UNIVERSITY OF GOTHENBURG

DEPARTMENT OF PHILOSOPHY, LINGUISTICS AND THEORY OF SCIENCE

RESPONSIBLE WOMEN AND ANALYTICAL MEN

Developing Swedish Gendered Lexica for Detection of Gender Bias in Job Advertisements

Saga Hansson & Konstantinos Mavromatakis

Master's Thesis:	30 credits
Programme:	Master's Programme in Language Technology
Level:	Advanced level
Semester and year:	Spring, 2021
Supervisor	Peter Ljunglöf
Examiner	Eleni Gregoromichelaki
Report number	(number will be provided by the administrators)
Keywords	Gender Bias Detection, Machine Translation, Fairness in AI

Abstract

In this research, we examine gender bias in Swedish job advertisements with the use of gendered lexica. Gendered lexica can be employed to ascertain whether job advertisements are written with words that are associated with more masculine or feminine traits. The main purpose of this study is to investigate translating English gendered lexica using different machine translation methods: Google Translate, frequency-based and word-embedding-based. In the absence of a gold standard, we evaluated the translations by conducting quantitative and qualitative experiments. The embedding-based translation was evaluated as the most consistent method for the development of gendered lexica. Further testing of the embedding-based lexicon showed that Swedish job advertisements seem to be written with more feminine coded words, regardless of the gender of the majority of the workers in the advertised occupation. Advertisements for technical universities, specifically, tend to be written with more masculine coded words, while advertisements for universities that offer a wider range of education contain more feminine coded words.

Acknowledgements

We would like to thank our supervisor Peter Ljunglöf for guiding us through our research, providing us with helpful advice and deciphering our thoughts and ideas.

Furthermore, we would like to thank Ann, Nikolai, Helena and Christian for helping out with various tasks.

Finally, we want to thank our friends, family and classmates. You have all helped us and kept us motivated with your kind words and feedback.

Contents

1 Introduction

1.1 Focus

The main goal of our thesis is to investigate if it is possible to translate publicly available gendered lexica to Swedish, and to see whether these lexica can be used to detect gender bias – prejudice and unfair treatment that a person faces on the basis of their gender – in job advertisements. A gendered lexicon is a collection of words that can be categorised as feminine or masculine coded, based on whether they are associated with typically feminine or masculine traits. It can be utilised to score a document by aggregating gendered scores of all the gendered words in the document.

To translate English gendered lexica, developed through previous research on the topic, four different translation methods, which capture word meaning in different ways, were utilised and evaluated qualitatively and quantitatively. Then, the Swedish version of the gendered lexicon with the most consistent performance, according to the evaluation experiments, was employed to assess job advertisements for gender bias.

1.2 Motivation

Research within the field of natural language processing (NLP) has predominantly revolved around developing methods that detect and mitigate gender bias for English. However, similar research on languages other than English is limited. Therefore, we aim to bring the attention to gender bias in Swedish, and simultaneously motivate other researchers to explore gender bias across different languages, by providing tools and experiments to do so.

Previous work (Lindmark & Sundin, 2013) on gendered wording in job advertisements for the Swedish language employed linguistic intuition as a means of developing a gendered lexicon. The two main shortcomings of their approach are that the lexicon was not evaluated, and was only tested on a small sample of advertisements. Due to the lack of motivation behind Lindmark & Sundin's methodology regarding the creation of their Swedish gendered lexicon, we have developed Swedish gendered lexica, based on previous research, which is reported in this thesis.

1.3 Contributions

Our main contributions are listed below:

- In our work, we set out to investigate if it is possible to develop gendered lexica in a language other than English using different translation methods. We found that utilising more complex meaning representations in the translation of the lexica yields the most stable results.
- We developed a pipeline to automate the creation of gendered lexica and the detection of gender bias in a large corpus of job advertisements.
- In the absence of a gold standard, we propose a set of experiments, both quantitative and qualitative,

to evaluate the performance of each translation method. These experiments may well be used for the evaluation of gendered lexica developed in other languages.

- Using the best translation method, we generated two gendered lexica for Swedish, translated from English lexica created by Gaucher et al. (2011) and Pietraszkiewicz et al. (2019).[1]

1.4 Research questions

The questions we intend to examine in this research are:

- Can we develop gendered lexica in Swedish that perform similarly to the English lexica, despite the lack of a gold standard?
- Are more complex word representations able to capture gender bias more consistently?
- According to Gaucher et al. (2011), the majority of English job advertisements tend to contain more masculine coded than feminine coded words, irrespective of which gender dominates the occupations. Does the same hold for Swedish?
- Are function words as good an indicator of agency and communion as content words? By agency, we mean the pursuit of independence through assertiveness and control, while communion is defined as the participation in a large community through strong interpersonal relationships.

1.5 Delimitations

In the absence of sufficient research on gender neutral wording in job advertisements, the focus of this thesis is feminine and masculine coded wording. Thus, the use of terms, such as women and men, female and male, feminine and masculine, also follows a binary gender division.

[1] The two Swedish gendered lexica can be found here: https://github.com/sagahansson/mlt-thesis/tree/main/swedish-lexica.

2 Background

In this Section, we present background information on the topics of sex and gender, bias, gender bias and previous research into gender bias. We start by discussing the difference between sex and gender, the term *gendered* and the difference between gendered language and gendered wording. Following that, we detail bias and, more specifically, gender bias. Finally, we present the work of Gaucher et al. (2011) and Pietraszkiewicz et al. (2019), who both explored the area of gender bias in job advertisements.

2.1 Sex & gender

2.1.1 Differentiating between sex & gender

Biological sex is related with the external and internal anatomy of a person and is assigned at birth (American Psychological Association, 2014). According to the most common taxonomy, there are two biological sexes: female and male. However, in certain countries (for instance, India and Malta) a third sex, namely intersex, is recognised. Intersex people possess characteristics of both sexes, suggesting that a binary sex division does not adequately reflect modern society.

Sometimes misunderstood as the same thing as sex, gender is a term introduced in the 1960s that pertains to cultural, and psychological characteristics that distinguish femininity from masculinity through particular social contexts (Pilcher & Whelehan, 2004). Two fundamental concepts of gender are *gender identity* and *gender expression*: the former represents our internal feelings and self-perception; the latter is defined as how we communicate our gender by choice of clothes, hair style and behaviour (Gender Spectrum, 2019). A common misconception is that a person's gender always aligns with our preconceived notions of gender and biological sex. That is, one might assume that if a person wears a dress, which is a piece of clothing typically associated with women or girls, that person's gender is feminine and, therefore, their biological sex is female. This may not be necessarily a true statement – a person's gender identity, biological sex and how we interpret their way of expressing gender might align in specific situations, but this is not always the case.

2.1.2 The term *gendered*

Gendered is a term typically attributed to items or matters that possess specific characteristics based on the gender associated with them and strongly reflect the experiences related to a particular gender more than the others (Pilcher & Whelehan, 2004). The description of an object as *gendered* places this object in a fixed state and essentially separates it from an item belonging to a different gender. Many concepts and items of our everyday life are gendered. For instance, a dress and a suit are considered to be gendered pieces of clothing, since they are associated with femininity and masculinity respectively.

2.1.3 Gendered language and gendered wording

The vast majority of existing research on *gendered language* and *gendered wording* has been conducted with a focus on the English language. Below we present a summary of critical work on gendered language and gendered wording for English and Swedish.

Gendered language – in a non-grammatical sense – refers to the differences in language use between women and men. For English, the concept of gendered language was introduced in the 1970s. Based on the analysis of linguistic data with discourse analysis methods, Lakoff (1973) showed that each gender of the binary model (i.e., feminine and masculine) has its own vernacular.[2] According to her, women tend to produce more polite expressions and tag questions in their speech, while also being more flexible in their conversations. Women are more inclined to distinguish more fine-grained details of their surroundings, such as different shades of colours (Lakoff, 1973). Women's utterances are longer, display more complex structure and contain a greater number of 'evaluative adjectives' (Haas, 1979, p. 622). Men are more crude and direct, likelier to interrupt their interlocutor and frequently use slang words (Labov, 1966). Furthermore, Haas (1979) noted that men tend to swear more, as opposed to women who are conservative in their speech (e.g. avoiding the use of profanities) and often opt for euphemisms and paraphrasing on topics they are sensitive about (e.g. body parts). Utterances of both genders can be expressive, but unlike women's, that tend to be more enthusiastic, men's utterances convey anger. Assertiveness in one's speech is a trait common to both genders, but while men are praised for it (Delamater & Mcnamara, 1986), women are looked down upon and characterised as manipulative (Haas, 1979). To avoid such discrimination, women try to be more tentative. Naturally, individual exceptions from the noted patterns do occur (i.e., some women swear more than some men).

On the other hand, *gendered wording* is based on the differences between linguistic expressions that are used to describe behavioral traits of women and men according to gender stereotypes (Rudman & Kilianski, 2000). The former group is associated with *communion*, while the latter is linked with *agency*. *Agentic* wording is used to describe a state of independence, self-advancement, control, confidence and assertiveness, while *communal* wording expresses a strong sense of community and belonging and a value for interpersonal relationships (Pietraszkiewicz et al., 2019). There are, of course, individual exceptions to the rules: agency is not exclusive to men, as communion is not exclusive to women (Phelan et al., 2008).

In the context of Swedish, research on language and gender has mainly revolved around gendered language, as opposed to gendered wording. Einarsson (2009) and Norrby & Håkansson (2015) provide extensive overviews of the research on gendered language in Swedish. Differences between women/girls and men/boys have been noted in spoken language (Eriksson, 1997, Nordberg, 1972; Nordenstam, 1990, Norrby, 1998), written language (Hultman & Westman, 1977, Allén, 1977) and classroom interactions (Einarsson & Hultman, 1984).

Research on gendered wording has been conducted by Einarsson (1979) and Sundgren (2007), who both examined how men and women that share the same traits are perceived by others. In Einarsson's study, the man was perceived as more trustworthy and competent, whereas the woman was perceived as more humane. Sundgren found that the man was seen as more educated and logical, as opposed to the woman, who was seen as more reliable and committed.

[2] Lakoff used the term sex when talking about linguistic differences among men and women, however, she stresses that the variation in linguistic behavior is not biologically based. 'It is sometimes claimed that there is a biological basis for this behavior difference, though I don't believe conclusive evidence exists that the early differences in behavior that have been observed are not the results of very different treatment of babies of the two sexes from the beginning' (Lakoff, 1973, p. 8).

2.2 Bias

A person's upbringing and environment are highly likely to make them susceptible to certain biases and stereotypes. On an interpersonal level, the unfair and prejudicial treatment of an individual or a group based on features they acquire as part of said group is known as *individual bias*. *Institutional bias*, on the other hand, is when any kind of discrimination towards members of particular groups is embedded in institutional policies, procedures, practices and laws (Henry, 2010).

Data is produced, collected and handled by humans, and thereby often reflects their biases, especially when there has not been enough effort to mitigate the bias. A very detailed categorisation of data bias can be found in Mehrabi et al. (2019). They listed 23 different definitions of bias from previous literature and proposed a three-way categorisation of bias, consisting of bias in data, algorithms and user interaction, which generates a feedback loop.[3] Crawford (2017) proposed a two-way taxonomy of the possible harm bias can cause. The first category is *allocational harm*, which occurs "when a system [unintentionally] allocates or withholds certain groups an opportunity or a resource" (Crawford, 2017) because of the way that system has been trained. Crawford presented an example of allocational harm in which women or those under the age of 30 are denied a mortgage by a biased system. The implications of this type of harm are rather straightforward – certain groups are denied access to certain resources, creating an unfair environment. The second category is *representational harm* – this occurs "when systems reinforce the subordination of some groups along the lines of identity – [...] race, class, gender, etc" (Crawford, 2017). As an example of representational harm, Crawford brought up research by Sweeney (2013), who discovered racial bias in advertisements provided by Google when querying for different names. Querying for names typically born by black people yielded advertisements that are "suggestive of arrest" (Sweeney, 2013, 22) to a much larger extent than when querying for names typically born by white people. This is an example of how representational harm is dangerous given its influence on society and our views – Crawford suggested that if a potential employer would come across these search results, they could have damning consequences for the applicant whose name is unfairly associated with an arrest.

2.3 Gender bias

Gender identity and expression are two of the most common grounds for facing prejudice, also known as *gender bias*. Most frequently, gender bias negatively affects women (Pilcher & Whelehan, 2004). On the other hand, cis-gender men seem to benefit the most from such implicit stereotypes. The majority of the research on gender bias revolves around bias in professional environments, but unfair treatment of women has also been documented in the household and in the health and education sectors, according to Rothchild (2007).

Evidence of gender bias can also be found in NLP. Sun et al. (2019) stated that both data and algorithms reflect the prejudice of their creators. From the underrepresentation of female entities in datasets to gender-biased predictions, the possible effects of gender bias in NLP are devastating: machines reinforce stereotypes

[3] A feedback loop is when the output of a model is used as input for future tasks.

and fail to represent all groups equally, preventing the development of an equal society.

2.4 Gendered wording in English job advertisements

Gaucher et al. (2011) investigated whether or not the wording of job advertisements is biased with regards to gender, by conducting a series of studies involving gendered wording and advertisements collected from two Canadian job listing websites (1493 advertisements) and from the University of Waterloo (3640 advertisements). All advertisements (5133 in total) in their studies came from either female or male dominated areas, which enables testing for the presence of covariance between gendered wording and the gender of the majority of the workers in a field.

A gendered lexicon, containing 42 agentic and 40 communal words, was created and used to assess which gender an advertisement is associated with. Agentic words, such as 'independent' and 'decisive', are commonly associated with masculine traits, whereas communal words, like 'polite' and 'responsible', are associated with feminine traits. Therefore, agentic and communal are used interchangeably with masculine and feminine coded, respectively.

Gaucher et al. examined the distribution of the agentic and communal words in relation to the length of each collected advertisement and assigned feminine and masculine scores to an advertisement based on the percentage of masculine or feminine associated words. They calculated mean feminine and masculine scores for all available data and tested these scores for correlation. As there is no mention of how advertisements with no gendered words or advertisements that contain an equal number of gendered words were dealt with, we assume that they were assigned a gendered score of 0 (i.e., neither feminine, nor masculine).

The studies conducted by Gaucher et al. (2011) show that, irrespective of which gender dominates the occupation, the majority of the job advertisements tend to contain more masculine coded words than feminine coded words, with a slightly higher frequency of masculine coded words in the advertisements for male dominated occupations. Feminine coded words were effectively as prevalent in advertisements for female dominated occupations as in advertisements for male dominated occupations.

In an attempt to explain the underlying reasons behind the differences in distribution of feminine and masculine coded vocabulary in job advertisements for male dominated professions, Gaucher et al. (2011) used two social theories: social dominance theory (SDT; Sidanius & Pratto, 1999) and social role theory (SRT; Eagly, 1987). SDT is based on the observation that distinct "group-based social hierarchies" (Sidanius & Pratto, 1999, p. 31) govern different human social groups, often leading to power inequities. SRT relates all differences and similarities of women's and men's roles, behaviors and traits to a society's division of labour by gender.

In line with SDT, the use of gendered wording in job advertisements may unintentionally maintain a hierarchical system and contribute to social inequalities. Such practices, despite not being overt, can still discourage the other groups from attempting to gain access to male dominated fields, thus limiting their stance in society.

SRT, on the other hand, attributes the existence of gendered wording to traditional gender roles. This theory suggests that men and women tend to follow career paths that line up with these traditional gender roles, leading to more women in caretaking professions and more men in professions involving manual labour. Female and male dominated fields are described as requiring communion and agency respectively, which inevitably causes the wording of job advertisements for male dominated occupations to be more agentic, and job advertisements for female dominated occupations to be worded in a more communal manner.

Gaucher et al. (2011) conducted three additional studies to examine how people perceive advertisements for male and female dominated occupations that have been modified to include more agentic or more communal words. Specifically, they investigated if gendered wording reinforces preconceptions about the gender of the majority of workers of a female or male dominated field and whether the use of such wording affects how women feel about applying to a job position based on appeal, personal skills and belongingness. Their results indicate that the research subjects expected fewer women to be employed within occupations advertised with more masculine than feminine wording, and that traditionally male dominated occupations are likely to employ fewer women. With regards to the appeal the modified job advertisements had to the research subjects, advertisements for male dominated occupations appeared less interesting, regardless of the gender of the study participant. Furthermore, advertisements with more masculine coded words were less attractive to women. The study participants also expected to experience a stronger sense of belongingness in female dominated occupations, as opposed to neutral and male dominated occupations. Specifically for women, a stronger sense of belongingness came from job advertisements that are worded with more feminine coded words. Finally, according to the female participants, the skills required for an occupation were considered to be the same, regardless of the wording of the advertisements.

2.5 Agency and communion dictionaries for gender bias detection

Pietraszkiewicz et al. (2019) developed Agency and Communion dictionaries following the method used for creating Linguistic Inquiry and Word Count (LIWC) dictionaries (Pennebaker et al., 2015). LIWC dictionaries represent different psychological concepts, such as happiness, sadness, relationship etc., and are used to analyse various types of text by examining the distribution of words from different categories. In Pietraszkiewicz et al. (2019), the words were collected and evaluated by a group of eight experts in social psychology and language. The Agency dictionary contains 192 words, whereas the Communion dictionary contains 184 words. As opposed to the lexicon created by Gaucher et al. (2011) that mainly contains adjectives, the dictionaries created by Pietraszkiewicz et al. (2019) also contain verbs, since "agency [...] is better captured in verbs than in adjectives and nouns" (Pietraszkiewicz et al., 2019, p. 873).

Semantic similarity scores were calculated between agentic and communal occupations provided by Fiske & Dupree (2014), and the entries of the Agentic and Communion dictionaries. These semantic similarity scores were compared with the communal and agentic ratings given to the occupations by participants in Fiske & Dupree (2014) and were found to be strongly correlated, suggesting that the Pietraszkiewicz dictionaries effectively capture Agency and Communion.

Finally, the Agency and Communion dictionaries were tested against job advertisements for female and male dominated occupations to examine whether the occupations are presented differently with respect to gender.

