sua città e ama molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il più bel commesso della sua città e gli piace molto il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il venditore più dolce della sua città e ama il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il venditore più orgoglioso della sua città e ama il suo lavoro. Va d'accordo anche con i suoi colleghi.
Si chiama Liam e viene da Londra. È il venditore più determinato della sua città e ama il suo lavoro. Va d'accordo anche con i suoi colleghi.

Abstract (English)

In the last few years scholars have concentrated on gender bias in the field of machine translation (MT), namely the phenomenon why MT-tools produce sexist outputs. As a consequence, male and female-dominated professions are often rendered in the false gender form. The current work focuses on the effect of adjectives related to gender roles on the translation of professions. To this end, three corpora were created in English and they are composed of simple sentences with male and female subjects as well as of professions. The second corpus also comprises adjectives whereas the third corpus incudes longer sentences with more context. The corpora were translated into Italian with Google Translate and DeepL. Data on the amount of correct translations were collected and the performances of the two MT-tools were compared. Some of the results of the study were that: (1) both Google Translate and DeepL translations result in being gender bias, especially against women, (2) the use of adjectives can either improve or worsen translation outputs in terms of gender form, (3) outputs improved, when male-dominated professions were to be translated in the female form, (4) outputs got worse, when female-dominated professions were to be translated in the male form. The main conclusion of the work is that not only adjectives impact on translation outputs, but the related gender roles must also be taken into account (e.g. male-dominated professions are more likely to be translated in the female form if the co-occurring adjective also relates to a female role).

Abstract (Deutsch)

In den letzten Jahren haben sich ForscherInnen im Bereich Maschinelle Übersetzung (MÜ) auf Gender-Bias konzentriert, nämlich das Phänomenon aufgrund dessen MÜ-Systeme sexistische Übersetzungen liefern. Stereotypische Männer- und Frauenberufe werden deswegen oft mit der falschen Geschlechterform versehen. Die vorliegende Arbeit untersucht die Auswirkungen von sich auf Geschlechterrollen beziehenden Adjektiven auf die Übersetzung von Berufsbezeichnungen. Zu diesem Zweck wurden drei Korpora auf Englisch erstellt, die aus kurzen Sätzen mit weiblichen und männlichen Akteuren und Berufsbezeichnungen bestehen. Das zweite Korpus enthält auch Adjektive und das dritte weist hingegen längere Sätze mit weiterem Kontext auf. Die Korpora wurden ins Italienische mittels Google Translate und

DeepL übersetzt. Daten über die Anzahl von korrekten Übersetzungen wurden aufgenommen und die Leistungen der zwei verwendeten MÜ-Tools wurden verglichen. Aus der Studie hat sich unter anderem ergeben, dass (1) beide Google Translate und DeepL Gender-Bias aufweisen, insbesondere gegenüber Frauen, (2) die Verwendung von Adjektiven Übersetzungen verbessern oder verschlechtern kann, was die Geschlechtsform angeht, (3) die Verbesserungen stattgefunden haben, wenn stereotypische Männerberufe in die weibliche Form zu übertragen waren, (4) die Verschlechterungen stattgefunden haben, wenn stereotypische Frauenberufe in die männliche Form zu übertragen waren. Die wichtigste Schlussfolgerung der vorliegenden Masterarbeit ist nicht nur, dass die Verwendung von Adjektiven sich auf die Übersetzung auswirkt, sondern auch, dass die damit verbundenen Geschlechterrollen berücksichtigt werden müssen (z.B. werden Männerberufe öfter in die weibliche Form übersetzt, wenn sich das gemeinsam auftretende Adjektiv auf eine weibliche Rolle bezieht).