Similarly to the study by Gaucher et al. (2011), this analysis was performed by counting the total number of words of each advertisement, followed by counting the number of the agentic and communal words of each advertisement. Thereafter, the percentages of agentic and communal words per advertisement were calculated. The most significant finding of their analysis is that advertisements for male dominated occupations contain more agentic words, while advertisements for female dominated occupations are worded with more communal terms.

2.6 Gendered wording in Swedish job advertisements

In their bachelor thesis, Lindmark & Sundin (2013) examined differences in the wording of job advertisements for female and male dominated occupations by translating the English Gaucher et al. (2011) gendered lexicon into Swedish using Google Translate and their linguistic intuition, and performing discourse analysis on a total of 60 job advertisements. They noted that tasks in job advertisements for male dominated professions were often clearly explained, whereas tasks in job advertisements for female dominated professions were more vague. Personal qualities, qualifications and mentions about previous experience were extracted from the job advertisements and analysed, revealing, for example, that the demand for independence is higher in job advertisements for male dominated occupations and that the demand for the ability to cooperate is higher in job advertisements for female dominated occupations. However, there is a demand for both qualities in job advertisements for both female and male dominated occupations (the ability to cooperate and the ability to work in a team are the fifth and sixth most frequently sought-after skill mentioned in job advertisements for male dominated occupations, and independence is the third most sought-after quality mentioned in job advertisements for female dominated occupations). The top five most sought-after qualities mentioned in job advertisements for male dominated professions are independence, focus on end-product, flexibility, service mindedness and ability to cooperate. On the other hand, the top five for female dominated professions are ability to cooperate, flexibility, independence, caregiving and service mindedness. Similar research on the topic of gendered wording in Swedish job advertisements has been conducted by Andersson & Nilsson (2014), Chamera & Kühnemann (2016), Östlund (2018) and Martinsson (2021).

University	English ads	Swedish ads	Cleaned English ads	Cleaned Swedish ads	Time Period
CTH	5406	1731	4841	1632	2010-2021
GU	2549	9235	2426	8806	2017-2021
KTH	1	5736	1	5633	2014-2021
LIU	1203	7492	1157	7171	2013-2021
LNU	749	2664	687	2383	2011-2021
LTU	911	2948	902	2878	2011-2021
LU	1853	5696	1823	5285	2015-2021
SLU	527	1454	525	1415	2017-2021
SU	2059	7582	2057	7375	2015-2021
UU	18	9532	17	9522	2013-2021
Total	15276	54070	14436	52100	

Table 1: Distribution of job advertisements before and after cleaning. Cleaning is described in Section 4.1.1.1

3 Existing Resources

In the following sections we present the existing resources that were used in this project. First, we describe our data: the job advertisements and the publicly available gendered lexica. Thereafter, we present the translation methods we used to generate Swedish gendered lexica. We then describe the Gender Decoder tool used to assign a gender to the job advertisements. Finally, we detail statistics from Statistics Sweden used to determine if an occupation field is male or female dominated.

3.1 Job advertisements

A collection of job advertisements (66536 in total), written in English and Swedish, was gathered from ten different universities across Sweden. Metadata, such as date of release, area of expertise, advertisement id (corresponding to a project that the applicant will work with), is also included. The list of universities which posted the job advertisements include Chalmers University of Technology (CTH), University of Gothenburg (GU), KTH Royal Institute of Technology (KTH), Linköping University (LIU), Linnaeus University (LNU), Luleå University of Technology (LTU), Lund University (LU), Swedish University of Agricultural Sciences (SLU), University of Stockholm (SU) and Uppsala University (UU). The distribution of advertisements from the 10 universities can be seen in Table 1. As shown in Table 1, all universities but one, CTH, have a greater amount of Swedish than English job advertisements.

As seen in Table 1, CTH advertisements date back to 2010. Data from LNU and LTU date back to 2011, while the advertisements from UU and LIU are from 2013 onwards. Job advertisements from KTH are from 2014 onwards, whereas the advertisements from LU and SU are from 2015 onwards. Finally, the advertisements from GU and SLU are more recent, from 2017 onwards.

This collection of job advertisements covers a wide range of data from different institutions that provide education in various fields. CTH, KTH and LTU mainly provide education within STEM (science, technology, engineering, and mathematics) fields of study, but, as opposed to LTU and KTH, CTH also provides education for maritime studies. Additionally, LTU offers education in health science, economics and teaching, amongst others. SU provides education within the areas of human and natural science. Slightly more extensive is the range of education offered by GU, providing programmes and courses in the medical field, human sciences, law and economics, among others. The education provided by LNU and LIU is similar to that of GU, but in contrast to LIU and GU, LNU also provides police education. Furthermore, UU, as well as LU, offer a very extensive range of programmes and courses, including medicine, law and technology. SLU provides, as the name suggests, education within the field of agriculture.

Unlike the other universities, KTH categorised the job advertisements, based on the area of the occupation, into 16 different groups, which can be seen in Table 5.

3.2 Publicly available gendered lexica

The gender assessment of the collection of job advertisements was performed with gendered lexica created by

- Gaucher et al. (2011),
- Lindmark & Sundin (2013) and
- Pietraszkiewicz et al. (2019),

along with the translations we generated from the Gaucher and the Pietraszkiewicz lexica, detailed in Section 4.1.2.2. Both the gendered lexica by Gaucher et al. (2011) and by Pietraszkiewicz et al. (2019) contain words that end with asterisks (e.g. 'courag*').[4] Denoting "the acceptance of all letters, hyphens, or numbers following its appearance" (Gaucher et al., 2011, Appendix), the asterisks allow the extension of the gendered lexica with longer words from truncated words (e.g. from 'courag*' to 'courage' and 'courageous'). Most of these items (e.g. 'compet*') can be equated to a bound base morpheme, i.e., a word that cannot stand free on its own and, as a result, needs a suffix or another base added to it to make it into a word and have a full meaning – but they still capture certain concepts. However, there are some truncated items (e.g. 'assert*') that are free morphemes and, while they can stand alone with a full meaning, they can also be combined with suffixes to create other words and expressions with different meanings (e.g. 'assertive'). Hence, all words that start with the same suffixed-with-an-asterisk root and capture a similar concept to what the root denotes can be detected as agentic or communal terms.

[4]From hereon, words suffixed with an asterisk are referred to as *truncated words*, whereas those without are referred to as *full-form words*.

3.3 Machine translation methods employed for creating gendered lexica

Four different translation methods were employed to generate corresponding Swedish meaning representations for the available English gendered words:

- *Baseline translation*
- *Word-embedding-based translation*
- *Frequency-document-based translation*
- *Frequency-sentence-based translation*

With the advent of Google's Neural Machine Translation (GNMT) system in 2016, Google switched from a statistical to a neural approach in translation. When evaluated on various benchmark test sets for machine translation, the models developed by Google have achieved state-of-the-art results (Johnson et al., 2017; Vaswani et al., 2017). As Google Translate is a state-of-the-art method in machine translation, the *googletrans* library,[5] is employed to perform our baseline translation.

When performing the word-embedding-based translation, English and Swedish pretrained and prealigned embeddings from fastText (Joulin et al., 2018; Bojanowski et al., 2017) were employed. As described in Bojanowski, the embeddings are trained on Wikipedia text using a Skip-Gram model that also takes sub-word information into account, producing embeddings for a total of 294 languages. As the sub-word information allows vector representations of rare words, fastText was chosen over its predecessors, Word2vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014), both of which fail to create representations for words that are not in the original vocabulary. Joulin et al. align several of these monolingual embedding spaces by learning a linear mapping between the source and target languages.

Two types of frequency-based translation were used: document-based and sentence-based. For the document-based translation, we created a *bilingual document corpus* by first aligning the English job advertisements with their Swedish equivalents using a project id.[6] Then, for the sentence-based translation, we aligned the English sentences of each advertisement with their Swedish translation using NLTK's implementation of the Gale & Church (1991) algorithm, creating the *bilingual sentence corpus*. The two approaches are explained in more detail in Section 4.1.2.2.

3.4 Gender Decoder tool

The Gender Decoder is a tool created by Kat Matfield[7] that identifies which gender a job advertisement is associated with. To employ the tool, a string representation of an advertisement and a gendered lexicon are required, whose entries are indicators of agency and communion.

[5] Available at: https://py-googletrans.readthedocs.io/en/latest/
[6] The project id's are part of the metadata of the job advertisements.
[7] Available at: https://github.com/lovedaybrooke/gender-decoder

In the Gender Decoder tool, a variation of the gendered lexicon from Gaucher et al. (2011) is used to assess the advertisements. Matfield extended the lexicon with a few new words (e.g. 'battle', 'champion', 'shar'[8], 'inclusive' etc.) using her linguistic intuition. The tool was implemented to treat all words as truncated (i.e., it disregards Gaucher's distinction between truncated and full-form words). Based on the frequency of gendered words and the ratio of feminine to masculine coded words, a textual document can be classified as belonging to one out of six possible categories: feminine, masculine, strongly feminine, strongly masculine, empty or neutral. An advertisement is considered strongly either feminine or masculine coded if the difference between the number of feminine and masculine coded words is greater than 3. Unless the aforementioned condition holds, a greater number of feminine coded words indicates that an advertisement is feminine coded. On the other hand, more masculine coded words would lead to an advertisement being classified as masculine coded. Finally, an advertisement that does not contain any gendered words is categorised as empty, while an advertisement containing an equal number of masculine and feminine coded words is considered neutral.

3.5 SCB employment statistics

The Swedish Occupational Register from Statistics Sweden (SCB, Statistiska Centralbyrån, 2012) details employment statistics along various different dimensions, including, but not limited to sex, orientation of education and region of birth. Occupations can be categorised according to three different levels – SSYK2, SSYK3 and SSYK4 – starting from a more general categorisation to a more specific one (Statistiska Centralbyrån, 2018). For example, primary school teacher would be classified as the SSYK2 category *23 Yrken med krav på fördjupad högskolekompetens inom utbildning* (Professions with requiremens of advanced university qualifications in education), the SSYK3 category *234 Grundskollärare, fritidspedagoger och förskollärare* (Primary school teachers, recreational pedagogues and preschool teachers) and the SSYK4 category *2341 Grundskollärare* (Primary school teachers). For this project, all three levels of classification, along with the sex dimension, were used.

[8] We assume that 'shar' refers to 'sharing' etc.

4 Implementations

In this section, we describe the data pre-processing and modifications to the tools that were utilised. We start with pre-processing raw texts from our English and Swedish advertisements. As our next step, we describe how the English gendered lexica were translated to Swedish, and how all versions of the lexica were instantiated with different settings. We then present our modifications to the Gender Decoder tool. Finally, we describe how we mapped KTH occupation areas to SCB occupations in order to retrieve statistics for the relevant areas.

4.1 Data preparation

4.1.1 Job advertisements

This section describes the pre-processing of the job advertisements in terms of general data cleaning and lemmatisation.

4.1.1.1 Data cleaning

The advertisements were separated into English and Swedish by using Compact Language Detector v3,[9] a neural network model trained to identify languages. Duplicates, advertisements shorter than 1200 characters and advertisements that are not recognised as either English or Swedish were filtered out, along with announcements of positions that have been filled and internal decisions. Finally, entities that denote personal information (e.g., email addresses, telephone numbers, website URLs) were anonymised.

4.1.1.2 Lemmatisation

The original advertisement text was lemmatised using Stanza (Qi et al., 2020), a Python-based library developed by Stanford that contains various NLP tools. As Stanza provides models for both English and Swedish, these were used for lemmatisation.[10]

The original English gendered lexica, along with all generated Swedish gendered lexica, were also lemmatised.

4.1.2 Gendered lexica

In the following Sections, we outline the processing of the gendered lexica – from the original English gendered lexica, to the different Swedish translations and versions. As illustrated in Figure 1, the process involves multiple steps, which are described below, with the exception of lemmatisation, as that is described in Section 4.1.1.2 above.

[9] CLD3 GitHub repository: https://github.com/google/cld3
[10] Available on Stanza's website: https://stanfordnlp.github.io/stanza/available_models.html

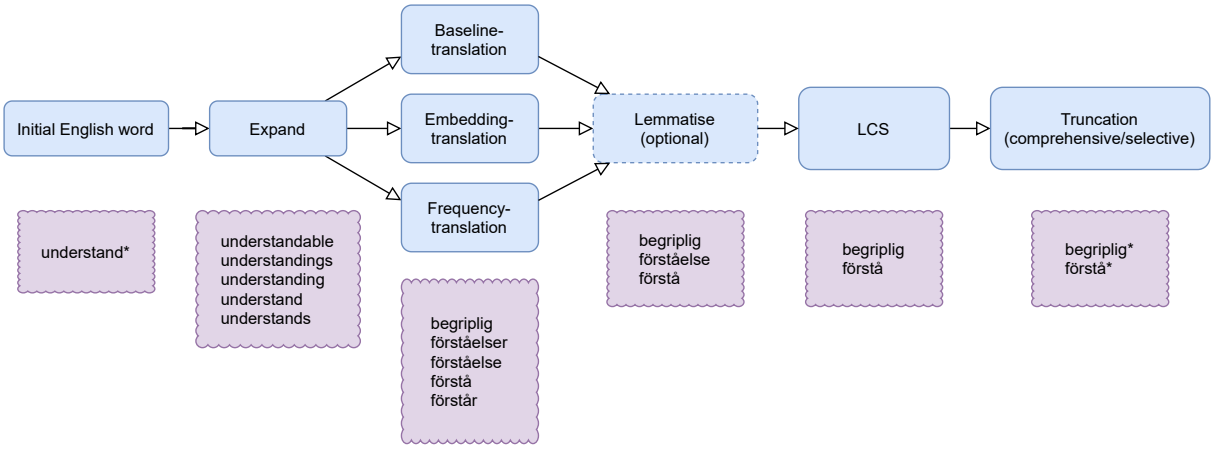

Figure 1: A flowchart illustrating the processing of the words. The blue boxes indicate processes, whereas the purple boxes provide an example of how each process affects the words. That is, the initial English word 'understand*', from the gendered lexicon by Gaucher, is expanded into five words. These five words are then translated (only the baseline method is shown in the chart), after which they are either lemmatised or stay inflected (see Section 4.1.1.2). Thereafter, the list of words is shortened using our implementation of the Longest Common Substring (LCS) (see Section 4.1.2.3). Finally, the gendered lexica are either comprehensively or selectively truncated (see Section 4.1.2.3).

4.1.2.1 Word expansion

As some words in both the Gaucher et al. (2011) and Pietraszkiewicz et al. (2019) gendered lexica are truncated (and suffixed with an asterisk), expanding those into full words was a necessary procedure to create the Swedish gendered lexica. Simply translating truncated words, such as 'courag*' and 'trust*', would either result in meaningless items, or exclude the translation of words such as 'trusting'. For that reason, the truncated words were expanded into full words by scanning the English job advertisements for words that start with the truncated words, allowing the expansion from truncated words to only include relevant words that appear in the advertisements. The words without asterisks (e.g. 'passionate') were not expanded, as Gaucher and Pietraszkiewicz deliberately left these without asterisks.

4.1.2.2 Translations of the English gendered lexica

Baseline translation Building on the gendered lexica by Gaucher et al. (2011) and Pietraszkiewicz et al. (2019), Swedish baseline gendered lexica were created. Using the googletrans library,[11] the expanded gendered lexica were automatically translated, resulting in the Swedish baseline gendered lexica.

Word-embedding-based translation For the embedding-based method, English and Swedish pretrained and pre-aligned embeddings from fastText (Bojanowski et al., 2017; Joulin et al., 2018) were employed, as they generated the most accurate translations. However, for some words, the embeddings did not produce a translation, as there were no representations for these words in the pretrained embeddings. To remedy this, the baseline translation was used as a fallback method.

[11] Available at: https://py-googletrans.readthedocs.io/en/latest/

A multitude of other embeddings, including pretrained embeddings and embeddings trained on the job advertisements, were manually assessed and compared to the fastText embeddings by Bojanowski et al. (2017), and Joulin et al. (2018). The fastText library was used to train embeddings, employing both Skip-Gram and CBOW. With the exception of the embeddings already aligned by Joulin et al., all embeddings were aligned using an algorithm provided by Babylonhealth.[12] Using a different method of embedding alignment than the method used by Joulin et al. allowed us to indirectly compare the two methods, as the embeddings Joulin et al. aligned are available both in their aligned and not-aligned form.[13]

Comparing the translations produced by the embeddings, the pretrained and prealigned embeddings by Joulin et al. produced the most accurate translations, which we attribute to three factors. Firstly, by comparing the translations made by the Bojanowski et al. embeddings aligned by Joulin et al. with the same embeddings aligned by us, using the Babylonhealth algorithm, we deduced that the Joulin et al. alignment method seems to work better, as the embeddings aligned with the method provided by Babylonhealth did not produce accurate translations in as many cases as those aligned by Joulin et al. (2018). Secondly, the Swedish job advertisements used as training data for the Swedish embeddings trained by us seem to contain some English sentences. The embeddings we trained on the job advertisements produced mostly inaccurate translations (for instance, many English gendered words were often translated to an English word, instead of a Swedish word as we would expect). Lastly, the Swedish and English data used to train the embeddings are not equal in size – the vocabulary of the Swedish advertisements is approximately four times the size of the English vocabulary, meaning that the alignment of the two vector spaces is likely to generate pairs of vectors that do not necessarily capture the same semantic relations.

In order to translate a word in the English embedding space to a word in the Swedish embedding space, the embeddings of the Swedish subspace were normalised.[14] Then, the cosine similarity of all the normalised Swedish embeddings and the normalised embedding of the English gendered word was calculated. The Swedish word representation most similar (i.e., with the highest cosine similarity) to the English word representation was chosen as the Swedish translation.

Frequency-based translation An English to Swedish translation was also generated using frequency-based methods. Two types of frequency-based translations were explored: sentence-based and document-based. These two types of frequency-based translation are detailed below.

Sentence-based translation English and Swedish term-sentence matrices were created from the bilingual sentence corpus, which is described in Section 3.3. In the English term-sentence matrix (shown in the upper part of Table 2), each row represents a gendered word from the gendered lexicon and each column represents each separate sentence from the English advertisements. Each row of the Swedish term-sentence matrix (shown in the lower part of Table 2) represents each unique word that occurs in the Swedish advertisements,

[12] Available at: https://github.com/babylonhealth/fastText_multilingual

[13] The aligned embeddings are available at https://fasttext.cc/docs/en/aligned-vectors.html and the not-aligned are available at https://fasttext.cc/docs/en/pretrained-vectors.html

[14] The normalisation was done using a script which can be found at https://github.com/babylonhealth/fastText_multilingual/blob/master/align_your_own.ipynb.

	Sentence #1	Sentence #2	Sentence #3	Sentence #4
English				
analysis	1	1	1	0
cooperation	0	0	0	0
together	1	1	1	1
connection	1	0	1	1
Swedish				
analys	1	1	1	0
samarbete	0	1	0	0
tillsammans	1	1	1	1
anslutande	0	0	0	0

Table 2: An example of a term-sentence matrix for four English and four Swedish words in four English and four Swedish sentences. Each cell contains the number of times the (row) word occurs in the (column) sentence.

while the columns represent each separate sentence of a Swedish advertisement that is the translation of an English sentence. An example of eight frequency vectors representing eight words from our data is shown in Table 2.

The English vectors that occurred in the data were linked to Swedish vectors based on their cosine similarity – a frequency vector representing an English gendered word was linked to the most similar vector created from the Swedish sentences. As the gendered lexicon created by Pietraszkiewicz et al. (2019) is much larger than the lexicon developed by Gaucher et al. (2011), it proved impossible to create sentence-based frequency vectors for the entirety of the Pietraszkiewicz et al. gendered lexicon due to computational limitations. Thus, the frequency-sentence based method was only applied to the lexicon developed by Gaucher.

Document-based translation English and Swedish term-document matrices were created from the bilingual document corpus, described in Section 3.3. In the English term-document matrix (shown in the upper part of Table 3), each row represents a word from the gendered lexicon and each column represents an English advertisement. As demonstrated in Table 3, each row of the Swedish term-document matrix (shown in the lower part of Table 2) represents each word that occurs in the Swedish advertisements, while the columns represent each Swedish advertisement that is the translation of an English advertisement. A sample of eight frequency vectors representing eight words from our data is shown in Table 3.

The vectors of the English gendered words were linked to the most similar vector representing a Swedish word, using cosine similarity. Because the document-based vectors were of significantly fewer dimensions in comparison with the sentence-based vectors, we were able to generate document-based translations for both the Gaucher and the Pietraszkiewicz-based gendered lexica.

Eventually, we used the original dimensions of the frequency vectors to translate words from English to Swedish. However, it should be noted that in an attempt to combat the sparseness of the frequency vectors, we performed dimensionality reduction (to 200, 100, 50 & 20 dimensions) using Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) on the English and the Swedish vectors. We then calculated the cosine similarity of each English gendered word vector with all Swedish word vectors, but

	Advertisement #1	Advertisement #2	Advertisement #3	Advertisement #4
English				
analysis	5	3	1	1
cooperation	1	0	3	2
together	2	1	1	1
connection	1	1	0	1
Swedish				
analys	3	2	1	1
samarbete	4	0	4	5
tillsammans	1	2	1	0
anslutande	1	0	0	0

Table 3: An example of a term-document matrix for four English and four Swedish words in four English and the corresponding four Swedish advertisements. Each cell contains the number of times the (row) word occurs in the (column) advertisement.

the most similar Swedish word vector was an inaccurate translation, in the majority of the cases. As a result, we decided to not perform dimensionality reduction on the data since it apparently caused a substantial information loss.

4.1.2.3 Instantiating different versions of the gendered lexica

Each gendered lexicon resulting from the four translation methods (baseline, frequency-document-based, frequency-sentence-based and embedding-based), along with each original English gendered lexicon, was instantiated four times with different settings. The settings can be seen as dimensions: the first dimension is *lemmatised* (L) or *inflected* (I), and the second is *comprehensively truncated* (CT) or *selectively truncated* (ST). When combined, these settings produce four different versions of the gendered lexicon, per translation, resulting in a total of 28 different Swedish gendered lexica and 8 different English gendered lexica, as shown in Table 4.

The comprehensively truncated/selectively truncated dimension was created to allow the indication of certain words as extendable (those suffixed with an asterisk). In the selectively truncated versions of the lexica, the Swedish translations of truncated words in the original papers are suffixed with an asterisk. As for the comprehensively truncated versions, asterisks are added to all words, meaning that all words are extendable when employed with the truncation-sensitive Gender Decoder tool described in Section 3.4.

As mentioned in Section 4.1.1.2, the lemmatisation was done using Stanza. The words in the the lemmatised versions of the gendered lexica were altered to their base form, while the inflected gendered lexica consist of translated words that have not been processed further.

To limit the size of the gendered lexica, we implemented a partial solution to the longest common substring problem.[15] Our implementation finds the longest common *string-initial* substring, *if* the length of that string

[15]The longest common substring problem is the task of finding the longest string that two (or more) strings share. For example, in the string pair *zodiac* and *diagram*, the longest common substring is *dia*.

Gaucher-based			
Gendered lexicon	**Version**	**Count**	
		F	**M**
English	LCT	40	42
	ICT		
	LST		
	IST		
Baseline	BLCT	67	56
	BICT	71	56
	BLST	67	56
	BIST	71	56
Embedding-based	ELCT	83	73
	EICT	84	74
	ELST	83	73
	EIST	84	74
Frequency-document-based	FdLCT	88	74
	FdICT	88	75
	FdLST	88	74
	FdIST	88	75
Frequency-sentence-based	FsLCT	82	74
	FsICT	82	75
	FsLST	82	74
	FsIST	82	75

(a) The Gaucher-based gendered lexica.

Pietraszkiewicz-based			
Gendered lexicon	**Version**	**Count**	
		F	**M**
English	LCT	186	194
	ICT		
	LST		
	IST		
Baseline	BLCT	257	277
	BICT	270	285
	BLST	256	277
	BIST	270	285
Embedding-based	ELCT	291	306
	EICT	297	311
	ELST	290	306
	EIST	297	311
Frequency-document-based	FdLCT	411	385
	FdICT		
	FdLST		
	FdIST		

(b) The Pietraszkiewicz-based gendered lexica.

Table 4: Overview of the different versions of the gendered lexica. For each lexicon version in the tables, there is a feminine and a masculine word count. Table a shows the Gaucher-based gendered lexicon, while Table b shows the Pietraszkiewicz-based gendered lexicon. The version codes indicate which translation was used, whether the words in the lexica are selectively (ST) or comprehensively truncated, and whether the words are lemmatised (L) or inflected (I). For example, FdICT indicates that the gendered lexicon is translated using the frequency-document-based method (Fd), that the words are inflected (I) and comprehensively truncated (CT). In the two tables, counts of the masculine and the feminine words in each gendered lexicon are also given.

is greater than or equal to three. This task, finding the longest common string-initial substring, is performed for each word group translated from the same truncated word in the original English gendered lexicon. An example of this is shown in Figure 1.

The comprehensively truncated and the selectively truncated versions of the gendered lexica (both inflected and lemmatised) were run with the truncation-sensitive version of the Gender Decoder, described in Section 4.2.

4.2 Truncation-sensitive Gender Decoder tool

Matfield's Gender Decoder tool treats all gendered words as truncated, as opposed to what Gaucher et al. (2011) suggests. As a result, all items in a text that start with a word from the gendered lexicon count as gendered words. For instance, Gaucher et al. would not consider 'logically' a gendered word since 'logic' is not one of the truncated words. However, when using the Matfield Gender Decoder tool, 'logically' would count as a masculine coded word. A distinction between truncated and full-form words is even more important in the assessment of data with the Pietraszkiewicz et al. (2019) gendered lexicon. Their lexicon contains some short words (e.g. 'we', 'you', etc.) that would inflate the results of the distribution of gender coded advertisements, since the Gender Decoder tool counts all words that start with 'we' and 'you' as gendered.

To take into consideration the algorithm and the form of the original gendered lexica by both Gaucher et al. and Pietraszkiewicz et al. and prevent such misinterpretations of the data, we implemented a modified version of the Gender Decoder tool (from hereon, this version of the Gender Decoder tool will be referred to as *truncation-sensitive Gender Decoder* and will be used for the assessments of the advertisements). In our version, a distinction was made between truncated and full-form words. The former category of words was matched with all tokens in the advertisements that start with them, while the latter category requires exact matches between the words it contains and words that appear in the advertisements (that is, in this version, 'logically' would not count as masculine coded). The classification of an advertisement as either neutral, feminine or masculine coded is based on gendered word frequency. For our experiments, following the lines of Gaucher et al., no distinction was made between masculine coded and strongly masculine coded, and feminine coded and strongly feminine coded, i.e. there are only three categories: feminine, masculine and neutral/empty.

4.3 KTH occupations to SCB pairing

As mentioned in Section 3.1, the KTH advertisements are labeled with one out of 16 occupation areas. For example, an advertisement for *Systemadministratör* is categorised into the group *Data/IT*. These categories were matched to the SCB SSYK-categories (Statistiska Centralbyrån, 2018) using a one-to-many string matching algorithm.[16]

To ensure a good pairing between KTH and SCB, the advertisements were reviewed, after which some of the area labels were changed: *hantverksyrken* (trade work) was changed to *hantverkare* (tradespeople); *naturvetenskapligt arbete* (scientific work) was changed to *naturvetenskap* (science); *tekniskt arbete* (technical work) was changed to *teknik* (technology); *pedagogiskt arbete* (educational work) was changed to *universitets- och högskolelärare* (university and college teachers); *industriell tillverkning* (industrial manufacturing) was changed to *industri tillverkning* (industry manufacturing); *sanering och renhållning* (sanitation and cleaning)

[16]The SSYK occupation classification system contains different levels, each level containing multiple categories of occupations. The lower level categories are more general, while the higher level categories are more specific. See https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/ for a detailed explanation of the SSYK-categories and levels.

Occupation Area	Women %	Men %	Gender	# KTH advertisements
Construction and building	13%	87%	male	39
Installation, operation, maintenance	21%	79%	male	35
Industrial manufacturing	22%	78%	male	9
Data/IT	23%	77%	male	322
Security work	30%	70%	male	10
Technical work	34%	66%	male	656
Total # for male dominated areas	**29%**	**71%**		**1071**
Trade work	42%	58%	equal	1
Scientific work	44%	56%	equal	749
Managers and organisation directors	44%	56%	equal	51
Educational work	47%	53%	equal	2174
Sales, purchasing, marketing	49%	51%	equal	43
Art, culture, media, design	55%	45%	equal	72
Administration, finances, law	56%	44%	equal	1048
Sanitation and cleaning	57%	43%	equal	15
Hotel, restaurant, institutional household	58%	42%	equal	2
Total # for equal areas	**49%**	**41%**		**4155**
Health care	73%	27%	female	8
Total # for female dominated areas	**73%**	**27%**		**8**

Table 5: Occupational statistics for the KTH areas in English. The original Swedish version of this table is available in the Appendix, Section 9.2.

was changed to *städ* (clean). Subsequently, the pairings were inspected and extended with SSYK-categories, as deemed necessary by comparing the KTH advertisements and the descriptions of SCB SSYK-categories.

For all SCB occupations that were matched to each KTH occupation, we first calculated the proportions of women to men employed in all occupations that were retrieved using the string matching algorithm. With the aim of examining the dominant sex in each KTH area, the average of these proportions was calculated. Based on the sex of the qualified majority (60% and above), an area was coded as female or male dominated. An area was coded as neutral if there is no dominant sex (between 40 to 60%). A table displaying the occupational statistics of the areas advertised from KTH can be seen in Table 5.

5 Experiments

In this section, we describe the experiments we conducted to evaluate the Swedish gendered lexica that were created using the four translation methods, and investigate which lexicon captures agency and communion most accurately. We present the quantitative, followed by the qualitative experiments.

5.1 Quantitative evaluation experiments

We conducted a series of quantitative experiments to find the version of the Swedish gendered lexicon that captures agency and communion most similarly to the corresponding English lexicon. In the absence of a gold standard (for instance, previous psycho- or sociolinguistic research on gendered wording in Swedish, or advertisements that have already been assigned a gender by annotators) that could be used to evaluate the performance of our gendered lexica, our evaluation methods are inevitably limited. In the lack of a ground truth, we assume that the English and the Swedish gendered lexica will perform similarly during the assessment of advertisements, since both languages are Germanic and share some structural similarities. Thus, we assume that a Swedish gendered lexicon yields the most accurate results if its results are the most similar to the results yielded by the original English lexicon. We decided to use all available advertisements for the evaluation of the gendered lexica, since the only translation methods that handled the advertisements during development, the frequency-sentence-based and the frequency-document-based method, only utilised a very small portion (6%).

The truncation-sensitive Gender Decoder tool was employed to assess the job advertisements. The experiments were run with all available versions of the gendered lexica, which can be seen in Table 4. We compared the assessments on the English advertisements yielded by the original English Gaucher lexicon with the assessments on the Swedish advertisements produced by the different Gaucher-based Swedish gendered lexica, and correspondingly for the Pietraszkiewicz gendered lexica.

5.1.1 Experiment 1 – crosslingual comparison of distributions of feminine and masculine coded advertisements

For the first evaluation experiment, we ran the truncation-sensitive Gender Decoder tool on the English and Swedish job advertisements using all available gendered lexica (Gaucher- and Pietraszkiewicz-based) and compared the difference in distribution of the masculine to feminine coded advertisements between the English and the Swedish job advertisements. We assumed that the most accurate Swedish gendered lexicon would produce the most similar scores to the scores of the English lexicon.

5.1.2 Experiment 2 – crosslingual comparison of frequency distributions of gendered wording scores

For the second evaluation experiment, we gave each advertisement a gendered wording score between -1 and +1, where a score of -1 is given to an advertisement with exclusively masculine coded words, a score of 0 is given to an advertisement with either no gendered words or with an equal number of masculine and feminine coded words, while a score of +1 is given to an advertisement with only feminine coded words.

The general formula for calculating the gendered wording score of an advertisement is shown in Eq. 1.

$$s(f, m) = \frac{(f - m)}{(f + m)} \tag{1}$$

In this formula, s is the gendered wording score of an advertisement, while f and m represent the counts of feminine and masculine coded words in that advertisement.

We calculated gendered wording scores in the English advertisements using the English gendered lexica by Gaucher et al. (2011) and Pietraszkiewicz et al. (2019). After calculating the same scores for the Swedish data, using all available Swedish gendered lexica, we compared the distributions of the scores of the English gendered lexicon with the scores of the different Swedish lexica. We assumed that the most similar Swedish distribution to the English distribution is the distribution of the most accurate gendered lexicon. It should be noted that the terms negative and positive gendered wording scores are used interchangeably with the terms masculine and feminine gendered wording scores respectively.

5.1.3 Experiment 3 – crosslingual comparison of the frequency distributions of gendered word counts

Our third evaluation experiment was conducted to examine if the findings of the second experiment were consistent with the results from this experiment. We calculated the frequency of the English gendered words provided by Gaucher et al. (2011) and Pietraszkiewicz et al. (2019) in the English advertisements and the frequency of Swedish gendered words in the corresponding Swedish advertisements using the different versions of the Swedish gendered lexica. The most accurate gendered lexicon was chosen by comparing the frequency distributions of gendered word counts using the original English lexicon in the English job advertisements with the corresponding frequency distributions produced by the different versions of the Swedish lexica; we deemed the Swedish lexicon with the most similar frequency distribution to the English frequency distribution to be the most accurate.

5.1.4 Experiment 4 – crosslingual comparison of the ten most frequent gendered words

In the fourth and final quantitative evaluation experiment, we compared the ten most frequent gendered words in the English data with the ten most frequent gendered words, according to the different versions of the Swedish gendered lexica, from the Swedish data.

5.2 Qualitative evaluation experiments

For the qualitative analysis, one Swedish native speaker evaluated the Swedish gendered lexica that are based on Pietraszkiewicz et al. (2019) and Gaucher et al. (2011). Because of the size of all combined gendered lexica, [17] three other Swedish native speakers were asked to evaluate a small subset of words (50 feminine

[17] The Gaucher and the Pietraszkiewicz lexica contain a total of 1363 English words, which had either three or four translations each.

and 50 masculine coded) from the lexica based on Gaucher.[18] Each word of the baseline, the frequency-based and the word-embedding-based translations was evaluated and scored on a grading scale of -1, 0, +1, where -1 indicates a bad translation, 0 indicates an acceptable translation and +1 indicates a good translation. The gendered lexicon assessed with the highest score was assumed to be the best performing quality-wise.

Inter-annotator agreement was also employed to measure the similarity between the four annotators' decisions regarding the correctness of the translations. To measure the inter-annotator agreement, we used Cohen's kappa coefficient (κ) (Cohen, 1960), a very popular and simple-to-use statistic for qualitative analysis. It "measures the proportion of observed agreement over the agreement by chance and the maximum agreement attainable over chance agreement considering pairwise agreement" (Bhowmick et al., 2008, p. 1).

5.3 Sample study: are gendered wording and field of work correlated

Following the studies conducted by Gaucher et al. (2011) and Pietraszkiewicz et al. (2019), we used the best performing Swedish gendered lexicon on the Swedish KTH advertisements[19] to examine if there is a correlation between the dominant gender of the field of work an advertisement is intended for and its gendered wording score (introduced in the second quantitative experiment, Section 5.1).

Pietraszkiewicz's study confirmed the hypothesis that the advertisements for male dominated areas contain more masculine coded words while the advertisements for female dominated areas contain more feminine coded words. Gaucher, on the other hand, found that in advertisements for both male and female dominated areas, masculine coded words are more prevalent. However, we did not expect to get the same results while testing the gendered lexicon on our data for various reasons. First and foremost, all advertisements labeled with the area they are destined for came from the same employer, KTH University, which limits the possibility of variation in language use. On top of that, out of the 16 areas of employment, only one (health care) is female dominated (73% women), six are male dominated ($24 \pm 7\%$ women), while nine have an even number of female and male workers ($50 \pm 6\%$), which does not adequately represent the Swedish society. In addition, because the sample of advertisements for female dominated areas is not significant (8 advertisements), we investigated the gendered wording scores in the advertisements for neutral and male dominated areas. In total, the number of advertisements used in this experiment is 5226, of which 1071 are from male dominated areas and 4155 are from neutral areas (see Table 5 for more details). While this is a significant number of advertisements, the findings of this study are to be taken with a grain of salt, given that no advertisements for female dominated areas were tested.

5.4 Assessment of advertisements per university

After the quantitative and qualitative evaluation, the best performing versions of both the English and the Swedish gendered lexica were used to assess the English and Swedish advertisements of each university separately.

[18] The instructions given to the study participants can be found in Appendix, Section 9.1.
[19] As there is only one English advertisement from KTH in the data, it could not be used.

English advertisements from KTH and UU are not analysed in this study as there was insufficient data for these two universities. The Swedish advertisements from KTH and UU, however, are still examined.

For the sake of this study, we assumed that the advertisements from the universities that offer education within the STEM fields (CTH, LTU) are for occupations within male dominated areas, while the advertisements from the remaining universities were assumed to be for occupations within neutral areas (that is, neither male, nor female dominated areas).

The end goal was to investigate if there is a correlation between the education provided by each university and the portion of masculine and feminine coded advertisements. Furthermore, we aimed to examine how the distribution of feminine and masculine coded advertisements changes from the English to the Swedish advertisements, when employing the gendered lexica we generated.

6 Results

This section presents the results from the quantitative and qualitative evaluation experiments. At this point, it should be noted that the overlap between the Gaucher and Pietraszkiewicz gendered lexica tends to be rather insignificant (with a maximum of 40 common words between each Gaucher-based and Pietraszkiewicz-based lexicon with the same dimensions). It is apparent that the two gendered lexica differ greatly, and, thus, we expect that their results will also differ to a great extent.

6.1 Quantitative evaluation experiments results

6.1.1 First experiment: comparing distributions

As stated in Section 5.1, the first experiment of our quantitative analysis compared the distribution of feminine and masculine coded English advertisements with the corresponding distributions of the Swedish data.

As Table 6 shows, the Swedish gendered lexicon created with the embedding-based translation method scored the most similar results on the Swedish advertisements to the results returned by all versions of the English Gaucher gendered lexicon on the English advertisements. It is apparent that the percentages of the English feminine coded advertisements are consistently higher than the corresponding Swedish percentages. Conversely, the percentages of the English masculine coded advertisements are significantly lower than the Swedish ones.

Gendered lexicon	**IST**		**ICT**		**LST**		**LCT**	
	F	M	F	M	F	M	F	M
English	52%	26%	58%	20%	53%	25%	59%	21%
Baseline	41%	42%	40%	43%	18%	72%	23%	67%
Embedding	**43%**	**41%**	**51%**	**34%**	**46%**	**39%**	**55%**	**32%**
Frequency-document	25%	63%	31%	56%	27%	61%	33%	55%
Frequency-sentence	28%	62%	34%	55%	23%	69%	28%	62%

Table 6: Percentage distributions of feminine (F) to masculine (M) coded advertisements yielded by the first quantitative experiment using the Gaucher-based gendered lexica. The percentages of equal advertisements are not shown in the Table, but can be calculated by subtracting the female and male percentages from the total (100%). The English results are shown in the top row, and the Swedish percentages closest to the English are boldfaced.

The percentages of the feminine and masculine coded advertisements scored by the Pietraszkiewicz gendered lexica displayed more variation, which is demonstrated in Table 7. Firstly, it is worth noting that the Swedish advertisements assessed with ST versions of the gendered lexicon generally appear to have a higher ratio of feminine coded advertisements, compared to the English advertisements. The same cannot be said for the Swedish advertisements assessed with CT versions of the gendered lexicon, which seem to lean in the opposite direction, having a higher ratio of masculine coded advertisements, compared to the English advertisements.

Gendered lexicon	IST		ICT		LST		LCT	
	F	M	F	M	F	M	F	M
English	36%	47%	74%	5%	36%	47%	80%	8%
Baseline	60%	35%	11%	86%	72%	23%	**52%**	44%
Embedding	**57%**	**37%**	14%	83%	**59%**	**35%**	47%	49%
Frequency-document	78%	17%	**80%**	**0.18%**	84%	13%	30%	**0.01%**

Table 7: Percentage distributions of feminine (F) to masculine (M) coded advertisements yielded by the first quantitative experiment using the Pietraszkiewicz-based gendered lexica. The percentages of equal advertisements are not shown in the Table, but can be calculated by subtracting the female and male percentages from the total (100%). The English results are shown in the top row, and the Swedish percentages closest to the English are boldfaced.

The EIST and ELST gendered lexica produced the most similar results to the corresponding versions of the English gendered lexicon. When comparing the results of the English and Swedish ICT versions of the Pietraszkiewicz gendered lexicon, the most similar performance was achieved by the FdICT gendered lexicon. On the other hand, the LCT version of the Pietraszkiewicz gendered lexicon produced conflicting results, in the sense that the feminine coded part of the Swedish baseline lexicon produced the most similar results to the feminine coded part of the original English lexicon, while the masculine coded part of the frequency-document-based lexicon yielded the closest results to the masculine coded part of the original English gendered lexicon. However, it should be noted that the results of the Swedish frequency-document-based gendered lexicon are quite different, compared to the results of the other Swedish gendered lexica, which calls for some skepticism.

We also examined the discrepancies of the results yielded by the inflected against the lemmatised version of each gendered lexicon, and correspondingly for the CT against the ST versions of each lexicon. The goal was to investigate to which extent lemmatising and treating all words in the gendered lexica as truncated affects the results of our experiments.

As Figure 2 shows, the performance of most of the Gaucher-based gendered lexica tends to stay the same before and after lemmatisation. The only lexicon where there is a notable discrepancy between the inflected and the lemmatised version is the Swedish baseline, both the comprehensively and selectively truncated versions.

Figure 3 illustrates the discrepancies between the results of the lemmatised and the inflected versions of the Pietraszkiewicz-based gendered lexica. The results of comprehensively truncated gendered lexica show a larger variation between their inflected and lemmatised version than the selectively truncated gendered lexica. Specifically, the BICT and EICT lexica resulted in a notably lower amount of masculine coded advertisements, compared to results of the lemmatised versions. Table 7 demonstrates that the FdICT gendered lexicon, on the other hand, resulted in the majority of the advertisements being masculine coded, while the lemmatised version resulted in the majority being feminine coded.

The distribution of the advertisements assessed with the ST versions of the Pietraszkiewicz-based gendered lexicon is fairly even, both prior to and after lemmatisation. Similarly to the selectively truncated Gaucher-

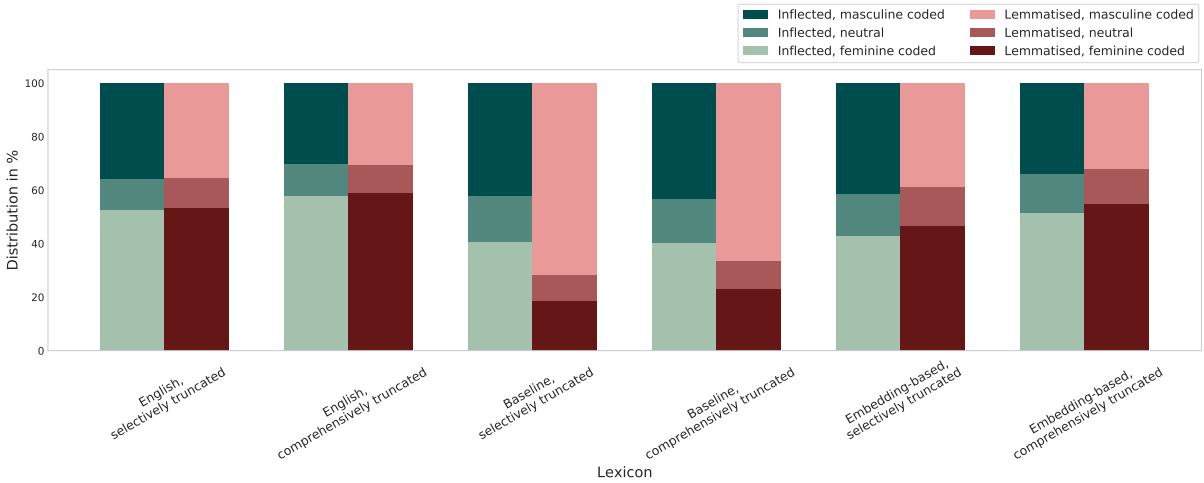

Figure 2: Difference in percentage distribution between lemmatised and inflected feminine coded, neutral and masculine coded advertisements assessed with the Gaucher gendered lexica. As most of the lexica displayed similar results, only a sample of the most interesting results is shown here.

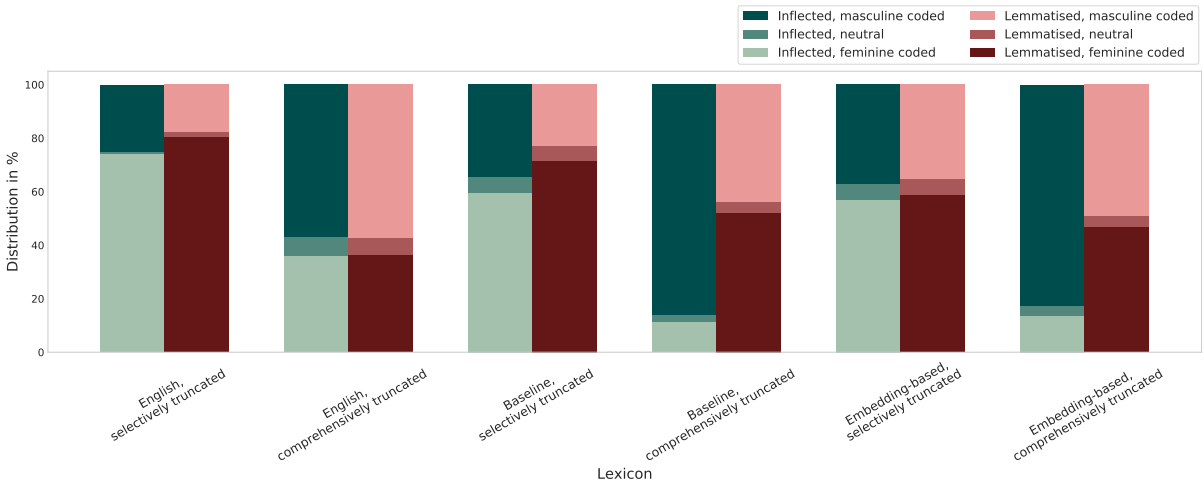

Figure 3: Difference in percentage distribution between lemmatised and inflected feminine coded, neutral and masculine coded advertisements assessed with the Pietraszkiewicz gendered lexica. As most of the lexica displayed similar results, only a sample of the most interesting results is shown here.

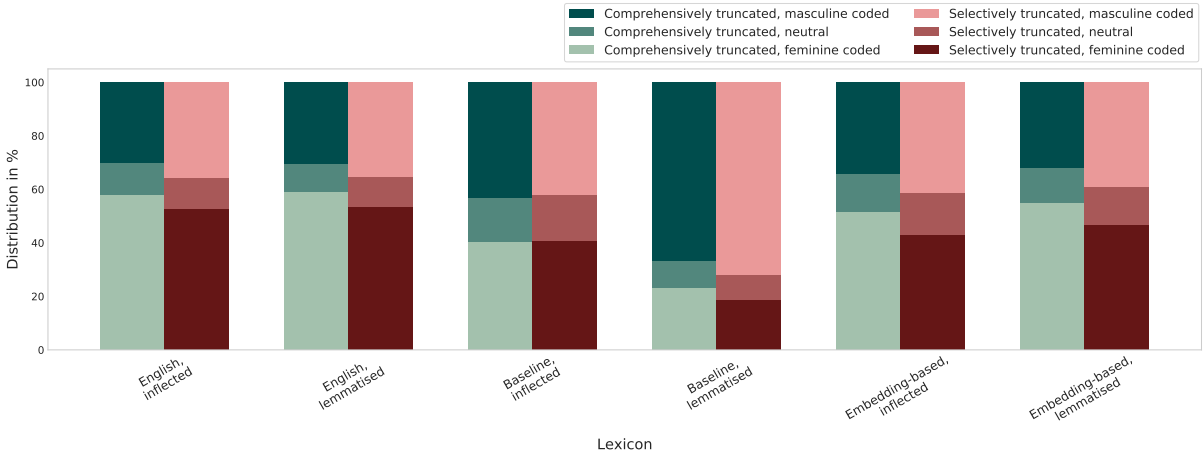

Figure 4: Difference in percentage distribution between feminine coded, neutral and masculine coded advertisements assessed with the Gaucher gendered lexica. Selectively truncated and comprehensively truncated versions of the same gendered lexicon are compared here. As most of the lexica displayed similar results, only a sample of the most interesting results is shown.

based gendered lexica, the largest discrepancy between inflected and lemmatised is found in the baseline versions, although much smaller and in the opposite direction; in this case, the lemmatisation resulted in a larger amount of feminine coded advertisements.

As shown in Figure 4, employing the ST over the CT versions of the Gaucher-based gendered lexicon to assess the advertisements only alters the distribution of the gender coded advertisements marginally. The Pietraszkiewicz-based CT and ST gendered lexica, on the other hand, resulted in larger discrepancies in the distribution of gender coded advertisements, as shown in Figure 5. Moreover, the ST versions of the Pietraszkiewicz-based gendered lexicon seem to yield a larger portion of neutral advertisements, compared to the CT versions of the gendered lexicon.

Tables 6 and 7 show that the ICT and LCT versions of the gendered lexicon produced the most extreme results for English and Swedish. Given the discrepancies in the results between the English results and the Swedish results, we decided to not use the CT versions of the gendered lexicon in the remaining quantitative experiments. Furthermore, given that the results of the IST and LST versions of the gendered lexicon were quite similar, only the latter was utilised in the other experiments.

6.1.2 Second experiment: comparing gendered wording scores

As mentioned in Section 5.1, the second quantitative evaluation experiment examined the differences in frequency distributions of gendered wording scores produced by the original English gendered lexicon, and the different Swedish gendered lexica, in the respective job advertisements.

As Figure 6 illustrates, the distribution of gendered wording scores yielded by the Gaucher-based embedding-based gendered lexicon is the most similar to the distribution of the original Gaucher-based lexicon. Regarding the embedding-based gendered lexicon and the English gendered lexicon, the amount of advertisements with a score of 0.00 or lower is around 57% and 50%, respectively. For the other gendered lexica,

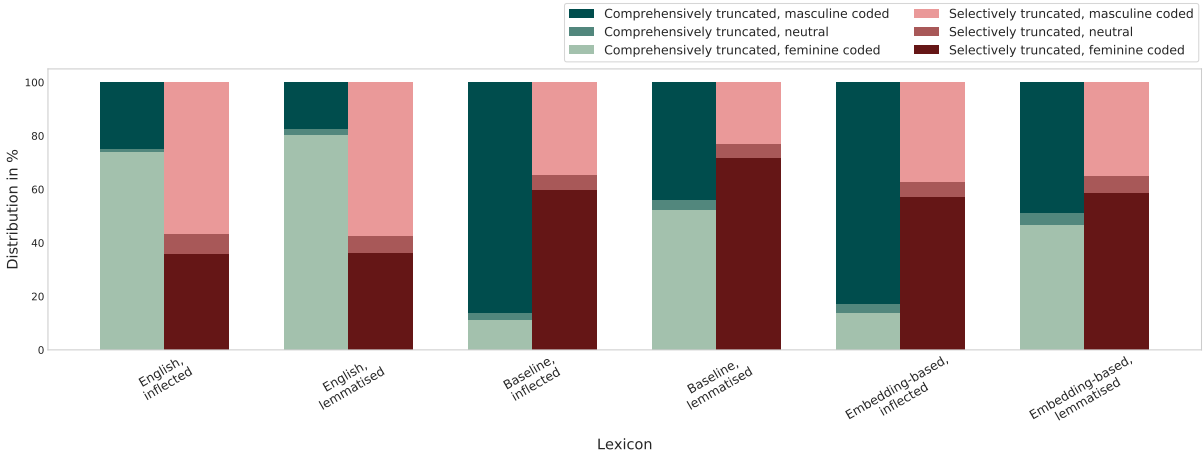

Figure 5: Difference in percentage distribution between feminine coded, neutral and masculine coded advertisements assessed with the Pietraszkiewicz gendered lexica. Selectively truncated and comprehensively truncated versions of the same gendered lexicon are compared here. As most of the lexica displayed similar results, only a sample of the most interesting results is shown.

the amount of advertisements with a score of 0.00 or lower is much higher, ranging from around 77%, using the frequency-sentence-based gendered lexicon, to almost 90%, using the lexicon by Lindmark & Sundin.

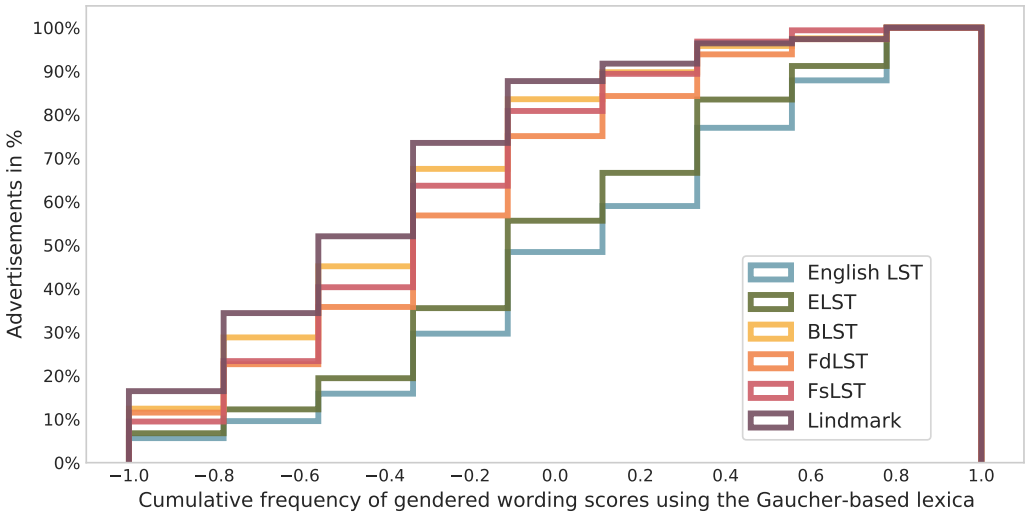

Figure 6: Cumulative frequency distribution of gendered wording scores yielded by the English and the Swedish Gaucher-based gendered lexica.

Similarly to the Gaucher-based distributions, as shown in Figure 7, the evaluation of the Pietraszkiewicz-based distributions shows that the embedding-based distribution is most similar to the distribution of the English gendered lexicon. In this case, as opposed to the Gaucher-case, all of the Swedish gendered lexica show a lower amount of advertisements with a negative score than the English lexicon. That is, the amount of advertisements with a score of 0.00 or lower, for the English lexicon, is around 70%, whereas for the Swedish gendered lexica, the amount of advertisements with a score of 0.00 or lower ranges from just over 20% to around 55%. This is particularly noticeable around the -0.25-+0.25 mark.

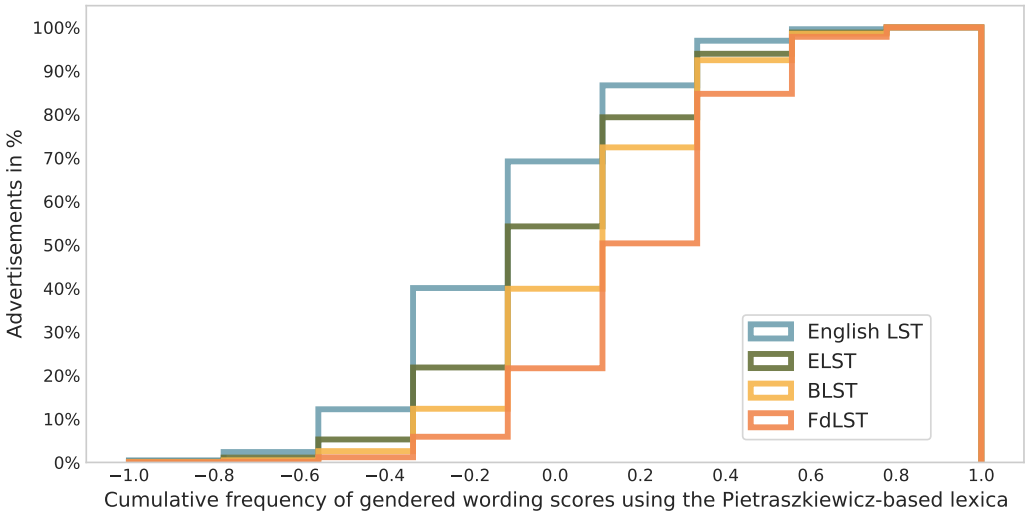

Figure 7: Cumulative frequency distribution of gendered wording scores yielded by the English and the Swedish Pietraszkiewicz-based gendered lexica.

Based on these results, it is apparent that the embedding-based versions of both the Gaucher and Pietraszkiewicz gendered lexica perform the best in the second experiment, as their distributions of gendered wording scores across the Swedish advertisements are most similar to the distributions of the English Gaucher and Pietraszkiewicz lexica.

6.1.3 Third experiment: comparing frequency of gendered words

As described in Section 5.1, the third evaluation experiment compared the frequency distributions of gendered word counts, as assessed by the English gendered lexicon, with the corresponding frequency distributions produced by the different Swedish lexica. By conducting this experiment, we aim to find out whether there is a consistency in the results of the second and the third experiment.

Figure 8 confirms that the frequency distribution of gendered word counts yielded by the embedding-based lexicon is most similar to the distribution of the original English Gaucher-based lexicon. However, as seen in the figure, the line of the embedding-based gendered lexicon is generally above the English line, indicating that a slightly higher percentage of advertisements assessed by the embedding-based gendered lexicon have the same number of gendered words as the advertisements assessed using the English lexicon. The gendered lexicon that generated the second closest distribution of gendered words to the distribution of the Gaucher-based English gendered lexicon is the frequency-document-based lexicon. The frequency-sentence-based gendered lexicon, however, seems to have generated the most dissimilar distribution.

In Figure 8, we can see that approximately 90% of the English advertisements assessed with the original Gaucher gendered lexicon and the Swedish advertisements assessed with the embedding-based gendered lexicon contain around 12 gendered words or less, whereas for the Swedish advertisements assessed with the other Swedish lexica, approximately 80% contain the same number of gendered words. It is apparent that the distributions generated by the different gendered lexica are most dissimilar around the 5-word-mark.

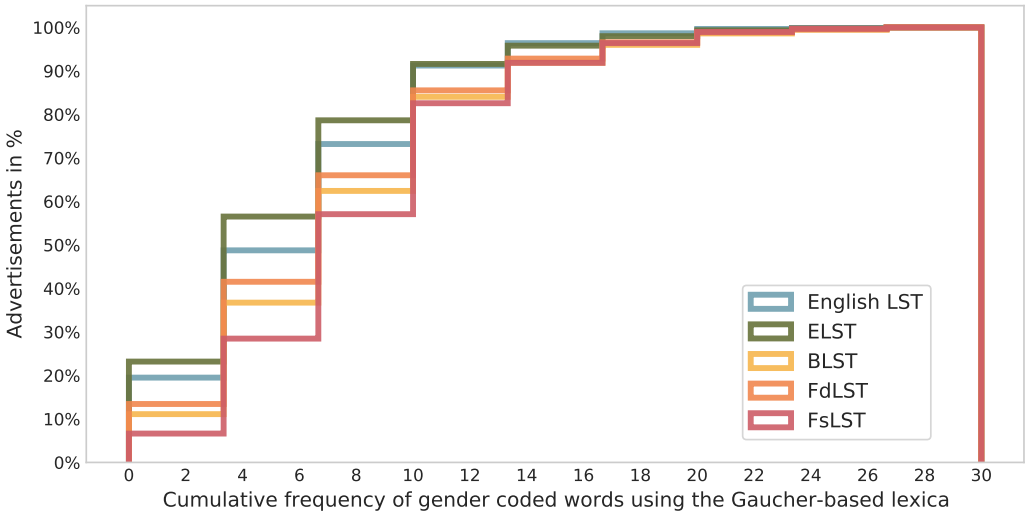

Figure 8: Cumulative frequency distribution of gendered word counts yielded by the English and the Swedish Gaucher-based gendered lexica.

Regarding the Pietraszkiewicz-based gendered lexica, the frequency distribution of gendered word counts yielded by the baseline lexicon seems to be most similar to that of the original English Pietraszkiewicz-based lexicon up to approximately the 25-word-mark (see Figure 9). However, the distribution generated by the embedding-based lexicon appears to be closest to the English distributions for more than 25 gendered words. The most dissimilar distribution to the English one was generated by the frequency-document-based lexicon, indicating that a frequency-based translation is often not the most successful. This is in conflict with what was noted regarding the distributions of the Gaucher-based gendered lexica, where the frequency-document-based lexicon appeared to be the second best choice. By examining Figure 9 further, we can note that around 80% of the English advertisements and the Swedish advertisements, assessed with the embedding-based gendered lexicon, have the same number of gendered words (~30).

6.1.4 Fourth experiment: comparing top k gendered words

As described in Section 5.1, the fourth quantitative evaluation experiment compared the ten most frequent English and Swedish gendered words in the English and Swedish job advertisements respectively.

In Table 8, displaying the ten most frequent words in the job advertisements using each of the different Gaucher-based gendered lexica, we can see that the majority of the most frequent English words can be found (translated) among the most frequent Swedish words. For example, 'individual' is the sixth most frequent word in the English advertisements, and the Swedish word, *individuell*, is the seventh most frequent word in the Swedish advertisements using the baseline and the frequency-document-based gendered lexica, and the sixth most frequent word using the embedding-based lexicon. The English words not present among the most frequent Swedish words, according to either gendered lexica, are 'support', 'decision' and 'understanding'. As for Swedish words not present among the most frequent English words, the three Swedish gendered lexica all have *samarbete* ('collaboration') as either the second or third most frequent word – additionally, according to the embedding-based gendered lexicon, *samarbetsförmåga*, 'ability to cooperate', is

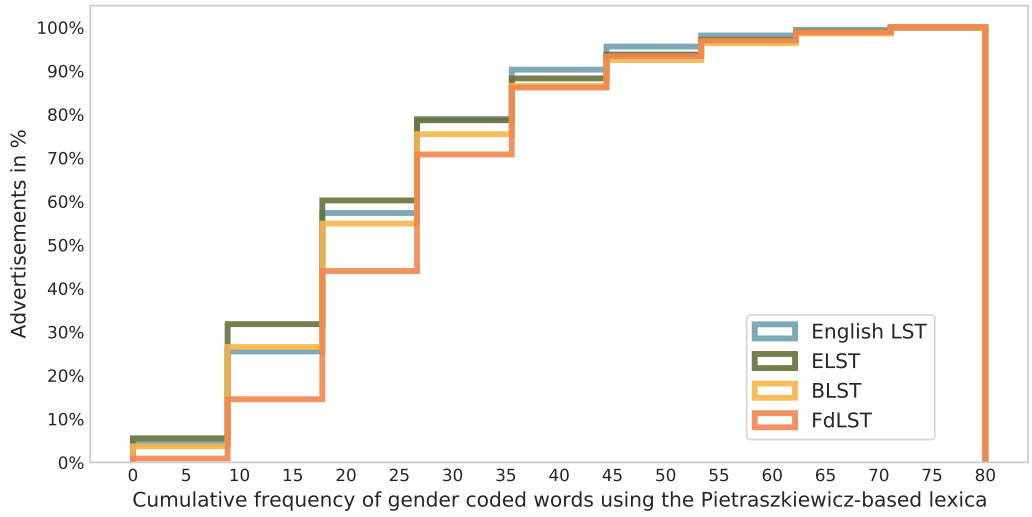

Figure 9: Cumulative frequency distribution of gendered word counts yielded by the English and the Swedish Pietraszkiewicz-based gendered lexica.

the tenth most frequent word – but 'collaboration' is not present among the most frequent English words. A few other words, *komma*, 'come' or 'will', *motsvarande*, 'equivalent', and *utveckla*, 'develop', can be found among the top ten Swedish words, according to the different gendered lexica, but not among the top ten English words.

As with the most frequent words retrieved using the Gaucher-based gendered lexica, the majority of the ten most frequent English words can be found translated among the most frequent Swedish words, using the Pietraszkiewicz-based gendered lexica. However, some of the English words also found among the top ten Swedish words are function words, for example *du*, 'you', *vi*, 'we' and *skola*,[20] 'shall'. *Utbildning*, 'education', and *kunskap*, 'knowledge', both content words, can also be found among both the English and the Swedish most common words. A relatively larger number of the English words, compared to those of the Gaucher-based most frequent words, can *not* be found among the Swedish words. 'Collaboration', for instance, which appeared among the top ten words using all of the Swedish Gaucher-based gendered lexica, does not appear among the top ten words according to either of the Swedish Pietraszkiewicz-based lexica. Mostly due to erroneous translations, a number of Swedish words cannot be found among the top ten English words. For example, in the baseline and the embedding-based gendered lexica, 'you' is translated to *ha*, 'have', and in the frequency-document-based lexicon, 'take' is translated to *och*, 'and'.

6.2 Qualitative evaluation experiments results

As mentioned in Section 5.2, four participants evaluated the quality of the translations of a small subset of the Gaucher et al. (2011) gendered lexica. Calculated with Cohen's Kappa (Cohen, 1960), the inter-

[20] *Skola* which coincidentally also means 'school', is an archaic way of translating 'shall'. The most common ways to indicate that something will occur in the future in Swedish are using *ska* or *kommer (att)*.

Top k~	English LST		BLST		FdLST		ELST	
1	respon-sibility	11.2%	komma	18.7%	motsvarande	7.1%	utveckla	7.4%
2	analysis	9.6%	behörighet	4.8%	samarbete	4.7%	samarbete	5.6%
3	support	6.2%	samarbete	4.1%	kompeten[a]	4.4%	kompeten	5.2%
4	lead	6.2%	kompeten	3.8%	självständig	4.4%	självständig	5.2%
5	decision	4.3%	självständig	3.8%	självständig[b]	4.4%	ansvara	4.8%
6	individual	3.9%	ansvara	3.5%	ansvara	4.0%	individuell	4.7%
7	active	3.4%	individuell	3.4%	individuell	3.9%	leda	4.5%
8	indepen-dently	3.3%	leda	3.3%	leda	3.8%	aktiv	4.1%
9	understanding	3.0%	aktiv	3.0%	aktiv	3.5%	analys	3.7%
10	competence	2.8%	analys	2.8%	analys	3.1%	samarbets-förmåga	3.3%

Table 8: In this table, we display the ten most frequent English and Swedish gendered words from Gaucher, along with the percentages of the top ten gendered vocabulary.

[a]The word *kompeten* is truncated, either because it has been lemmatised to this form or because it occurs in this wrong form in the advertisements.

[b]Some words appear twice in the list of the ten most frequent words due to them appearing in both the masculine and feminine parts of the gendered lexica.

Top k~	English LST		BLST		FdLST		ELST	
1	you	10.5%	för	10.9%	och	28.8%	du	12.4%
2	will	9.9%	som	10.0%	du	9.0%	ha	8.3%
3	we	7.2%	du	8.5%	anställning	5.0%	vi	5.3%
4	education	3.5%	ha	5.7%	vi	3.8%	skola	4.4%
5	should	3.4%	vi	3.6%	komma	3.2%	sökande	3.5%
6	skill	3.2%	skola	3.0%	ansökan	3.0%	utbildning	3.2%
7	teaching	3.0%	komma	3.0%	utbildning	2.3%	förmåga	3.1%
8	knowledge	3.0%	kunna	2.3%	kunskap	1.7%	kunskap	2.3%
9	group	2.6%	utbildning	2.2%	undervisning	1.5%	skicklighet	2.2%
10	collaboration	2.1%	förmåga	2.1%	vetenskaplig	1.2%	skicklighet[19]	2.2%

Table 9: In this table, we display the ten most frequent English and Swedish gendered words from Pietraszkiewicz, along with the percentages of the top ten gendered vocabulary.

English word	Baseline	Word embeddings	Frequency-document	Frequency-sentence
decisions	beslut	beslut	regelboken	regelboken
competency	kompetens	kompetens	landsbygdsmiljöer	urbanism
leads	ledare	leder	leder	forskningscenter
determinations	bestämmning	tillämplighet	kromogena	strukturbestämningar
challenged	utmanad	utmanade	utmanas	tillverkningsteknologin

Table 10: Random sample of translations generated by the different methods. The English words are masculine coded words retrieved from the Gaucher-based gendered lexicon.

Swedish Gaucher-based gendered lexicon	F	M
Baseline	89%	93%
Embedding-based	77%	88%
Frequency-document based	47%	58%
Frequency-sentence based	45%	61%

Table 11: Average qualitative evaluation scores in % by all participants for the translations of the feminine (F) and the masculine (M) coded words of the gendered lexica. A score of 100% would indicate that all participants gave all words a score of 1.

annotator agreement score for the feminine coded words (k = 0.52) showed moderate agreement between the annotators, while for the masculine coded words there seems to be substantial agreement (k = 0.68). All participants evaluated the baseline lexicon as the most successful, followed by the embedding-based lexicon. The frequency-document- and sentence-based gendered lexica were evaluated as the least successful by all annotators.

Examples of translations are provided in Table 10. The baseline and the embedding-based translation methods correctly translate most of the sampled English words, with the exception of *ledare*, 'leader', by the baseline and *tillämplighet*, 'applicability', by the embedding-based method. In contrast, the frequency-based translation methods translate almost all of the sampled English words incorrectly, aside from *utmanas* and *leder*.

All annotators evaluated the translations of the masculine coded words of the gendered lexica with higher scores than the translations of the feminine coded words of the lexica. As demonstrated in Table 11, the baseline translation of both feminine and masculine coded words of the gendered lexica was evaluated as the most accurate, followed by the embedding-based translation. The accuracy scores of the frequency-document- and sentence-based translation methods were significantly lower in comparison with the other two methods.

In all qualitative evaluations, the baseline and the embedding-based translation methods were evaluated with significantly higher scores than the frequency-based translations. The average difference in performances between the baseline and the embedding-based methods is rather small (7%).

The one participant, who evaluated all words from both Gaucher et al. (2011) and Pietraszkiewicz et al.

Swedish gendered lexicon		F	M
Gaucher-based	Baseline	89%	89%
	Embedding	82%	83%
	Frequency-document	47%	59%
	Frequency-sentence	51%	57%
Pietraszkiewicz-based	Baseline	87%	86%
	Embedding	81%	79%
	Frequency-document	46%	47%

Table 12: Qualitative evaluation scores in % by the one participant who evaluated all words from all feminine (F) and masculine (F) coded gendered lexica (Gaucher & Pietraszkiewicz). A score of 100% would indicate that the participants gave all words a score of 1.

(2019), also evaluated the baseline as the most accurate translation method, followed by the embedding-based and the frequency-based methods for both the Gaucher and the Pietraszkiewicz gendered lexica (see Table 12).

The translations of the masculine coded words from Pietraszkiewicz with the baseline, the embedding-based and the frequency-document-based method were given evaluation scores of 86%, 80% and 47% respectively by the one annotator who evaluated all gendered lexica. The corresponding scores for the feminine coded words from Pietraszkiewicz were 87%, 82% and 46%.

6.3 Collective results: which are the best gendered lexica?

After the quantitative and qualitative evaluation of all available gendered lexica, we drew the conclusion that the lemmatised, selectively truncated versions of both the English and the embedding-based Swedish gendered lexica yielded the most reliable results.

The Swedish gendered lexica generated with the embedding-based translation method performed the best in the quantitative analysis, since they consistently yielded the most similar results to the corresponding English lexicon in all four quantitative experiments.

As described in Section 6.2, the baseline and the embedding-based translations were evaluated with significantly higher scores than the frequency-based translations. Given the significantly better performance of the embedding-based gendered lexicon in the quantitative analysis and the small difference in performance between the baseline and the embedding-based gendered lexica in the qualitative analysis, we propose the gendered lexicon created with the embedding-based translation as the most reliable candidate for assessments of job advertisements with respect to the gender they are associated with.

6.4 Testing gendered wording and field of study for correlation: sample study results

In this sample study, the best performing Swedish gendered lexica were used to produce gendered wording scores for the Swedish advertisements provided from KTH. Then, the gendered wording scores were tested for correlation with the dominant gender of the workers in the area that an advertisement is destined for, as described in Section 5.3. Results for the two sets of gendered lexica are available in Figure 10.

Both histograms in Figure 10 demonstrate that there is significant overlap between the gendered wording scores in advertisements for neutral and male dominated areas, which denotes that the two distributions are similar to a great extent. There seems to be a common pattern for advertisements of both areas, irrespective of the dominant gender. More advertisements for male dominated and neutral areas have a positive gendered wording score than they do a negative score. This pattern denotes that the advertisements tend to be written with more feminine coded words.

Although the distributions of the gendered wording scores of the advertisements for male dominated and neutral areas overlap, the two distributions display certain differences. In general, the results of the Gaucher-based gendered lexica demonstrate that a large number of advertisements for male dominated areas appear to have a gendered wording score closer to +1, while a significantly greater number of advertisements for neutral areas seems to have negative gendered wording scores. It should also be noted that the distribution of the Gaucher-based gendered lexica has two peaks of the same height, meaning that the distribution is bimodal. There are two separate values corresponding to two gendered wording scores (0.0 and 0.5) that most of the advertisements appear to have.

Despite the overlap of the two distributions of gendered wording scores generated by the Pietraszkiewicz-based gendered lexica, there are some ways in which the distributions of advertisements for the male dominated and the neutral areas differ. Regardless of the dominant gender of an area, more advertisements are assessed as having more positive than negative gendered wording scores. The distributions show that most of the advertisements, assessed with the Pietraszkiewicz-based ELST lexicon, have a rather neutral-to-positive gendered wording score between 0.0 and 0.25. According to the histogram, no advertisements appear to have a negative gendered wording score between -1.0 and -0.75. It appears that more advertisements for neutral areas have more masculine coded words in comparison with the advertisements for male dominated areas.

The results yielded by the Pietraszkiewicz-based gendered lexica are quite similar to those of the Gaucher-based lexica. Regardless of the dominant gender of an area, the advertisements are assessed as having a higher feminine than masculine wording score, indicating that there is no correlation between gendered wording and dominant gender of area. More advertisements for male dominated areas have positive gendered wording scores, compared to the advertisements for neutral areas. Conversely, it appears that more advertisements for neutral areas have negative gendered wording scores in comparison with the advertisements from male dominated areas.

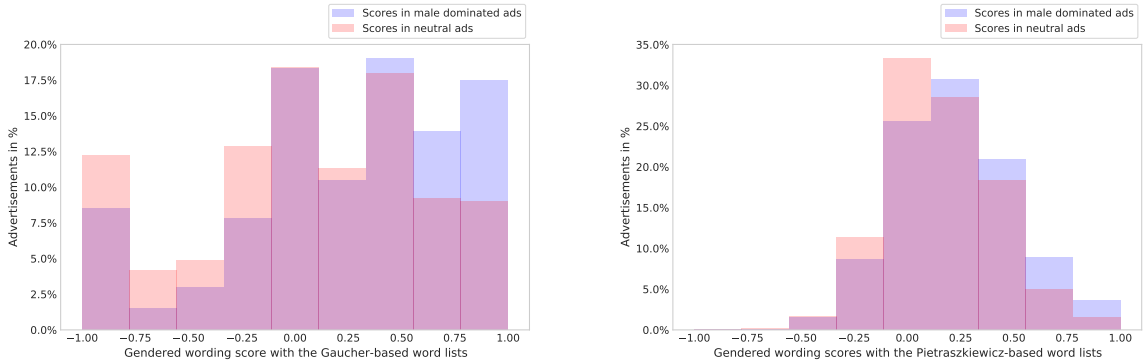

Figure 10: Frequency distribution of gendered wording scores across KTH Swedish advertisements for neutral and male dominated areas, using the Gaucher (left) and the Pietraszkiewicz (right) gendered lexica.

6.5 Job advertisement assessment per university

Using the best performing gendered lexica, we assessed the job advertisements of each university separately, with the goal of examining whether the distribution of gendered advertisements correlates with the type of studies offered by each university. Moreover, studying the distribution of gendered advertisements per university for both Swedish and English would theoretically allow us to discover patterns across languages.

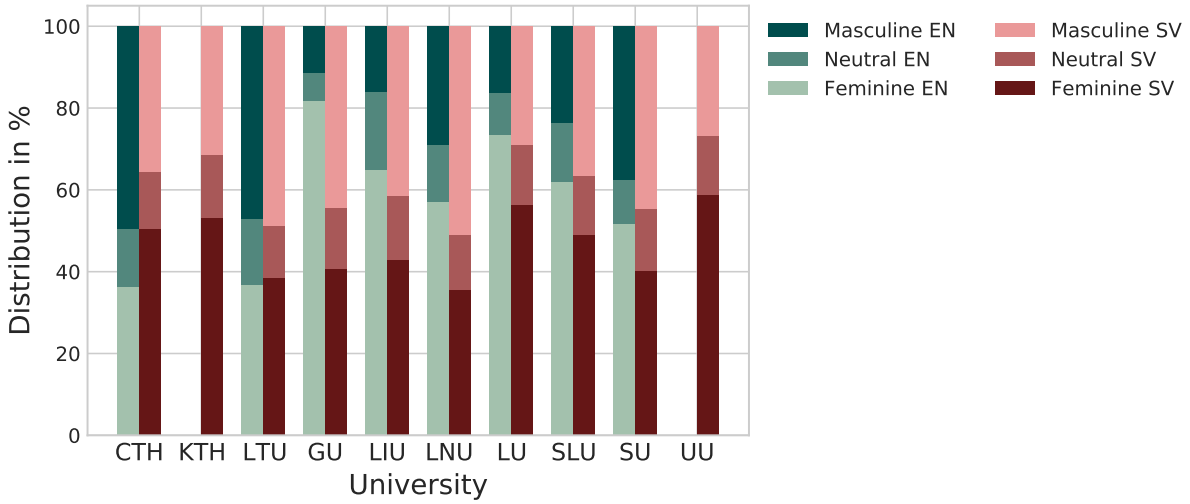

Figure 11: Percentage distribution of feminine coded, neutral and masculine coded advertisements per university using the Gaucher gendered lexica for English and Swedish. As previously mentioned, we only examine the available Swedish advertisements from KTH and UU.

Assessing the advertisements of each university with the original English Gaucher et al. (2011) gendered lexicon returns quite interesting results, which can be seen in Figure 11. Two of the eight universities (CTH & LTU) appear to have more masculine than feminine coded advertisements. Regarding the gender association distribution of the advertisements of the remaining six universities, it appears that more advertisements are feminine coded than masculine coded.

When employing the Gaucher embedding-based lexicon, the results, which can be seen in Figure 11, are more interesting. In four (GU, LNU, LTU, SU) out of ten universities, there are more masculine coded than

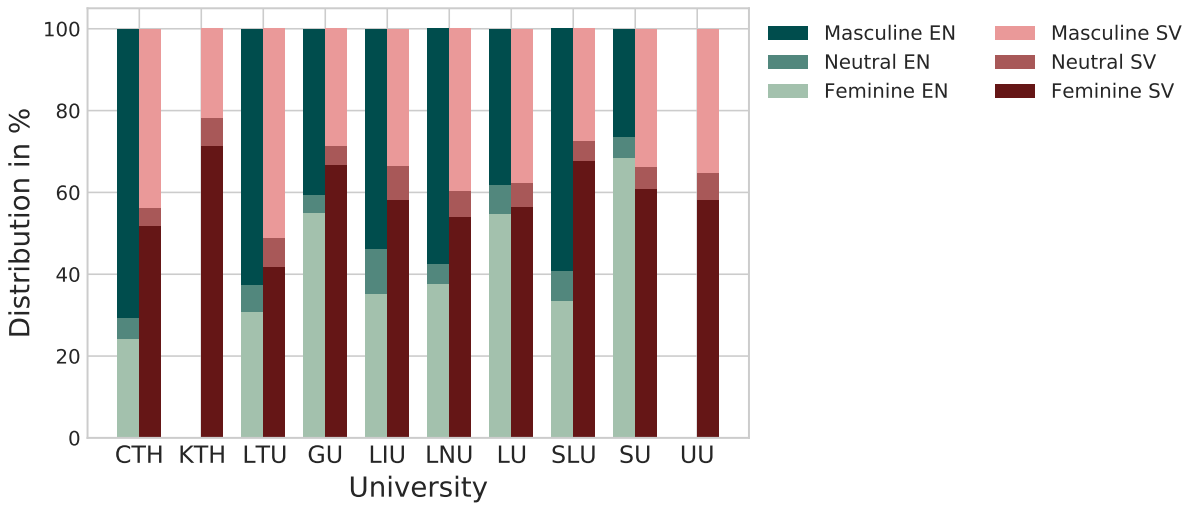

Figure 12: Percentage distribution of feminine coded, neutral and masculine coded advertisements per university using the Pietraszkiewicz gendered lexica for English and Swedish. As previously mentioned, we only examine the available Swedish advertisements from KTH and UU.

feminine coded advertisements. When comparing the distribution of English gender coded advertisements from LNU with the distribution of Swedish gender coded advertisements from the same university, the results seem to be reversed for Swedish. On the other hand, in six (CTH, KTH, LIU, LU, SLU, UU) out of the ten universities, there are more feminine than masculine coded advertisements. Swedish advertisements from LIU, LU and SLU followed the same pattern as they did for English; their advertisements contained a greater number of feminine than masculine coded words. Finally, the CTH advertisements display an opposite distribution for Swedish compared to English.

Assessments of the advertisements of each university with the original Pietraszkiewicz gendered lexicon provide equally interesting results, as seen in Figure 12. Out of the eight universities, five (CTH, LIU, LNU, LTU, SLU) appear to have more masculine than feminine coded advertisements. Conversely, most of the advertisements from GU, LU and SU, appear to be feminine coded.

Assessing Swedish advertisements with the Swedish Pietraszkiewicz embedding-based gendered lexicon produces conflicting results, which can be seen in Figure 12. Only one university (LTU) provided advertisements, the majority of which are assessed as masculine coded (51%); approximately 41% of the available advertisements are assessed as feminine coded. There seems to be a common pattern among the remaining nine universities: the portion of the feminine coded advertisements surpasses the portion of the masculine coded ones.

The English LIU, LNU and SLU advertisements are coded as masculine in the majority of assessments. However, the majority of the Swedish advertisements from these three universities are assessed as feminine coded by the Swedish gendered lexicon. The portion of feminine coded advertisements is greater than the portion of masculine coded advertisements for Swedish in CTH.

7 Analysis & discussion

In this section, we analyse the translation methods we employed, along with the results yielded by the quantitative and the qualitative experiments. We also examine the findings of the two studies that looked into the advertisements of each university separately. Finally, we discuss the limitations of our research.

7.1 Discussion of quantitative results

As described in Section 3.3, four different translation methods (Google Translate, frequency-document-based, frequency-sentence-based, word-embedding-based) were employed to develop gendered lexica in the Swedish language. The success of the translation methods was tested in five different experiments, culminating in the word-embedding-based method being selected as the most appropriate. Before we start investigating the results of the different experiments, it should be stated again that because of the small overlap of the Gaucher-based and the Pietraszkiewicz-based gendered lexica, their results do not align.

In the first experiment, we compared the distributions of job advertisements assessed by different versions of the gendered lexica, which showed that the results are rather stable for all of the Gaucher-based gendered lexica. However, one thing to note is that the English percentages of feminine coded advertisements are consistently higher than the corresponding Swedish percentages. This may be due to a number of reasons – if we assume that the Swedish lexicon denotes agency and communion accurately, *and* that the Swedish advertisements are a perfect translation of the English advertisements, we can draw the conclusion that the Swedish advertisements do in fact contain less feminine coded words and more masculine coded words compared to the English advertisements, as defined by the Gaucher-based lexica. However, it may be the case the Swedish lexicon denotes agency and communion accurately, but that the Swedish advertisements are *not* a perfect translation of the English advertisements. In that case, we would not know whether the crosslingual differences in our results are valid, or whether these differences stem from stylistic differences of the various authors of the Swedish and English advertisements. Further, it may be the case the Swedish lexicon does not represent agency and communion accurately in Swedish, in which case we would, yet again, not be able to ascertain the validity of the crosslingual differences. As mentioned in Section 6.2, the translations were not rated with perfect scores during annotation, indicating that certain words might not denote agency or communion, yet are employed in the assessments. Because the translations are not perfect, they could be partly (or fully) responsible for the crosslingual differences in the results. In order to determine the cause of the ratio discrepancies, all of the mentioned factors would need to be examined in detail.

Regarding distributions of job advertisements assessed using the Pietraszkiewicz-based gendered lexica, significant discrepancies can be seen in the results. Due to the extremeness of the results when employing the CT versions of the Pietraszkiewicz-based lexica,[21] we decided that any efforts at analysing these results would most likely be fruitless, as non-gendered wording could potentially be captured as gendered and skew

[21] The CT versions allow all words starting with short words from the gendered lexicon to count as gendered words in the advertisements. For example, the word *vi*, 'we', is the start of the interrogative/relative pronoun *vilken/-et/-a*, 'which', which would be counted as a gendered word.

the results. Generally, the results of the ST versions of the Pietraszkiewicz-based gendered lexicon indicate that feminine coded words are more prevalent in the Swedish advertisements than in the English advertisements. The Swedish distributions tend to vary similarly, however, the contrast between the feminine and the masculine coded advertisements is significantly larger when assessed with the frequency-document-based gendered lexicon. Finally, it should be stated that both before and after lemmatisation, the results of the ST versions of the gendered Pietraszkiewicz-based lexicon remain stable.

Some of the reasoning presented as a means of understanding discrepancies in the results of the Gaucher-based gendered lexica may also provide some insight into the discrepancies found in the results of the Pietraszkiewicz-based lexica. For example, the quality of the translation may affect the Pietraszkiewicz-based lexica in the same manner as the Gaucher-based lexica. Like the Gaucher-based lexica, the translations of the Pietraszkiewicz-based lexica are not given perfect scores, indicating that some words in the lexica are not representative of agency or communion, in which case ascertaining the validity of the crosslingual variation in the results would require further investigation. Assuming that the advertisements are translated correctly and that the lexica denote agency and communion accurately, it could also be the case that the Swedish advertisements are simply worded with more feminine and less masculine coded words, as defined by Pietraszkiewicz, while the English advertisements contain more masculine than feminine coded words.[22] As with the discrepancies in the Gaucher-based results, the advertisements may not be translated entirely correctly, in which case writing style could be the cause of the variation across languages. The different number of feminine and masculine coded, truncated words in the lexicon – there are approximately ten more feminine than masculine words that are full-form and 30 more feminine than masculine coded words that are truncated in the lexicon – could lead to a larger number of words being captured as feminine coded words.

A big part of the first experiment was determining which settings (inflected or lemmatised, comprehensively truncated or selectively truncated) of the gendered lexica generate the most reliable and consistent results. Lemmatising allows inflected forms of a word to be grouped into one lemma, saving on computational resources as the gendered lexica are smaller, while simultaneously allowing more items to be captured from one lemma. Besides, given that the inflected versions of the gendered lexica include different forms of the same lemma, they can also skew the results. For example, if *påverkas*, 'be influenced by', is in the gendered lexicon, but *påverka*, 'influence', is in the text, the latter would not count as a gendered word. Conversely, if a word, *sammanslutning*, and its plural form, *sammanslutningar*, are in the gendered lexicon, and the plural form is in the text, this word would be counted twice – once for the singular form, and once for the plural form.[23]

As for truncation, treating all words as truncated clearly skews the results of the assessments (see Table 7), since random words are treated as gendered words, if they start with any of the words from the gendered lexica. While this does not have a big impact on the Gaucher-based gendered lexica, it greatly affects the Pietraszkiewicz-based gendered lexica, which contain a substantial number of function and short words.

[22] As the gendered lexica by the two authors differ substantially, it is possible that the Swedish advertisements contain more feminine than masculine coded words according to the Pietraszkiewicz-based lexica, and simultaneously contain more masculine than feminine coded words according to the Gaucher-based lexica.

[23] That is, if the shorter word in the gendered lexicon is truncated.

Selectively truncating is employed to prevent unpredictable results, since it captures the words that were intended by Gaucher and Pietraszkiewicz to be captured while also limiting the risk of capturing non-agentic and non-communal words by mistake. For instance, the selectively truncated version requires exact matches for words, such as 'we' and 'you' from Pietraszkiewicz et al. (2019), and their Swedish translations, *vi* and *du*, all of which are not suffixed with an asterisk. It captures only these exact words and not other words that would be captured if we were to use the comprehensively truncated version of the gendered lexicon, which would allow non-agentic and non-communal words (e.g., *vibrationen*, *dubbel*) to incorrectly count as gendered. All aspects considered, the lemmatised and selectively truncated versions of the gendered lexica were chosen, as they yielded the most consistent results.

The second and the third quantitative experiment, which investigated the frequency distributions of gendered wording scores and gendered word counts of the English and the different Swedish gendered lexica respectively, showed relatively consistent results.

For the Gaucher-based gendered lexica, the word-embedding distribution was consistently most similar to the English distribution in both experiments, proving to be the most reliable translation method. On the other hand, the most dissimilar distribution was generated by the Lindmark & Sundin (2013) gendered lexicon in the second experiment. In the third experiment, the most dissimilar distribution to that of the English gendered lexicon was generated by the frequency-sentence-based lexicon. Figure 6 shows, as previously mentioned, that the original English and the embedding-based gendered lexicon yield a much lower percentage of advertisements with a score of 0.00 than the other gendered lexica. This suggests that the baseline, the frequency-document-based, the frequency-sentence-based and the Lindmark & Sundin gendered lexica assess a larger amount of advertisements as having a negative or neutral gendered wording score.

For the Pietraszkiewicz-based gendered lexica, as shown in Figure 7, the word-embedding distribution displayed the most similar behavior to the English distribution in the second experiment, while the distributions of the other two Swedish lexica are further away. This proves how different the translations are – were the translations more similar, the results would have been more similar. In the third experiment, there are some inconsistencies with both the baseline and the embedding-based gendered lexicon displaying most similar performances to the English lexicon in different parts. As with the discrepancies in the Gaucher-based results, the greatest Pietraszkiewicz-based discrepancies can be observed between the distributions of the English and the frequency-document-based gendered lexica in both experiments.

From these two experiments, we can conclude that the dissimilarities of the gendered lexica, and, more specifically, the translations of the English gendered lexica, are more substantial than what could be gathered from the first experiment. The cause of the discrepancies is the differences of the translation methods employed. We theorise that the embedding-based method is superior due to its ability to produce and connect more complex meaning representations. That is, we can obtain a word that is semantically related to our target word by employing monolingual aligned embeddings, since words with similar meaning have similar coordinates, despite not being direct translations. Besides, aligned word embeddings have proved to provide appropriate translations, even for words of distant language pairs (Joulin et al., 2018). Additionally, as the embeddings are pretrained on millions of textual documents, the embedding-based method will

undeniably have an advantage over the frequency-based methods, which were trained on a corpus of 5000 documents. In theory, the frequency-based method should work in a similar fashion to the embedding-based method, however, perhaps due to the limited size, sparsity and domain-specificity of the data, the frequency-based method generated a large number of nonsensical translations. Furthermore, we theorise that the embedding-based method outperforms the baseline as we translate words without context – Google Translate is likely modelled to translate words within a sentence by looking at the sentence as a whole, i.e., by using contextual information. As a result, Google Translate leads to literal translations for a great number of out-of-context words. However, words convey meaning in different manners crosslingually. Thus, our baseline word-to-word translation also lead to Swedish misinterpretations of certain English words.

An interesting aspect of our experiments is the discrepancies between the English, the baseline and the Lindmark & Sundin (2013) gendered lexica. We would expect their results to be similar since Lindmark & Sundin created their Swedish gendered lexicon based on Gaucher by using Google Translate and manually choosing the best translation that fit the purpose of their research. However, from the results of the second experiment, we gather that that is not the case: as shown in Figure 6, the line of the Lindmark & Sundin lexicon is significantly higher than that of the English and the baseline lexica, indicating that the former lexicon over-represents masculine coded words in the Swedish advertisements. The irregularities between the two Swedish versions of the gendered lexica can be attributed to the fact that the architecture of the models that Google Translate uses has changed from 2013 to 2021, along with their difference in size (the baseline gendered lexicon contains 15 more gendered words in comparison with the Lindmark & Sundin gendered lexicon), which is a result of the word expansion process described in 4.1.2.1 (the words in the Lindmark & Sundin lexicon are not expanded since they are already in Swedish, and the expansion was implemented to allow for as reasonable translations as possible).

In the Gaucher-based assessments, both the Swedish and the English advertisements are mostly feminine coded, however, the number of Swedish masculine coded advertisements is higher than the number of English masculine coded advertisements. The tendency of the Gaucher-based Swedish gendered lexica (especially the frequency-document-based, the frequency-sentence-based, the baseline lexica) to assess more advertisements as having a negative or neutral score, compared to the assessment using the English gendered lexicon, can be attributed to the simple fact that, according to the Gaucher-based lexica, the English advertisements contain more feminine coded words and less masculine coded words than the Swedish advertisements, as suggested for the first experiment. It may also be the case that the wording of the English and Swedish advertisements differ due to writing style of the authors of the advertisements. Furthermore, the translation of the words is not absolute – the Swedish translation may not convey agency and communion in the same manner that the original English lexicon does, causing a discrepancy between the English and Swedish results.

Additionally, there is a tendency of the Pietraszkiewicz-based Swedish gendered lexica to assess more advertisements with a positive gendered wording score, compared to the Pietraszkiewicz-based English gendered lexica, which assess more advertisements as having a negative gendered wording score. This can be explained by similar reasoning as proposed for the Gaucher-based discrepancies. That is, the crosslingual discrepancies may be due to authorship differences and/or translation. Furthermore, since the Gaucher-based lexica and the Pietraszkiewicz-based lexica differ greatly, it is possible that the Swedish advertisements con-

tain more feminine coded words than the English advertisements, as defined by the Pietraszkiewicz-based lexica, while simultaneously containing more masculine coded words than the English advertisements, as defined by the Gaucher-based lexica.

As demonstrated earlier in this section, the results yielded by the Gaucher-based gendered lexica have been consistently more stable than the results of the Pietraszkiewicz-based gendered lexica. After investigating the ten most frequent gendered words in the English and Swedish job advertisements in the fourth experiment, according to the different versions of the Gaucher- and the Pietraszkiewicz-based gendered lexica, we made some interesting observations. In general, the ten most frequent English Gaucher-based words correspond better to the different Swedish translations, in comparison with the Pietraszkiewicz-based gendered lexica. There is a strong tendency for semantically similar words to appear among the most frequent Gaucher-based gendered words for all different translations, suggesting that the different gendered lexica have some similarities and are able to capture some common fundamental concepts of agency and communion. The Swedish Gaucher-based gendered lexica differ to a certain extent because of erroneous translations – for example, *utveckla*, 'develop', is a mistranslation of 'individual*' that only appears among the most frequent words generated using the embedding-based gendered lexicon. The most frequent Gaucher-based words denote more abstract concepts, related to behavioral traits that are expected of applicants (such as *självständig*, 'independent', *ansvara*, 'be responsible (for)'), whereas the most frequent Pietraszkiewicz-based words are more domain-specific to academia (for example, *utbildning*, 'education', *undervisning*, 'teaching') and to the process of job application (for example, *anställning*, 'employment', *ansökan*, 'application'). Additionally, function words appear to be among the most frequent Pietraszkiewicz-based words (e.g., *du*, 'you', *skola*, 'shall', *ha*, 'have'). Regarding *skola*, it may be the case that this word is counted incorrectly as it is a homonym – *skola* is semantically ambiguous and can mean either 'shall/will' or 'school'. This issue is mainly due to lemmatisation – had *skola* not been lemmatised, it would have taken the form *ska* and would thus have been unambiguous. Finally, a similar tendency observed in the most frequent Gaucher-based words can be found among the most frequent Pietraszkiewicz-based words as well but to a lesser degree; semantically related words appear in all translations.

As shown in the previous experiments, the Gaucher-based gendered lexica performed more consistently in both English and Swedish data, in comparison with the Pietraszkiewicz-based gendered lexica, whose assessments displayed rather significant irregularities. The Gaucher-based gendered lexica consist of content words that denote agentic and communal traits. Unlike the Pietraszkiewicz-based gendered lexica, they do not contain any function words, which, according to linguistic theory, have little meaningful content. Since function words have been said to not carry any meaning, we can also theorise that they are not good indicators of agentic or communal meaning either. On top of that, function words would skew the results with their extremely high frequency of occurrence. As a result, we conjecture that the function-word-free, Gaucher-based gendered lexica with their consistent performance are more apt for detecting gender bias in job advertisements.

7.2 Discussion of qualitative results

As reported in Section 6.2, the baseline translation method was evaluated as having the best translations, followed closely by the embedding-based translation. The two variations of the frequency-based translations

were consistently rated with lower scores.

The higher quality of the baseline gendered lexicon can be explained by the fact that Google Translate, a state-of-the-art translation system, performed the translation of the English items. Even when dealing with out-of-context words and expressions, most translations were reasonable. The embedding-based gendered lexicon also performed well in the qualitative analysis. Expectedly, as shown in Conneau et al. (2018), aligning monolingual word embedding subspaces in an unsupervised way produces translations of high quality between multiple language pairs. The frequency-based translation proved to be the least successful, which was expected, as it is a more naive approach, using only word counts. The poor performance can be attributed to the really low frequency of certain gendered words in the English data, which were inevitably represented with very sparse count vectors. When linking a very sparse English vector to a Swedish vector using cosine similarity, the Swedish vector was, similarly, quite sparse. However, because many of the English gendered words were rare, they were all linked to the same, equally rare Swedish vector. For example, approximately 20% of the masculine coded words and 28% of the feminine coded words were all translated to the word *forskningsinriktningar*, 'research specialisation/objective'. The vector representing *forskningsinriktningar* was as sparse as the vectors of the rare, English gendered words it had been linked to. Consequently, the frequency-sentence- and document-based translations started with a disadvantage. However, the frequency-based methods returned the correct translations in some cases, where the other two translation methods did not. For instance, the English word 'independence' was correctly translated to *självständighet* with the frequency-sentence- and document-based methods, but not with the baseline method – which returned the word *oberoende*, which may be semantically correct, but is of a different part of speech. As with most, if not all, machine-based translation systems, the performance was not perfect, but satisfactory regardless.

7.3 Testing gendered wording and area for correlation

As described in Section 5.3, we conducted a small sample study to investigate if there is a correlation between the gendered wording score of an advertisement and the dominant gender of the area that the advertisement is intended for.

The histograms in Figure 10 show that there is a common pattern for advertisements of neutral and male dominated areas. Regardless of the gender of the area, more advertisements appear to have positive gendered wording scores, i.e., they tend to be written with more feminine than masculine coded words. When assessed with the Gaucher-based embedding-generated lexicon, more advertisements for male dominated areas seem to have positive gendered wording scores, while a greater number of advertisements for neutral areas appear to have negative gendered wording scores. On the other hand, when assessed with the Pietraszkiewicz-based embedding-generated gendered lexicon, most advertisements of both male dominated and neutral areas are assigned gendered wording scores that range from 0.0 to 0.25, indicating that they are worded with more feminine than masculine coded words. There is also a tendency for advertisements for neutral areas to have more masculine coded words in comparison with the advertisements for male dominated areas.

The main observation from this study is that the Swedish advertisements from KTH tend to be worded with

more feminine coded words. Such results go against the findings of the studies conducted by Gaucher and Pietraszkiewicz. The lack of correlation between the gendered wording score of an advertisement and the dominant gender of the area is in line with our initial hypothesis (see Section 5.3). We did not expect to get similar results to Gaucher and Pietraszkiewicz, since all KTH advertisements are standardised and are created by the same employer, namely KTH. As a result, it is more than likely that there is minimal variation in word use across all advertisements, which would explain how most advertisements seem to have positive gendered wording scores.

Research has shown that advertisements for male dominated areas that are worded with more feminine coded words appear as more appealing to women than strongly masculine worded advertisements for male dominated areas (Gaucher et al., 2011). We theorise that the advertisements for male dominated areas are worded with more feminine wording in an effort, perhaps subconscious, to encourage women to become part of the traditionally male dominated STEM fields.

It has also been noted that women tend to compose their resumes using more communal wording, as opposed to men, who use more agentic wording (Ng et al., 2020). We speculate that this may be the case when writing job advertisements as well; the wording of a job advertisement may depend on the gender of the author. However, as we do not have access to the gender of the author of each advertisement, this cannot be further examined.

7.4 Assessment per university

As stated in Section 5.4, we assessed the English and Swedish advertisements of each of the eight universities that had a sufficient number of advertisements in both languages separately. Our goal was to investigate if the distribution of gendered advertisements correlates with the type of studies offered by each university.

Yet again, assessing the job advertisements with the Gaucher-based gendered lexica yielded more consistent results than the Pietraszkiewicz-based gendered lexica. CTH and LTU had more masculine coded advertisements, whereas the other six universities were assessed as having more feminine coded advertisements. CTH and LTU having more masculine than feminine coded advertisements is expected, as such distribution is in line with the findings of Gaucher et al. (2011) and Pietraszkiewicz et al. (2019). As stated in Section 3.1, both CTH and LTU offer education within the traditionally male dominated STEM fields. We theorise that there is a correlation between the existence of more masculine coded words in an advertisement and the gender of the majority of students that study within the area that the advertisements are for. On the contrary, the type of education provided by universities, such as GU, LIU, LNU, LU, SLU, SU, is very wide, ranging from human science to medicine to law, which usually attracts an equal amount of women and men (Statistiska Centralbyrån, 2012). Thus, we can also assume that there is a correlation between the existence of more communal words in an advertisement and the gender diversity across the fields of study that the advertisements are for. Specifically, for GU and SU, which provide education on a wide variety of fields, where both women and men are employed evenly, according to SCB (Statistiska Centralbyrån, 2012), there is only a 4% difference between the masculine and the feminine coded advertisements assessed with the English Gaucher-based gendered lexicon.

However, when assessing the Swedish advertisements, GU, LNU, LTU and SU appear to have more masculine coded advertisements. The majority of the advertisements from the remaining universities is feminine coded. An unexpected finding was that most of the Swedish CTH advertisements were assessed as feminine coded, while the English CTH advertisements appeared to be mostly masculine coded. It is also interesting that CTH is the only university, whose portion of Swedish feminine coded advertisements is greater than the portion of English feminine coded advertisements. As this irregularity is specific to the advertisements from one out of eight universities, we can infer that the cause must be the advertisements, not the gendered lexicon. A possible explanation for the reversal of the feminine to masculine ratio of the CTH advertisements is that the English and the Swedish data are not translations of each other. The same reasoning can be applied to the other universities where there is a large discrepancy between the English and Swedish distributions, that is, GU, LIU, LNU LU, and SLU. Another possibility is that the advertisements from universities with large discrepancies in the distributions are authored by different people in Swedish and in English, while the advertisements from universities with similar distributions across the two languages, from LTU and SU, are authored by the same person.

When employing the original English gendered lexicon by Pietraszkiewicz, again, CTH and LTU, universities that offer education within the traditionally male dominated STEM fields, have more masculine coded advertisements, along with LIU, LNU, and SLU. The fact that LIU, LNU, and SLU are also assessed as masculine coded can be attributed to the frequency of the agentic word 'you', according to Pietraszkiewicz et al. (2019) ('you' appears 1.45 times as much as the communal word 'we' in the English advertisements). Conversely, GU, LU and SU advertisements are assessed as feminine coded to a greater degree. Again in this case, we assume that there is some correlation between the stronger presence of feminine coded wording in the advertisements and the range of courses they provide.

Assessing the advertisements with the Swedish Pietraszkiewicz-based gendered lexicon, on the other hand, results in seven of the universities having more feminine coded advertisements. Such a reversal in the distributions across the two languages could be explained if we take into consideration the reasoning we proposed in Section 7.1: if we assume that the Swedish lexicon is a good representation of agency and communion, and that the Swedish advertisements are translated correctly from English, the conclusion is that the crosslingual variation is due to language specific features. Thus, it could be the case that the Swedish advertisements simply do contain more feminine and less masculine coded words, while the English advertisements contain more masculine and less feminine coded words, according to the Pietraszkiewicz-based gendered lexicon. The only university that does not align with this reasoning is LTU, which has more masculine coded advertisements in Swedish. This exception might be related to the fact that LTU provides education within the STEM fields. On the other hand, as suggested regarding the first experiment, if the Swedish lexicon does not denote agency and communion accurately and/or if the Swedish advertisements are not good translations of the corresponding English advertisements, it would not be possible to attribute the crosslingual differences to a specific reason. The cause of the crosslingual differences could be due to stylistic differences of the authors of the advertisements or simply due to the translation of the lexicon.

Moreover, it is worth noting that the portion of neutral advertisements is significantly smaller when assessing with the Pietraszkiewicz-based gendered lexicon than the Gaucher-based gendered lexicon. This can be attributed to the frequency with which certain Pietraszkiewicz-based words (e.g. function words) ap-

pear in the advertisement text. As the Pietraszkiewicz-based gendered lexicon contains a number of short, frequent words such as *vi*, 'we', and *ha*, 'have', which mostly express grammatical relationships between content words, an advertisement is more likely to be assessed as either masculine or feminine coded. Finally, the significantly bigger size of the Pietraszkiewicz-based gendered lexicon in comparison with the smaller Gaucher-based gendered lexicon and the lack of overlap between the two may also contribute to the discrepancies of the assessments as conducted by the Pietraszkiewicz-based and the Gaucher-based gendered lexica.

7.5 Limitations and weaknesses

As with most research, some limitations should be noted in the development of the gendered lexica, in the assessments of the advertisements and, consequently, in the evaluation of the gendered lexica.

The translation methods we employed are all limited to some extent. First, translating words out of context can be problematic. Google Translate is not always an entirely reliable translation service; it offers only a literal translation since there is no context to take into account, which occasionally results in nonsensical Swedish words. Frequency-based translations are also limited since the bilingual document- and sentence-based corpora are not big enough to represent semantic relationships between words in a reliable manner that is informative enough for the purposes of our research. Regarding the limitations of the embedding-based translation method, it is crucial to remember that the fastText embeddings have shown to be biased in various NLP applications (Sahlgren & Olsson, 2019). Thus, the embedding-based gendered lexica will also be biased. However, it might be required for the Swedish lexica to consist of biased words so that gender bias can be accurately captured. For instance, the masculine coded word 'athletic' was translated as *idrottsman*, 'athlete-man'. The bias in this particular word would likely not be an issue in our study, as 'athlete' is a masculine coded word – a masculine coded word that contains the inherently masculine gendered word 'man'. However, if 'athlete' was a feminine coded word that was translated to *idrottsman*, the situation would be different.

Another issue that should be raised is that certain words from both the Gaucher and the Pietraszkiewicz gendered lexica, such as 'submissive', 'feminine', 'aggressive', do not appear in the advertisements. As a result, it could be argued that these words can be excluded from the Swedish gendered lexica, especially since they did not occur in the advertisements. Making a gendered lexicon as concise as possible is a necessary procedure to render our pipeline as computationally efficient as possible.

It should also be noted that the lack of transparency behind the creation of the original English gendered lexica is problematic to some extent. We argue that the reasoning behind the creation of the English gendered lexica is necessary and could also aid in the development of gendered lexica in other languages. The same holds for the use of asterisks; the fact that certain words are suffixed with an asterisk, and others are not, is not properly explained by neither Gaucher, nor Pietraszkiewicz. Having some insight on how the addition of asterisks was determined would also facilitate the development of gendered lexica in other languages.

Another limitation concerns the differences between the gendered lexica created by Gaucher and Pietraszkiewicz. The Gaucher lexicon is easier to translate as it exclusively consists of content words. The Pietrasz-

kiewicz lexicon contains a significant number of function words that can potentially skew the results. As function words, like 'you' and 'we', are used in computational authorship attribution (Kestemont, 2014), one can argue that the Pietraszkiewicz gendered lexicon is to an extent capturing writing style instead of agency and communion. The frequency of such function words in the text of job advertisements is a good indicator as to why the Pietraszkiewicz lexicon is harder to translate, and thereby to employ, in comparison with the Gaucher lexicon. The translatability of the Gaucher lexicon renders it as the most suitable tool for assessing the gender of an advertisement. To sum up, it seems to be more efficient to translate and utilize only content words as a means of capturing agency and communion in text.

Weaknesses can also be found in the tools we used. The Gender Decoder tool is not context sensitive. It is very simplistic, as it uses frequency counts to assess an advertisement with respect to gender association. Neither semantic nor syntactic information is employed. As a result, in a rare case, where an advertisement presents undesirable traits, a negated agentic term (e.g. 'not responsible') would still count as a masculine coded word. Similarly, semantically ambiguous words, such as *skola*, 'shall' and 'school', complicate the assessments. In some cases the truncation-sensitive Gender Decoder tool captures the wrong word sense ('school') of a polysemous word (*skola*) as gendered, which inevitably skews the results. In this particular example, the lemmatisation was the cause of the wrong assessment, suggesting that lemmatisation might not always yield the most reliable results. However, it is likely that similar issues would arise when choosing not to lemmatise. Word sense disambiguation and taking into consideration dependency relations between words could easily account for the varied linguistic options of different authors.

Moreover, the frequency of feminine and masculine coded words in the advertisements were low in some cases, allowing just a couple of words to make a difference in the assessments of the advertisements with respect to gender. However, as Madera et al. (2009) argue, even the smallest difference in distribution is significant enough to make a difference in a quantitative analysis.

One more limitation in our research was the lack of a gold standard in the form of gender coded advertisements or publicly available Swedish gendered lexica. Having a gold standard available would allow us to evaluate the gendered lexica we developed with the different translation methods. However, in the absence of a ground truth, we were limited to conduct the quantitative and qualitative evaluation experiments we presented in Section 5.

Lastly, as we do not have a gold standard, it is impossible for us to ascertain whether agency and communion are expressed in the same way in English and Swedish. That is, there is no way of determining whether the gendered lexica we have created express agency and communion in Swedish in the same way that the gendered lexica developed and verified by Gaucher and Pietraszkiewicz do in English.

Despite the limitations of our research, the detection of gender bias in job advertisements – in languages other than English – is a really important task. We hope that our approach can be a first step in developing gendered lexica for detecting gender bias and spreading awareness with regards to the importance of inclusivity and gender neutral wording in job advertisements.

8 Conclusion & Future Work

In our thesis we have presented an evaluation of Swedish gendered lexica generated using different translation methods. The researched translation methods are Google Translate, frequency-based translation and embedding-based translation. After multiple quantitative and qualitative experiments, the lexica generated with the embedding-based translation were consistently evaluated as the best performing. Thereafter, the best performing gendered lexica were employed to assess advertisements to examine if there is a correlation between gendered wording in an advertisement and the dominant gender in the occupation that is advertised. Our study shows that regardless of the dominant gender in the occupation that is advertised, advertisements tend to be worded with more feminine coded words. Another study was conducted to investigate the distribution of gendered advertisements from each university separately. As we had advertisements from both technical universities and universities that offered a wider range of education, we were also able to examine if there was a correlation between gendered wording and the type of education that was offered by the university that created the advertisement. There was a tendency for advertisements from technical universities to be masculine coded. On the other hand, there were more feminine coded advertisements in universities that offered education on a wider range of fields.

There are many directions that our research could take in the future. To ensure the quality of the gendered lexicon, it would be fruitful make some additional changes to the best-performing embedding-based gendered lexicon, since it contains certain wrongly-translated words along with words that do not denote concepts relevant to agency and communion. This could be implemented by replacing translations that are annotated by all participants with a score of -1 or 0 in the qualitative study with translations that capture agentic and communal concepts. Most importantly, employing annotators and translators that are informed about the nature of the required translations could benefit our research.

It would also be of interest to explore more translation methods. Since there are so many types of publicly available pretrained embeddings, it would be interesting to align more embeddings and investigate if they provide translations of higher quality than the fastText embeddings we used. Specifically, we would like to utilise contextual word embeddings, as we believe that assigning each gendered word a representation based on its context would allow us to capture word use across varied contexts and encode knowledge that transfers across languages.

Moreover, regarding the embedding-based translation method, future research could continue to explore the possibility of extending the gendered lexicon by also considering the second nearest neighbor of the English gendered word as a possible translation. On the grounds of the distributional hypothesis, it is likely that in an embedding space, a neighboring word to an agentic or communal word would capture agency or communion as well.

Future experiments should certainly further test our tool and gendered lexicon on real-world (i.e., non-university, non-standardised) advertisements in order to investigate if the assessments would yield similar results.

Developing a more linguistically-informed pipeline for assessing advertisements is another challenging topic

for future work. As mentioned in Section 7.5, our tool employs a simple dictionary lookup method to locate gendered words in the advertisements. However, because of its simplicity, it fails to take into consideration complex semantic and syntactic relations between words. To achieve a more reliable performance, a dependency parser (for example, Stanza[24]) could be employed so that negated gendered words would be disregarded in the assessment of an advertisement. Additionally, utilising a semantic lexicon (e.g. Saldo[25]) could aid in capturing only the correct word sense of a semantically ambiguous word as gendered.

Further ways to evaluate the gendered lexicon include methods from Gaucher and Pietraszkiewicz: a) assessing the gender coded advertisements with human participants to determine whether a gendered advertisement appeals to some participants more than others, and b) consulting Swedish psychologists, psycholinguists and sociolinguists to get their opinion on whether the Swedish words accurately denote agency and communion.

Finally, future studies should investigate the use of topic modeling on job advertisements. Topic modeling is a statistical method that discovers some abstract "topics" that occur in a collection of documents. It would be interesting to examine if there is an intersection between the words that appear as topics in the advertisements and the gendered words in our gendered lexicon.

We hope that our research in developing Swedish gendered lexica can contribute to the detection and mitigation of gender bias in the workplace.

References

Allén, S. (1977). Några könsskillnader i joumalistspråk. In *Modersmålet i fäderneslandet: ett urval uppsatser under fyrtio år*, volume 25 of *Meijerbergs arkiv för svensk ordforskning* (pp. 393–395). Göteborg: Meijerbergs institut för svensk etymologisk forskning. 1999.

American Psychological Association (2014). Answers to your questions about transgender people, gender identity, and gender expression. https://www.apa.org/topics/lgbtq/transgender. Accessed: 2021-05-13.

Andersson, M. & Nilsson, T. (2014). *Den passionerade arbetssökaren: En textanalys av jobbannonser i svenska data och tvspelsbranschen*. Bachelor thesis, Umeå University, Umeå.

Bhowmick, P. K., Basu, A., & Mitra, P. (2008). An Agreement Measure for Determining Inter-Annotator Reliability of Human Judgements on Affective Text. In *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics* (pp. 58–65). Manchester, UK: Coling 2008 Organizing Committee.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.

[24]Stanza is available at: https://spraakbanken.gu.se/en/resources/stanza_synt
[25]Saldo is available at: https://spraakbanken.gu.se/resurser/saldo

Chamera, C. & Kühnemann, N. (2016). *Den kodade vinnaren: En diskursanalys om könskodning i IT-branschens jobbannonser*. Bachelor thesis, Uppsala University, Uppsala.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Conneau, A., Lample, G., Ranzato, M., Denoyer, L., & Jégou, H. (2018). Word Translation Without Parallel Data. In *International Conference on Learning Representations*.

Crawford, K. (2017). The Trouble with Bias. [Facebook video]. https://fb.watch/5ASOj30EMl. [YouTube video] https://youtu.be/ggzWIipKraM. Accessed: 2021-05-12.

Delamater, R. J. & Mcnamara, J. R. (1986). The Social Impact of Assertiveness: Research Findings and Clinical Implications. *Behavior Modification*, 10(2), 139–158. PMID: 3729911.

Eagly, A. (1987). *Sex Differences in Social Behavior: A Social-Role Interpretation*. Hillsdale, New Jersey: Lawrence Erlbaum.

Einarsson, J. (1979). Två roller söker sina egenskaper: en pilotstudie. *Könsroller i språk*, 3, 3–17.

Einarsson, J. (2009). *Språksociologi*. Lund: Studentlitteratur, 2nd edition.

Einarsson, J. & Hultman, T. G. (1984). *Godmorgon pojkar och flickor: om språk och kön i skolan*. Malmö: Liber, 1st edition.

Eriksson, M. (1997). *Ungdomars berättande: en studie i struktur och interaktion*. PhD thesis, Uppsala University, Uppsala.

Fiske, S. T. & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. *Proceedings of the National Academy of Sciences*, 111(Supplement 4), 13593–13597.

Gale, W. A. & Church, K. W. (1991). A Program for Aligning Sentences in Bilingual Corpora. In *Proceedings of the 29th Annual Meeting on Association for Computational Linguistics*, ACL '91 (pp. 177–184). USA: Association for Computational Linguistics.

Gaucher, D., Friesen, J., & Kay, A. (2011). Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality. *Journal of personality and social psychology*, 101, 109–28.

Gender Spectrum (2019). Understanding Gender. https://www.genderspectrum.org/articles/understanding-gender. Accessed: 2021-05-19.

Haas, A. (1979). Male and female spoken language differences: Stereotypes and evidence. *Psychological Bulletin*, 86, 616–626.

Henry, P. (2010). Institutional Bias. In *The SAGE Handbook of Prejudice, Stereotyping and Discrimination* (pp. 426–440). SAGE Publications Ltd.

Hultman, T. G. & Westman, M. (1977). *Gymnasistsvenska*. Lund: LiberLäromedel.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., & Dean, J. (2017). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351.

Joulin, A., Bojanowski, P., Mikolov, T., Jégou, H., & Grave, E. (2018). Loss in Translation: Learning Bilingual Word Mapping with a Retrieval Criterion. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* Brussels, Belgium: Association for Computational Linguistics.

Kestemont, M. (2014). Function Words in Authorship Attribution. From Black Magic to Theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)* (pp. 59–66). Gothenburg, Sweden: Association for Computational Linguistics.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

Lakoff, R. (1973). Language and Woman's Place. *Language in Society*, 2(1), 45–79.

Lindmark, S. & Sundin, J. (2013). *Den könskodade jobbannonsen: En diskursanalys av jobbannonser för kvinnodominerade och mansdominerade yrken*. Bachelor thesis, Umeå University, Umeå.

Madera, J., Hebl, M., & Martin, R. (2009). Gender and Letters of Recommendation for Academia: Agentic and Communal Differences. *The Journal of Applied Psychology*, 94, 1591–9.

Martinsson, S. (2021). *Engagerad eller driven?* Bachelor thesis, Lund University, Lund.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and Fairness in Machine Learning. *arXiv e-prints*, arXiv:1908.09635.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In Y. Bengio & Y. LeCun (Eds.), *1st International Conference on Learning Representations, ICLR* Scottsdale, Arizona, USA: Workshop Track Proceedings.

Ng, L., Madera, J., Zajac, S., & Hebl, M. (2020). When Words Matter: Communal and Agentic Language on Men and Women's Resumes. *Academy of Management Proceedings*, 2020, 20462.

Nordberg, B. (1972). Morfologiska variationsmönster i ett centralsvenskt stadsspråk. In B. Loman (Ed.), *Språk och samhälle [1] Språksociologiska problem* (pp. 14–44). Lund: LiberLäromedel.

Nordenstam, K. (1990). Hur talar kvinnor och män till vardags? *Kvinnovetenskaplig tidskrift*, 11(1990):1, s. 32-42, 60.

Norrby, C. (1998). *Vardagligt berättande: form, funktion och förekomst*. PhD thesis, University of Gothenburg, Gothenburg.

Norrby, C. & Håkansson, G. (2015). *Introduktion till sociolingvistik*. Lund: Studentlitteratur, 2nd edition.

Östlund, E. (2018). *Könskodade jobbannonser,snarare en regel änett undantag: En studie av jobbannonser i den svenska mediebranschen*. Bachelor thesis, Umeå University, Umeå.

Pennebaker, J., Booth, R., Boyd, R., & Francis, M. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates (www.LIWC.net).

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543).

Phelan, J. E., Moss-Racusin, C. A., & Rudman, L. A. (2008). Competent Yet Out in the Cold: Shifting Criteria for Hiring Reflect Backlash Toward Agentic Women. *Psychology of Women Quarterly*, 32(4), 406–413.

Pietraszkiewicz, A., Formanowicz, M., Gustafsson Sendén, M., Boyd, R. L., Sikström, S., & Sczesny, S. (2019). The Big Two Dictionaries: Capturing Agency and Communion in Natural Language. *European Journal of Social Psychology*, 49(5), 871–887.

Pilcher, J. & Whelehan, I. (2004). *Fifty Key Concepts in Gender Studies*. SAGE Key Concepts. London: SAGE Publications.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 101–108). Online: Association for Computational Linguistics.

Rothchild, J. (2007). Gender Bias. In G. Ritzer (Ed.), *The Blackwell Encyclopedia of Sociology* (pp. 1843–1845). Oxford: Blackwell. https://onlinelibrary.wiley.com/doi/10.1002/9781405165518.wbeosg011.pub2.

Rudman, L. A. & Kilianski, S. E. (2000). Implicit and Explicit Attitudes Toward Female Authority. *Personality & Social Psychology Bulletin*, 26(11), 1315–1328.

Sahlgren, M. & Olsson, F. (2019). Gender Bias in Pretrained Swedish Embeddings. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 35–43). Turku, Finland: Linköping University Electronic Press.

Sidanius, J. & Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press.

Statistiska Centralbyrån (2012). MIS 2012:1, SSYK 2012 standard för svensk yrkesklassificering. https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/.

Statistiska Centralbyrån (2018). Anställda (yrkesregistret) 16-64 år efter kön och yrke (SSYK 2012). https://www.statistikdatabasen.scb.se/sq/99307.

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating Gender Bias in Natural Language Processing: Literature Review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1630–1640). Florence, Italy: Association for Computational Linguistics.

Sundgren, E. (2007). Stereotypa uppfattningar om kvinnor och män? ingvar och ingvor lind i ny tappning. In B.-L. Gunnarsson, S. Entzenberg, & M. Ohlsson (Eds.), *Språk och kön i nutida och historiskt perspektiv: studier presenterade vid Den sjätte nordiska konferensen om språk och kön, Uppsala 6-7 oktober 2006*, volume 71 of *Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet* (pp. 253–263). Uppsala: Uppsala universitet.

Sweeney, L. (2013). Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising. *Queue*, 11(3), 10–29.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17 (pp. 6000–6010). Red Hook, NY, USA: Curran Associates Inc.

9 Appendices

9.1 Instructions for qualitative evaluation

9.1.1 Original instructions in Swedish

Hej!

Här kommer översättningsutvärderingen. Varje fil har 50 rader och 10 kolumner: den första kolumnen (som bara består av siffror) kan ni bortse från; kolumnen eng består av engelska ord som översätts i de resterande kolumnerna; kolumn 0 består av översättningar som ska utvärderas i kolumn eval 0; kolumn 1 består av översättningar som ska utvärderas i kolumn eval 1; kolumn 2 består av översättningar som ska utvärderas i kolumn eval 2; kolumn 3 består av översättningar som ska utvärderas i kolumn eval 3. Orden utvärderas på en skala från -1–+1, där -1 innebär att översättningen är dålig, +1 innebär att översättningen är bra, och 0 innebär att översättningen varken är bra eller dålig (exempelvis kan det vara så att en översättning inte är riktigt lika bra som den bästa översättningen, men inte riktigt lika dålig som den sämsta).

Om det engelska ordet som ska översättas är felstavat kan ni behandla det som om det vore rättstavat. Det kan hända att ordet som ska översättas inte är ett ord (exempelvis om två ord är ihopklumpade) och att alla översättningar är helt felaktiga – då kan ni sätta -1 på alla.

För att snabba på utvärderingen kan ni använda bokstäver istället för -1, 0 och +1 (valfritt såklart). Använd isåfall a = -1, s = 0, d = +1.

Filen ex.csv är ett exempel.

Om ni har några frågor är det såklart bara att höra av er.

Tack! Saga

9.1.2 Translation of the instructions

Hi!

Here is the translation evaluation. Each file consists of 50 rows and 10 columns: the first column you can ignore; the column eng consists of English words that are translated in the remaining columns; column 0 consists of translations to be evaluated in column eval 0; column 1 consists of translations to be evaluated in column eval 1; column 2 consists of translations to be evaluated in column 2 eval; column 3 consists of translations to be evaluated in column 3 eval. The words are to be evaluated on a scale from -1–+1, where -1 means that it's a poor translation, +1 means that it's a good translation, and 0 means that it's neither a good, nor a poor translation (for example, a translation may not be quite as good as the best translation, but not as poor as the worst).

If the English word to be translated is misspelled, you can treat it as if it were correctly spelled. The word to

be translated may not be a word (for example, two words may be clumped together), and thus, all translations are wrong – in that case, you can evaluate all words as -1.

To speed up the evaluation, you can use letters instead of -1, 0 and +1 (optional, of course). In that case, use a = -1, s = 0, d = +1.

The file ex.csv is an example file.

If you have any questions, please get in touch.

Thank you! Saga

9.2 Occupational statistics for the KTH areas in Swedish

Occupation Area	Women %	Men %	Gender	# KTH advertisements
Bygg och anläggning	13%	87%	male	39
Installation, drift, underhåll	21%	79%	male	35
Industriell tillverkning	22%	78%	male	9
Data/IT	23%	77%	male	322
Säkerhetsarbete	30%	70%	male	10
Tekniskt arbete	34%	66%	male	656
Total # for male dominated areas	**29%**	**71%**		**1071**
Hantverksyrken	42%	58%	equal	1
Naturvetenskapligt arbete	44%	56%	equal	749
Chefer och verksamhetsledare	44%	56%	equal	51
Pedagogiskt arbete	47%	53%	equal	2174
Försäljning, inköp, marknadsföring	49%	51%	equal	43
Kultur, media, design	55%	45%	equal	72
Administration, ekonomi, juridik	56%	44%	equal	1048
Sanering och renhållning	57%	43%	equal	15
Hotell, restaurang, storhushåll	58%	42%	equal	2
Total # for equal areas	**49%**	**41%**		**4155**
Hälso- och sjukvård	73%	27%	female	8
Total # for female dominated areas	**73%**	**27%**		**8**

Table 13: Occupational statistics for the KTH areas.